Logic, Epistemology, and the Unity of Science 59

Gabriele Usberti

Meaning and Justification. An Internalist Theory of Meaning

Springer

Logic, Epistemology, and the Unity of Science

Volume 59

Logic, Epistemology, and the Unity of Science aims to reconsider the question of the unity of science in light of recent developments in logic. At present, no single logical, semantical or methodological framework dominates the philosophy of science. However, the editors of this series believe that formal frameworks, for example, constructive type theory, deontic logics, dialogical logics, epistemic logics, modal logics, and proof-theoretical semantics, have the potential to cast new light on basic issues in the discussion of the unity of science.

This series provides a venue where philosophers and logicians can apply specific systematic and historic insights to fundamental philosophical problems. While the series is open to a wide variety of perspectives, including the study and analysis of argumentation and the critical discussion of the relationship between logic and philosophy of science, the aim is to provide an integrated picture of the scientific enterprise in all its diversity.

This book series is indexed in SCOPUS.

For inquiries and submissions of proposals, authors can contact Christi Lue at christi.lue@springer.com

Gabriele Usberti

Meaning and Justification. An Internalist Theory of Meaning

Springer

Gabriele Usberti
Department of Social, Political
and Cognitive Sciences
University of Siena
Siena, Italy

ISSN 2214-9775 ISSN 2214-9783 (electronic)
Logic, Epistemology, and the Unity of Science
ISBN 978-3-031-24604-3 ISBN 978-3-031-24605-0 (eBook)
https://doi.org/10.1007/978-3-031-24605-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

For Anne Rohozinski (Anne Zénour)

For Tommaso and Piero Usberti

To the memory of three friends who left too soon:
Pierangelo Miglioli (1946–1999)
Paolo Casalegno (1952–2009)
Edoardo Ferrario (1946–2017)

Preface

The aim of this book is to lay the foundations of a theory of meaning for natural languages inspired by a view of semantics as the study of the relations of language not with external reality but with internal, or mental, reality.

The fundamental motivation for such a theory—exposed in Chap. 1—is that traditional, externalist semantics is necessarily based on common-sense views of language and reality—what prevents it from being a scientific enterprise; in this vein, Chomsky's criticisms of Fregean reference are interpreted as an argument for the incompatibility of a computational-representational theory of language with the externalist notion of reference; an analogous argument against realist truth is formulated. On the contrary, the realism about the mental of contemporary cognitive psychology provides the basis for a scientific approach to internalist semantics; in this connection, a deep convergence between internalist and anti-realist reasons against externalism comes to light.

It is the tradition of semantic anti-realism that has introduced and studied the fundamental ingredients of an internalist theory of meaning, in particular the notions of proof (of a mathematical sentence) and of justification (of an empirical sentence); that's why in Chap. 2, different varieties of semantic anti-realism are considered, in order to individuate the most akin to internalist motivations: intuitionism is chosen, provided it is understood as a form of realism about the mental.[1]

In Chap. 3, the intuitive notion of justification relevant to the theory to be developed is distinguished from other kinds of justification through some characteristics: it must be epistemic, *ex ante*, defeasible, non-factive, and, most importantly, epistemically transparent. An argument is given to equate epistemic justifications with cognitive states, where cognitive states are characterised in computational terms. Finally, the notion of truth-ground is introduced in order to explain the assertibility conditions of empirical sentences. In the conclusion, a more explicit formulation is given of

[1] My intention in this book is to develop an anti-realist approach, not to argue directly against realist proposals. That's why I have chosen, in particular, *not* to discuss such an important work as Khlentzos (2004), even if its scope partially overlaps mine, but with a diametrically opposite and in a sense complementary intention: the research of a form of realism which «could answer the anti-realist challenge» (Khlentzos 2004: vii).

the sort of anti-realism about external world and realism about mental world I am advocating.

Chapters 4 and 5 are devoted to the definition of the theoretical notions of $\mathcal{C}$-justification for α and $\mathcal{C}$-truth-ground of α ('$\mathcal{C}$' for 'cognitive', but also for 'computational'), where α is (the formalisation of) a mathematical or an empirical sentence. The definition is by induction on the logical complexity of α. Atomic sentences are the object of Chap. 4. As a preliminary to the definition of justification for them, the notion of prelinguistic cognitive state is characterised, in terms of which the basic notions of an internalist ontology—$\mathcal{C}$-object and $\mathcal{C}$-concept—are defined. Linguistic cognitive states are then introduced, and with them the notions of $\mathcal{C}$-authorisation to use a singular term to refer to a given object and $\mathcal{C}$-authorisation to use a predicate to apply a manageable concept to $\mathcal{C}$-objects. Predication is conceived as functional application guided by inference to the best explanation, and this, in turn, is a particular kind of computation; in this way, the epistemic transparency of both justifications for and truth-grounds of atomic sentences is assured. It is shown how the 'uninteresting' problems generated, according to Chomsky, by the externalist assumption illustrated in Chap. 1 vanish within the internalist semantical framework outlined, and how, on the other hand, a direction can be suggested for the solution of some 'interesting' problems. The chapter terminates with the proposal of a theoretical *explicans* of the intuitive notion of synonymy, individuated through an analysis of our intuitions about synonymy of expressions of various categories; with a comparison of the present approach with the neo-verificationist model of sense, based on a distinction between direct and indirect methods of object identification; and with a discussion of the so-called frame problem as it has been formulated by Fodor.

Logically complex sentences are dealt with in Chap. 5. The definition is inspired by Heyting's inductive definition of the notion of proof of a mathematical sentence, once that definition is conveniently reinterpreted for the reasons adduced in Sect. 5.1; in Sect. 5.2, the notion of evidential factor for α is introduced, after having argued for the necessity of a notion of empirical negation besides intuitionistic negation; in Sect. 5.3, $\mathcal{C}$-justifications and $\mathcal{C}$-truth-grounds of logically complex sentences are defined; Sect. 5.4 contains a characterisation of the logical concepts as they are denoted by the logical constants and a definition of $\mathcal{S}$-validity and logical validity. In Sect. 5.5, it is shown that possession of an $\mathcal{S}$-justification for an arbitrary sentence is epistemically transparent by outlining a 'non-objectual' model of evidence possession; as a consequence, a way of justifying inference alternative to the neo-verificationist one is suggested.

Chapter 6 is concerned with the relation between the intuitive notions of truth and evidence or truth-recognition. While the intuitionists grant no space to the (intuitive) notion of truth of a mathematical sentence (Sect. 6.1), according to many supporters of anti-realist theories of meaning, in particular neo-verificationist ones, the intuitionistic attitude is unacceptable because, on the one hand, it is highly counterintuitive, and on the other hand, some notion of truth, irreducible to proof possession, cannot be avoided even within an anti-realist conceptual framework. In Sect. 6.2, two arguments for the necessity of a distinction between truth and truth-recognition are analysed and criticised; Sect. 6.3 discusses the neo-verificationist debate between temporalist and atemporalist conceptions of truth.

Chapters 7 and 8 propose an extension of the internalist approach to doxastic and epistemic reports, respectively. In the introduction to Chap. 7 (Sect. 7.1), it is argued that the *De Dicto/De Re* ambiguity conceals in fact two different ambiguities: the Transparent/Opaque one and the Epistemic Specific/Non-specific one. The former is analysed in Sect. 7.2; in Sects. 7.2.1–7.2.5, it is argued that the foundational puzzles concerning it (Frege's, Mates's, and Kripke's puzzles) do not admit of an optimal solution within the framework of externalist semantics; in Sects. 7.2.6–7.2.10, the distinction is analysed and formally represented, within the present internalist framework, as concerning not two kinds of belief but two different propositions semantically expressed by the subordinate clause of the belief report, for the believer and for the reporter, respectively. In Sect. 7.3, a solution to the Paradox of Analysis is suggested. The Specific/Non-specific ambiguity is analysed in Sect. 7.4; in Sect. 7.4.1, it is argued that it cannot be represented in terms of scope; in Sects 7.4.2–7.4.5, it is analysed and formally represented by taking advantage of the difference, within the present internalist semantics, between the meanings of $\exists x\alpha$ and $\neg\forall x\neg\alpha$.

The first part of Chap. 8 (Sects. 8.1–8.3) proposes an extension of the definition of $\mathcal{C}$-justification to epistemic reports and an analysis of Gettier problems. More specifically, in Sect. 8.1, the definition is given; in Sect. 8.2, a representative class of Gettier problem is introduced and analysed; in Sect. 8.3, a comparison is made between the present approach and J. Pollock's analysis. The second part is devoted to two strictly connected themes: assertibility conditions of empirical sentences (Sect. 8.4), defined in such a way as to result epistemically transparent; and Williamson's argument(s) against transparency (or luminosity, in his terminology) of knowledge (Sect. 8.5).

Chapter 9 discusses the significance of the Paradox of Knowability with respect to the question of how to conceive truth within an anti-realist conceptual framework. After a description of the so-called Paradox (Sect. 9.1), the intuitionistic equation of truth with knowledge is articulated (Sect. 9.2), first by putting into evidence (Sects. 9.2.1 and 9.2.3) the conditions at which the equation is acceptable—transparency of knowledge and 'disquotational property' of truth—then by showing (Sects. 9.2.4 and 9.2.5) how a charge of inconsistency can be resisted. In Sect. 9.3, the neo-verificationist approaches to the Paradox are discussed, and in Sect. 9.4, the Dummettian problem is tackled of how a debate between alternative logics can be rationally shaped.

The semantics I will develop is not intended to be applicable directly to a natural language, but to a first-order formal language whose atomic sentences are thought as either mathematical or empirical; the semantic interpretation of natural language sentences will be thought as resulting from the composition of the interpretation of the formal language with a translation from the natural language into the formal one; in most cases, this translation will be assumed to be the standard one, except in some crucial cases, which I shall explicitly examine and discuss (especially in Chap. 7).

The main reasons of this strategy are the following.

(i) One of the hallmarks of the semantics I will develop is the explanation of predication, which is different both from the externalist one based on the set theoretic relation of membership and from other internalist ones; the presentation of the proposed explanation is all the more transparent, the simpler the syntax of the reference language.
(ii) Another feature distinguishing the present approach from other internalist ones is the fact that it explains in detail how we know inferential relations, including logical ones, among sentences; to this purpose, it is very useful to make reference to a language much simpler than a natural language, a sort of language skeleton, of which the logical constants are nevertheless a salient constitutive element.
(iii) A third peculiarity of my approach is the confluence, in it, of motivations coming from apparently very distant, but in my opinion converging, conceptual frameworks: Chomsky's methodological internalism, on the one side, and Heyting's explanation of the meaning of the logical constants, on the other; again, the focus is on the logical constants.
(iv) An important application of the whole approach concerns belief reports, in particular the formal characterisation of the Epistemic Specific/Non-specific distinction, in which an essential role is played by the difference between classical and intuitionistic explanations of the existential quantifier; reference to a standard first-order language permits a more clear formulation of the basic ideas.

As it can be seen, the reasons of my choice are essentially practical; on the other hand, the choice may involve some oversimplification or some distortion, as for instance the assumption that a first-order language is a sort of logical skeleton of a natural language; but, as far as I can see, the internalist inspiration of my approach in no way entails the theoretical necessity of such oversimplifications or distortions.

As a matter of fact, I will introduce several formal languages, starting from a basic one, $\mathcal{L}$, and then extending it in various ways as needed.

(1) The Language $\mathcal{L}$

Primitive symbols

- An infinite set $\mathcal{V}$ of *C*-object variables $\xi, \xi_1, \xi_2, \ldots$
- An infinite set $\mathcal{N}_o$ of names of *C*-objects
- An infinite set $\mathcal{P}^n$ of n-place predicates, for all natural numbers n>0
- The logical constants: $\wedge, \vee, \rightarrow, \perp, \forall, \exists$
- The predicates of objectual identity $=$ and of conceptual identity $\equiv$

Singular terms

- If $\tau \in \mathcal{V}$, then τ is a singular term

- If $\tau \in \mathcal{N}_o$, then τ is a singular term

Formulas

- If $\pi \in P^n$ and $\tau_1,\ldots,\tau_n$ are singular terms, then $\pi(\tau_1,\ldots,\tau_n)$ is a formula
- If τ_1, τ_2 are singular terms, then $\tau_1 = \tau_2$ is a (n *atomic*) formula
- If π_1, $\pi_2 \in P^n$, then $\pi_1 \equiv \pi_2$ is a (n *atomic*) formula
- $\bot$ is a formula
- If α and β are formulas, then $(\alpha \wedge \beta)$, $(\alpha \vee \beta)$, $(\alpha \rightarrow \beta)$ are formulas
- If α is a formula and $\xi \in \mathcal{V}$, then $\forall \xi \alpha$, $\exists \xi \alpha$ are formulas

Definition 1.
$\neg \alpha =_{\text{def}} \alpha \rightarrow \bot$
Definition 2.
$\alpha \leftrightarrow \beta =_{\text{def}} (\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$
Sentences are formulas containing no free occurrences of variables.

(2) The Language $\mathcal{L}_\sim$

The extension of $\mathcal{L}$ obtained by adding the symbol ~ and the relative clause:
If α is a formula, then $\sim\alpha$ is a formula.

(3) The Language $\mathcal{L}_{\mathbf{Bel}}$

The extension of $\mathcal{L}$ obtained by adding to the clauses defining $\mathcal{L}$ the following new ones:

Primitive symbols

- An infinite set $\mathcal{N}_\mathcal{S}$ of names of subjects
- The 2-place predicate: B
- The propositional operators: $\text{that}_{\mathcal{Bel}}$, $\text{that}_{\mathcal{Rep}}$

Singular terms

- If $\tau \in \mathcal{N}_\mathcal{S}$, then τ is a singular term

Propositional terms (p-terms)

- If α is a sentence, then $\text{that}_{\mathcal{Bel}}\alpha$, $\text{that}_{\mathcal{Rep}}\alpha$ are p-terms

Formulas

- If α is a formula and $\tau \in \mathcal{N}_\mathcal{S}$, then $B(\tau, \text{that}_{\mathcal{Bel}}\alpha)$, $\text{Bel}(\tau, \text{that}_{\mathcal{Rep}}\alpha)$ are formulas.

(4) The Language $\mathcal{L}_{\mathbf{Bel,\ Kn}}$

The extension of $\mathcal{L}_B$ obtained by adding to the clauses defining $\mathcal{L}_B$ the following new ones:

Primitive symbols

- The propositional operator: that
- The 2-place predicate: Kn

Formulas

- If α is a formula and $\tau \in \mathcal{N}_S$, then Kn(τ, thatα) is a formula

Definition 2.

$K\alpha =_{def} \exists S(Kn(S,\alpha))$

Some general notational conventions.

Throughout this book, α[ε] is a sentence α containing at least one occurrence of the expression (singular term, predicate, or sentence) ε; α[ε'/ε] is the result of replacing, within α, *every* occurrence of ε by an occurrence of ε' (where ε' is an expression of the same syntactical category as ε); α[ε'//ε] is the result of replacing, within α, *some* occurrences of ε by an occurrence of ε'.

When necessary (in particular in Chap. 9), I will distinguish intuitionistic logical constants from classical ones, using the logical symbols of $\mathcal{L}$ for the former and the following for the latter: &, $+$, $\supset$, $\equiv$, $-$, $\prod$, $\sum$.

Siena, Italy
June 2023

Gabriele Usberti

Reference

Khlentzos, D. (2004). *Naturalistic realism and the antirealist challenge*. The MIT Press.

Acknowledgements

Thanks are due to many people for discussions, long letters, encouragements, advices, suggestions, severe but constructive criticism, or for their affection: they include Paola Arcari, Marina Astrologo, Carla Bagnoli, Ermanno Bencivenga, Beatrice Bernazzi, Valentina Bianchi, Andrea Bonomi, Andrea Cantini, Gianni Carmignani, Paolo Casalegno, Carlo Cellucci, Noam Chomsky, Annalisa Coliva, Cesare Cozzo, Marie-Noëlle de Rohozinska, Michael Dummett, Donata Feroldi, Pasquale Frascolla, Paul Horwich, Andrea Iacona, Pierre Jacob, Josiane Maillard, Diego Marconi, Enrico Martino, Pierangelo Miglioli, Francesco Montesi, Sabrina Mori, Enrico Moriconi, Julien Murzi, Marco Panza, Sacha Pavan, Talia Pecker Berio, Duccio Pianigiani, Antonio Piccolomini d'Aragona, Salvatore Pistoia Reda, Claudio Pizzi, Dag Prawitz, Giuseppe Maurizio Rossi, Luca San Mauro, Peter Schroeder-Heister, Emilio Servidio, Göran Sundholm, Luca Tranchini, Giuseppe Varnier, Alain Volut, Bernhard Weiss, Timothy Williamson, and Sandro Zucchi. Special thanks to Punchimenike Wijesinha Hearath Mudiyanselage for her generosity.

Chapter 1 is new.

Of Chap. 2, Sect. 2.1 is based on parts of Usberti (2006b); Sects. 2.2–2.4 contain material from Usberti (1990) and (1992), and from Chaps. 1 to 3 of Usberti (1995).

Chapter 3 is partially based on Usberti (2004).

Chapter 4 contains material from Usberti (2002), (2006a), and (2015), with substantial modifications.

Of Chap. 5, Sects. 5.1–5.3 are new; Sect. 5.4 contains material from Usberti (2019b); Sect. 5.5 is partially based on Usberti (2019a).

Sections 6.1 and 6.2 of Chap. 6 are based on Usberti (2012); Sect. 6.3 is new.

Chapter 7 is wholly new.

Of Chap. 8, Sects. 8.1–8.4 are based on parts of Usberti (2014); Sect. 8.5 is new.

Chapter 9 is partially based on Usberti (2016), with substantial modifications.

I thank the editors concerned for permission to use this material.

References

Usberti, G. (1990). Prior's disease. *Teoria*, *10*(2),131–138.
Usberti, G. (1992). *Constructivism in logic*. Unicopli.
Usberti, G. (1995). *Significato e conoscenza*. Guerini.
Usberti, G. (2002). Names: Sense, epistemic content, and denotation. *Topoi, 21*, 99–112.
Usberti, G. (2004). On the notion of justification. *Croatian Journal of Philosophy*, *4*(10), 99–122.
Usberti, G. (2006a). Towards a semantics based on the notion of justification. *Synthese, 149*(3), 675–699.
Usberti, G. (2006b). The semantic characterisation of the realism/anti-realism debates. In E. Ballo & M. Franchella (Eds.), *Logic and philosophy in Italy* (pp. 447–463). Polimetrica.
Usberti, G. (2012). Anti-realist truth and truth-recognition. *Topoi, 31*(1), 37–45.
Usberti, G. (2014). Gettier problems, C-justifications, and C-truth-grounds. In E. Moriconi & L. Tesconi (Eds.), *Second Pisa colloquium in logic, language and epistemology* (pp. 325–361). ETS.
Usberti, G. (2015). A notion of C-justification for empirical statements. In H. Wansing (Ed.), *Dag Prawitz on proofs and meaning* (pp. 415–450). Springer.
Usberti, G. (2016). The paradox of knowability from an intuitionistic standpoint. In T. Piecha & P. Schroeder-Heister (Eds.), *Advances in proof-theoretic semantics* (pp. 115–137). Springer.
Usberti, G. (2019a). Inference and epistemic transparency. *Topoi, 38*, 517–530.
Usberti, G. (2019b). A notion of internalistic logical validity. In L. Bellotti, L. Gili, E. Moriconi & G. Turbanti (Eds.). *Third Pisa colloquium in logic, language and epistemology* (pp. 389–406). ETS.

Introduction

Perhaps, the most concrete way to explain the motivation behind this book is to start with a case that exemplifies a situation in which a scholar of semantics and theory of meaning often finds himself. The scholar, in this case, is Barbara Partee, who in a 1973 paper discusses Carnap's 1954 response to Mates's 1950 counterexample to his proposed analysis of belief reports in terms of intensional isomorphism. Carnap's response (suggested by Church) consisted of two steps: (i) rejecting a subject's disposition to assent to (or dissent from) (the assertoric utterance of) a sentence as a criterion for attributing (or not attributing) a belief, accepting it only as inductive support for that attribution (or non-attribution), and (ii) stipulate that the sentence

(1) Whoever believes that D, believes that D'

is true for synonymous D and D', in spite of the dissent from (1) of any subject who doesn't know that D and D' are synonymous. Partee comments that

> Such a stipulation is quite appropriate for the philosopher engaged in rational reconstruction. But the linguist, although he may agree wholeheartedly that 'believes' is a term for whose correct application no single kind of observational evidence is criterial, is not thereby free to discount a priori whatever observational evidence happens to conflict with his favorite hypothesis. (Partee 1973: 316)

Let us take a step back and consider the situation with the attention it deserves, starting with the philosopher. According to Frege, «To discover truths is the task of all sciences; it falls to logic to discern the laws of truth» (Frege 1918: 352). Wittgenstein's concept of philosophy (of which logic is a constitutive part) seems to generalise this idea: it is not for philosophy to discover new truths, but to clarify our vision of what we see. In this sense, philosophy is essentially rational reconstruction, and this is how the theory of meaning should be understood, according to Frege and to a large part of actual analytical philosophers. It seems to me that, implicit in Partee's commentary, there is also the recognition that this philosophical enterprise has produced remarkable results. In particular, it has led to a semantics for natural languages that has been extended to doxastic and epistemic reports, exposing their intricacies and ambiguities. And yet, precisely in this crucial area, it seems to lead to outcomes that a linguist, as an empirical scientist, cannot accept. The essential reason why Carnap stipulates that (1) is true is that, as a consequence of the substitutivity principle, (1) is logically true when D and D' are synonymous; the reason is undoubtedly very serious: substitutivity is an immediate consequence of compositionality, thus of the fundamental principle of Fregean semantics; but the requirement that a theory of belief be adequate to observational evidence is an equally serious reason, for an empirical science such as linguistics or, more generally, psychology, for the opposite conclusion that (1) is not true.

The examples can be multiplied, and the detailed analysis that is conducted in Chap. 7 of the solutions proposed for Mates's puzzle and for the other foundational problems of the semantics of doxastic reports (Frege's and Kripke's puzzles, the *De Dicto/De Re* distinction, the Specific/Non-specific distinction, etc.) highlights a profound tension between the empirical predictions of the most accredited belief theories and the indispensable requirement of descriptive adequacy of such theories. More precisely, the analysis identifies as responsible for the tension a fundamental assumption common to theories that are in other respects very different, such as Frege's and Russell's: that the fundamental semantic relations—of denotation of a singular term and of satisfaction of a predicate—, and thus all notions that can be defined in terms of them—including sense and synonymy—are relations between linguistic expressions and entities of the external world, hence public. If we abandon this assumption, it becomes possible to approach Mates's puzzle on the basis of the idea that, if D and D' are synonymous *for us*, but a speaker S believes that D and does not believe that D', then D and D' are not synonymous *for S*; this is the most intuitively natural idea, and it resolves the tension described; but it requires a semantic paradigm shift, as it makes it necessary to develop the semantics of a language by understanding it as the study of relations between linguistic expressions and entities of the *internal*, i.e. mental, world of speakers.

The reasons for such a paradigm shift certainly have to do not only with the semantics of belief reports, but involve the very possibility of grounding semantics as a scientific enterprise. They are identified in the Chap. 1 and come from two traditions that are in other respects very different from each other: on the one hand, the study of linguistic competence elaborated since the middle of the last century by Chomsky and on the other hand, the anti-realist tradition represented in particular, in the philosophy of mathematics, by the intuitionism of Brouwer and Heyting and, in the theory of meaning, by the neo-verificationism of Dummett, Prawitz, Martin-Löf, and others. Chomsky made a series of objections to the externalist notion of reference placed by Frege at the basis of his semantics, which highlight the aporias and false problems arising from the acceptance of common-sense ontology; intuitionists and neo-verificationists rejected, with different arguments, the externalist-realist notion of truth as transcending human cognitive capacities; the two traditions can be seen as converging in criticising the externalism and realism of the semantic tradition for its inability to account in scientific terms for knowledge of meaning. If one accepts this criticism, the problem that naturally arises is whether and how it is possible to account for semantic competence from an internalist point of view.

The conceptual framework within which to address this problem is that of psychology as an empirical science. Frege's objections to psychologism, in particular to the possibility that linguistic expressions denote mental entities, are well known, together with his arguments in support of the necessity that thoughts expressed by sentences belong to a realm of external and therefore public entities, and consequently that language is a social institution. Nonetheless, the very development of generative linguistics as a part of psychology, i.e. of the empirical study of mind, constitutes a response to Frege's anti-psychologism; that representations need a bearer does not entail that they are private: the heart also needs a bearer, but it is not private in the sense (relevant here) that it cannot be studied scientifically. More precisely, Chomsky laid the methodological and conceptual foundations for a scientific study of the language faculty, and more generally of mind. The methodological basis is internalism, which is a consequence of the idea that the best type of scientific explanation in psychology is the computational-representational one: mental processes are computations, whose inputs and outputs are representations, hence complex symbols of some representational system internal to the mind; any possible connection between these representations and entities of the external world is simply irrelevant.

On the other hand, one of the fundamental theses of Chomskyan linguistics is that the study of linguistic competence as a scientific enterprise entails abandoning the idea that language is a social institution. A language, understood as what a speaker knows, cannot be a potentially infinite set of sentences with meaning: it can only be the grammar of that language.[2] How does one learn a grammar? A crucial argument, the argument from the poverty of the stimulus, shows how such learning cannot be inductive in nature, as most traditional theories of meaning assumed. Chomsky elaborated a radically different model, according to which the acquisition of a language by a child is actually a process of selection: the starting point, the 'initial state' of the mind, common to all individuals as part of the genotype, incorporates the universal grammar, containing a set of *principles* placing restrictions on possible grammars, and a set of open *parameters* whose values are set by exposition to linguistic data (e.g. may a sentence have a null subject?): setting the parameters in a certain way virtually amounts to selecting one grammar (I-language) among the vast class of possible ones.

Between universal grammar and individual grammar—the only two entities that have psychological reality—there is no room for languages understood as norms and conventions to which nations, societies, and communities conform: psychology, understood as a science that adopts the methods and standards of Galilean science,[3] 'sees' what is biologically determined, hence invariant across the species, and what is specific to the individual (the competence the individual has of her/his own language), not society. Chomsky calls E-languages the social institutions regulated by norms

[2] On this point, see the comments to Lewis (1975) in Chomsky (1980: 81–85).

[3] For example, the systematic use of mathematics in the creation of abstract models, the introduction of appropriate idealizations, the experimental method, inference to the best explanation, and so on.

and conventions and I-languages the grammar within the mind of each individual speaker.[4]

Within this conceptual framework, explaining semantic competence amounts to proposing models of the computational processes through which the speaker's mind processes the relevant linguistic data. For instance, it is an empirical fact that, on the basis of our linguistic competence alone, we 'know' that if x is an uncle then x is a boy, that if John is a boy then it is false that John is not a boy, that if John is a boy then John is a boy or Mary is a girl, and so on and so forth; how do we account for the knowledge we have of these entailment relations? In order to answer this question, it is necessary to state what an entailment relation is; in an externalist semantics, it is characterised in terms of the externalist notion of truth, which is obviously not available in an internalist semantics; how to characterise it within an internalist conceptual framework? Similar problems arise in relation to the notion of denotation: there are empirical data indicating that on the basis of our linguistic competence alone, we 'know' that, under certain circumstances, two linguistic expressions denote the same object; how do we account for this kind of knowledge? In order to answer this question, it is necessary to address how to characterise, in an internalist framework, the notions of denotation, object, and identity between objects.

It is in relation to these questions that certain ideas developed by the anti-realist tradition become pertinent. In fact, as I anticipated, this tradition has (at least) two components, which should be clearly distinguished, the intuitionist and the neo-verificationist. In Chap. 2, they are presented and discussed in depth, in order to choose on which to base the semantics of the internalist theory of meaning I propose to set out. In a nutshell: an essential merit of the neo-verificationist approach is that it has clearly set the debate between realism and anti-realism on a semantic basis, i.e. as concerning the theory of meaning and thus the definition and role of such basic notions as reference and truth; decisive limitations, in my view, are the conception of language as a social institution, whose central 'function' is communication, and the hybridisation of the intuitionist tradition it draws on with formalist and Wittgensteinian themes, in particular anti-mentalism and the idea that 'meaning is use'. On the other hand, intuitionism can be seen as a form of realism about the mental—a necessary prerequisite for any computational-representational explanation—, and has elaborated, in the restricted domain of mathematical sentences, a systematic explanation of the meaning of logical constants, alternative to the Fregean externalist explanation, whose central notion is that of proof of a mathematical sentence: understanding a mathematical sentence, i.e. knowing its meaning, is equivalent to being able to recognise its proofs; since proofs can be seen as mental entities, this idea can become the cornerstone of an internalist approach to semantics and meaning theory. What is lacking in the intuitionist tradition is, firstly, an extension of the basic idea to empirical sentences: their meaning cannot be explained in terms of the notion of proof, but in terms of a defeasible notion such as that of justification; secondly, a compositional explanation of the meaning of atomic sentences is lacking.

[4] «"I" to suggest internal and individual, and also intensional, in that L is a specific procedure that generates infinitely many expressions of L.» (Chomsky 2000: 169).

Filling these two gaps is the purpose of Chaps. 3–5. In Chap. 3, the intuitive notion of justification is first analysed and defeasibility is identified as its essential characteristic. Secondly, a fundamental property that a theoretical notion of justification must possess in order to be able to play the role of a key notion of meaning theory is introduced: epistemic transparency; if knowing the meaning of α is equated to being capable to recognise justifications for α, then a subject who understands α must be in a position to recognise what counts as a justification for α. Finally, an argument is proposed in support of the thesis that justifications must be conceived as cognitive states that fulfil certain conditions.

What these conditions are is the subject of Chaps. 4 and 5. The first concerns atomic statements, both mathematical and empirical. The strategy I adopt is compositional: if knowing the meaning of an atomic utterance is equivalent to being able to recognise its justifications, this knowledge must be explained as resulting from knowledge of the meaning of the names and predicate that constitute it. In this context, an internalist approach cannot disregard how psychology conceives of object knowledge and concept mastery, in the sense that the proposed semantics will have to elaborate a notion of denoted object and a notion of concept compatible with the way such notions are conceived by computational-representational psychological theories. For example, some theories suggest that the mind of non-human animals has the basic machinery for representing predicate-argument structures (Gallistel 2011); a semantic theory compatible with such theories should accordingly characterise objects and concepts in such a way that they can be harboured in a prelinguistic mind; this is what I have tried to do. Chapter 5 deals with logically complex sentences. Although this is the chapter in which the reference to Heyting's intuitionist explanation of the meaning of logical constants is most evident, the concern to maintain the compatibility of the approach with the assumptions of psychology is also present, in particular in the sense of guaranteeing, for each logical constant, the epistemic transparency of justifications for sentences having that constant as their main operator. In this perspective, the conception, developed in Chap. 3, of justifications as cognitive states is essential; from this point of view, Heyting's explanation is reinterpreted as a definition not of the notion of proof of α (or justification for α, in the case where α is empirical), but of the notion of an *evidential factor* of α; and from this reinterpretation, a paradigm of logical validity arises that is very similar to the intuitionistic one, but in which the *ex falso quodlibet* is not valid.

Chapters 7 and 8 extend the semantics to doxastic and epistemic reports, respectively. It seems to me that the relativisation of semantic reports to cognitive subjects, made possible by the internalist nature of the approach, allows, in Chap. 7, for what I have called optimal solutions to the problems mentioned at the beginning, i.e. for a kind of reflective equilibrium between the theoretical predictions derived from the Substitutivity Principle and the requirement of descriptive adequacy of theory represented by the Disquotational Principle. The solution to Gettier's problems proposed in Chap. 8, on the other hand, exploits the justificationist nature of semantics, in that it allows the reasons for a justified belief to be formally represented, and thus the restrictions necessary to guarantee knowledge to be imposed on them. Finally, the epistemic nature of the notions on which semantics is based makes it possible, again in Chap. 7, to account for the distinction between specific and non-specific interpretations of indefinites, misleadingly identified by Quine with the Opaque/Transparent distinction. As can be seen, these chapters foreground a way of arguing in favour of an anti-realist paradigm that does not consist in a (in my opinion impossible) knock-down argument against truth-conditional semantics, but in showing how the adoption of an internalist conceptual framework and a justificationist semantics can solve in an empirically adequate and philosophically enlightening way the central problems of the semantics of belief and knowledge reports.

Chapters 6 and 9 are devoted to the identification of truth with knowledge, which is obvious from an intuitionist point of view, but also quite natural from an internalist point of view. Chapter 6 answers some neo-verificationist objections to this identification, and critically discusses Prawitz's and Martin-Löf's atemporalist conceptions of truth. Chapter 9 argues that a well-known realist objection to epistemic conceptions of truth, the Paradox of Knowability, actually affects neo-verificationist conceptions, which identify truth with knowability, but not the identification of truth with knowledge, provided that the metalinguistic logical constants are read intuitionistically. In other words, knowledge is the notion of truth 'internal' to the intuitionist theory of meaning.

The critical discussion of neo-verificationism and the comparison between it and Brouwer's and Heyting's intuitionism is a theme that runs throughout the book. In addition to the points already indicated, Chap. 2 discusses the idea that the rules of introduction of natural deduction are constitutive of the meaning of logical constants; the distinction between canonical and non-canonical proofs; the architecture of justification as reduction to canonical form; and the problem of the epistemic transparency of proofs. Chapter 3 criticises the choice of an indefeasible notion of justification as a key notion of the theory of the meaning of empirical sentences. Chapter 4 criticises the assumption of the existence of a 'favoured' or 'direct' method of identification of concrete objects. In Chap. 5, the neo-verificationist conception of the possession of evidence and the justification of inference are discussed.

Finally, a widespread objection to Chomsky's idea that the object of linguistics is the I-language of the individual speaker is that a language that is not shared by a community would not account for communication, i.e. the transmission of a thought from one speaker to another: «A language is shared by many, as a thought can be grasped by many», Dummett writes (Dummett 2010: 83), echoing an argument by Frege in 'Thoughts'. In the following pages, I will not consider this objection, but here I would like to conclude by suggesting an explicit answer. The objection does not seem convincing to me: communication can be accounted for without postulating that speakers share a language, as has been repeatedly suggested by Chomsky (e.g. in Chomsky 2000: 30). If the selection of an I-language takes place, from an initial state common to all individuals, by parameter fixation, it is highly probable that two subjects $\mathcal{S}_1$ and $\mathcal{S}_2$ belonging to the same community speak two very similar I-languages, and that they believe they do so; thus, when $\mathcal{S}_1$ listens to $\mathcal{S}_2$ speak, (s)he proceeds by assuming that $\mathcal{S}_2$'s I-language is identical to her/his own modulo M, where M contains several components, including information $\mathcal{S}_2$ associates to the names and predicates (s)he is using; then $\mathcal{S}_1$ seeks to determine M: sometimes an easy task, sometimes hard, sometimes impossible; the outcome may vary from adjusting for small differences to requiring substantial reasoning. Insofar as $\mathcal{S}_1$ succeeds in these tasks, (s)he understands what $\mathcal{S}_2$ says as being what (s)he means by her/his comparable expressions; if M cannot be determined, communication fails. No reference to a common language is necessary. The only ‘shared’ structure is universal grammar, the initial state of the language faculty.

References

Chomsky, N. (1980). *Rules and representations.* Columbia University Press.

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambrige University Press.

Dummett, M. (2010). *The nature and future of philosophy*. Columbia University Press.

Frege, G. (1918). Thoughts. In *Frege* (1984) (pp. 351–372)

Frege, G. (1984). *Collected papers on mathematics, logic, and philosophy*. In B. Mc Guinness (Ed.). Blackwell.

Gallistel, C. R. (2011). Prelinguistic thought. *Language Learning and Development, 7*, 253–262.

Lewis, D. (1975). Languages and language, In K. Gunderson (Ed.), *Language, mind, and knowledge*. University of Minnesota Press.

Partee, B. (1973). The semantics of belief-sentences. In J. Hintikka et al. (Eds.), *Approaches to natural language* (pp. 309–336). Reidel.

Contents

Chapter 1
Motivations for an Internalist Semantics

Abstract In the first part of this chapter I illustrate objections to Fregean semantics coming from two very different traditions like Chomsky's methodological internalism, on the one hand, and semantical anti-realism, on the other. In the second part I explain how those objections can be seen as converging to motivate an internalist program for semantics, and I argue that semantical anti-realism can significantly contribute to that program. More specifically, in Sect. 1.1, after a brief review of the reasons of Chomsky's equation of a scientific approach to the study of language with an internalist one, two objections are distinguished: the lack of explanatory power of the externalist notions of reference and truth, and their giving rise to puzzling and uninteresting questions. In Sect. 1.2 two objections coming from the anti-realist side are expounded: Dummett's so-called anti-realist argument and a variant of it, and Prawitz's objection to Tarskian definition of logical consequence. In Sect. 1.3 two alternative views of the scope of an internalist semantics are introduced, and reasons are adduced to prefer the second, 'thick semantics', to the first, 'thin semantics'. In the Conclusion the possibility of a convergence between methodological internalism and anti-realism about the external world is individuated in a realist attitude about the 'internal' or mental world.

Keywords Internalism · Chomsky · Reference · Truth · Externalist Semantics · Anti-Realism

1.1 Chomskyan Objections to Externalist Semantics

In order to understand Chomskyan objections to externalist semantics it is necessary to understand the rationale for his methodological internalism, and to understand this it is necessary to start from the dramatic difference existing, according to him, between the points of view of science and of common sense.

From Galileo to Descartes a coherent mechanical explanation of the natural world was developed, based on the idea that the world is a machine of the kind that could be constructed by a skilled craftsman, and on the crucial assumption, drawn from

G. Usberti, *Meaning and Justification. An Internalist Theory of Meaning*,
Logic, Epistemology, and the Unity of Science 59,
https://doi.org/10.1007/978-3-031-24605-0_1

common-sense understanding (or from folk physics), that objects can interact only through direct contact. The boundaries of the natural world were clearly marked by the limits within which the mechanical model was applicable; beyond those limits the domain of the *res cogitans* began. There was therefore a clear divide between what is physical and what is mental. With Newton this picture is upset. Newton postulates a force—gravitation—acting at a distance, which appears to be necessary for scientific explanation but does not belong to the conceptual framework of the mechanical philosophy, and therefore seems (to Newton himself) to be mysterious. This produces a definitive crisis of the mechanical model and, with it, of our idea of what is physical (not necessarily in the sense that we have no intuition of what is physical, but in the sense that we are no longer inclined to take our common-sense understanding as the basis for a general explanation of what is physical). This crisis has only been made worse by contemporary physics with the introduction of such notions as gravitational field or ten-dimensional space, that have sanctioned the end of the 'visualizability' of physical phenomena, according to the expression of W. Heisenberg. Another consequence of Newton's revolution concerns epistemology. I said that gravitation seemed to be necessary to scientific explanation, but also mysterious; mysterious from the point of view of the intelligibility criteria elaborated by the seventeenth century mechanical philosophy. As soon as gravitation was accepted into the conceptual framework of natural science, the point of view of common sense had to be abandoned.

The moral Chomsky draws from this story is twofold. On the ontological side, since no clear notion of body and of the physical is any longer at hand, the very project of a materialist (hence monist) or of a dualist answer to the mind–body problem cannot be formulated: there simply is not a materialist thesis, nor a dualist thesis, to be formulated in a coherent way. On the epistemological side, the *sole* criterion of intelligibility of the natural world we are left with is represented by the methods of the empirical sciences; there is no *other* point of view from which we can decree that a scientific theory is right or wrong. It is interesting that Chomsky conceives the epistemological moral he draws as somehow deriving from Kant, according to whom

> Pure mathematics and pure science of nature had, for their own safety and certainty, no need for such a deduction as we have made of both. For the former rests upon its own evidence, and the latter [...] upon experience and its thorough confirmation. The pure science of nature cannot altogether refuse and dispense with the testimony of experience; because with all its certainty it can never, as philosophy, imitate mathematics. Both sciences, therefore, stood in need of this inquiry, not for themselves, but for the sake of another science: metaphysics. (Kant, 1783: §40)

In other words, it was not mathematics and natural science that needed to be justified, but metaphysics that needed to be reinterpreted, and in this reinterpretation consisted Kant's Copernican revolution, which overturned the Leibnizian conception of the relationship between physics and metaphysics: the metaphysical concepts of substance and causality no longer describe a realm that exists 'behind' phenomena located in space and time, but constitute the spatiotemporal framework within which phenomena can be empirically described objectively.

From this point of view the materialism of so large a part of contemporary cognitive science appears to be a sort of long leap backwards from Kant to Leibniz, with the crucial difference that Leibniz's realism, and particularly his appeal to Aristotelian substances and final causation, was an attempt to 'think' a central notion for the dynamics of his time—the notion of *vis viva*—, while contemporary materialism does not seem to be motivated or sustained by anything coming from cognitive science. As a consequence, the standpoint from which such a materialism is proposed turns out to be a «mysterious point outside of science itself» from which philosophy shows «that our scientific knowledge somehow 'mirrors' an independently existing reality.»[1]

If we consider externalist semantics, i.e. semantics conceived as the study of the relationship between language and external reality, in the light of Chomsky's neat distinction between science and common sense, we realize that it is a hybrid intellectual enterprise: on the one hand it is conceived as a science, on the other its ontology is not the ontology of any science, but of common sense, inhabited as it is by persons, towns, rivers, nations, sunsets, and so on and so forth, which are not the objects of any science; as a consequence its fundamental notions—denotation of a name, satisfaction of a predicate by an object, hence truth of a sentence—are common-sense notions. In order to understand Chomsky's objections to a semantics and a theory of meaning based on such common-sense notions it may be useful to consider the standard argument for realist-externalist semantics. The argument can be summed up in the following steps:

(1) Reference and truth are realist-externalist notions, i.e. notions whose explanation involves the relation between language and the external, particularly physical, world;
(2) Reference and truth play an essential role in the explanation of meaning, and therefore in the explanation of our linguistic competence;
(3) Therefore, the relation between language and the physical world plays an essential role in the explanation of our linguistic competence.

Chomsky questions (2). It is often assumed that the essential reasons he offers against (2) stem from some deep metaphysical puzzles that he raises about the existence of things in the world for words to refer to.[2] This seems to me an oversimplification. Consider the following passage:

> We now suppose that LI [lexical item] has no I-meaning but that it [...] *S-denotes a semantic value* SV(LI) that is external to the person [...]. [T]he original project isn't advanced, merely restated, with many new problems. We have learned nothing more about how expressions are used and interpreted. (Chomsky, 2000: 178)

Here it is clear that Chomsky has *two* objections to (2): (i) the appeal to reference and truth in the explanation of meaning leaves the problems a linguist typically is confronted with untouched; (ii) it creates many new problems, uninteresting for the

[1] Friedman (1993: 48), quoted in Chomsky (2000: 112).

[2] See for instance Casalegno (1997), 357 ff.

linguist. Taken together, (i) and (ii) show that the critique concerns both the possibility and the relevance or the *interest* of using the realist-externalist notions of reference and truth in the explanation of meaning and therefore of linguistic competence.

1.1.1 Externalist Notions Have No Explanatory Value

1.1.1.1 Reference

Let me try to explain Chomsky's first objection. Which are the problems a linguist is typically confronted with? In general, problems concerning «internally-determined properties of linguistic expressions» (Chomsky, 2000, 34). Here is a paradigmatic example involving reference. Let us consider the following sentences:

(4) His mother thinks that John is abroad
(5) He thinks that John is abroad.

We intuitively remark (and assume as a datum) that in (4) "his" *may* be coreferential with "John", whereas in (5) "he" cannot. The problem is: how do we explain this fact concerning our linguistic competence? A subtheory of the theory of syntax, Binding Theory, explains this datum by means of Principle C: "A referential expression cannot be c-commanded by an expression with the same referential index"; or, equivalently: "A referential expression must be free everywhere (in all categories)". Since, by the definition of c-command, in (5) "he" c-commands "John", it follows that in (5) "he" and "John" cannot have the same referential index; on the other hand, since in (4) "his" does *not* c-command "John", it follows that in (4) "his" and "John" *may* have the same referential index.

It is not difficult to see that in this explanation no appeal is made to the externalist notion of reference. The explanation mentions two theoretical notions: c-command and referential index. The former is a purely syntactic relation between expressions, since it is defined exclusively in terms of the notion of tree diagram. It is important to realize that the latter notion is also purely syntactic,[3] although at first sight it might seem essentially semantic,[4] since two expressions with the same referential index are two expressions denoting the same object, and denotation (or reference) is a fundamental semantic relation, adopted by formal semantics as a primitive relation, intuitively clear if we assume the naive, i.e. realist-externalist, standpoint. The essential point to notice is that it *never* happens that the linguist, or the language user,

---

[3] On this point see also Cecchetto (1998).

[4] For instance P. Jacob writes:

> Referential dependency, however, seems to presuppose that the antecedent of an anaphoric expression possesses a reference. Otherwise, it is hard to see what the anaphoric expression would inherit its reference from. (Jacob 2002: 45)

This seems to imply that coreference is an essentially semantic relation.

needs to know whether two expressions denote the same object *in order* to establish (and therefore *before* (s)he has established) whether they may (or must) receive the same index; the conceptual priority is exactly reversed: *first* we establish, on the basis of purely syntactic procedures, whether the two expressions may receive the same index, and *then* (but this no longer concerns the linguist) we say that they denote the same object if they are coindexed.[5]

It seems to me that the preceding justification of Chomsky's objection (i) to (2) can be further strengthened. According to Chomsky,

> For the present, C-R [computational-representational] approaches provide the best-grounded and richest naturalistic account of basic aspects of language use. Within these theories, there is a fundamental concept that bears resemblance to the common-sense notion 'language': the *generative procedure* that forms *structural descriptions* (SDs), each a complex of phonetic, semantic, and structural properties. Call this procedure an *I-language* [...]. (Chomsky, 2000: 26)

An I-language consists of two parts: a computational procedure (invariant through the species) and a lexicon (with individual variation). The lexicon is a collection of items, each a complex of features, i.e. phonetic or semantic properties. The computational procedure selects items from the lexicon and forms expressions. An expression can be conceived as a pair <PHON, SEM> of symbolic representations: a phonetic representation and a semantic representation. PHON and SEM are the interfaces between the language faculty and the performance systems external to it: the articulatory-perceptual systems and the conceptual-intentional ones. These systems interpret information contained in PHON and SEM, in the sense that they take it as input.

In a computational-representational theory, therefore, input and output of a computation must be *representations*, complex *symbols* of a representational system. ["Representation" is used here in a technical sense; «there is nothing 'represented' in the sense of representative theories of ideas, for example.» (Chomsky, 2000, 173)] The assumption that a lexical item LI denotes an object o of the external world simply gives no contribution to the explanatory strategies of the theory, unless the fact that LI denotes o can be 'reflected' by some modification of the form of PHON or SEM; but in that case what is relevant is the (new) form of the representations, not the fact that LI denotes o.

From these remarks I conclude that the deep motivation for Chomsky's internalism is the same as for his principle of the autonomy of syntax: the requirement of articulating a computational-representational approach to language. In this sense internalism, even if not logically necessitated, is at the core of Chomsky's approach to language, and more generally to the mind.

---

[5] As a matter of fact, the relation between coindexing and covaluedness is more complex: two coindexed expressions are covalued, but not vice-versa: it may happen that two expressions with different indices are covalued; but information that they are covalued does not come 'from language' but from some other source.

1.1.1.2 Truth

If my analysis is correct, Chomsky's objections to the notion of realist-externalist reference can be extended in a natural way to the notion of realist-externalist truth, thereby completing the justification of his refusal of (2). As far as I know Chomsky has never stated explicit objections to the appeal to realist truth in the explanation of meaning; I think that the reason for this is not that he thinks that realist truth is essential in such an explanation, but that he thinks that realist truth is of no use for essentially the same reasons why reference is: the intrinsic tension between that notion and the requirement of articulating a computational-representational approach to language. For example, another typical problem concerning "internally-determined properties of linguistic expressions" is the fact that between the sentences.

(6) John is painting the house brown
(7) John is painting the exterior surface of the house, not the interior.

There is a relation of entailment. Now, given two sentences α and β such that α entails β, it *never* happens that the language user needs to know the truth-value of α or of β in any given situation *in order* to establish whether α entails β, or that the linguist needs to assume that the competent subject has that knowledge in order to explain the subject's competence. As in the case of reference, the conceptual priority is reversed: first we establish that α entails β, and then we say that, whenever α is true, β is true. Therefore the assumption that a sentence is (realistically) true or false gives no contribution to the explanatory strategies of the theory, unless the fact that it is can be 'reflected' by some modification of the form of PHON or SEM; but in that case what is relevant is the (new) form of the representations, not the fact that the sentence is true or false.

It is frequently assumed that, since entailment relations belong to the "internally-determined properties of linguistic expressions", and entailment is usually defined in terms of (realist) truth, linguistics is necessarily concerned with (realist) truth. In itself, this conclusion is a *non sequitur*: it would follow only if it were true that entailment *cannot* be defined but in terms of (realist) truth—which is not the case.[6]

1.1.2 Externalist Semantical Notions Generate Uninteresting Problems

1.1.2.1 Singular Terms

Let us consider Chomsky's second objection. An example of the "many new problems" the introduction of semantic values creates is the well-known case of 'London': in the sentence

[6] See below, Sect. 1.3.3. For alternative definitions see Prawitz (1985) and (2015), and references therefrom.

(8) London is so unhappy, ugly and polluted that it should be destroyed and rebuilt 100 miles away. (Chomsky, 2000: 37)

"London" seems to refer to something both concrete and abstract, animate and inanimate. The difficulty Chomsky raises can be made explicit in the form of the following argument:

(i) In model-theoretic semantics a sentence of the form "P(n)" is true if the individual denoted by "n" belongs to the set denoted by "P". Let us call this the "standard account" of the truth-conditions of the sentence.
(ii) Externalist semantics assumes that the individual denoted by "n" is an object of the external world, and that the set denoted by "P" is a set of objects of the external world.
(iii) Sometimes it happens that two sentences "P(n)" and "Q(n)" are intuitively true, where "P" and "Q" denote disjoint sets of objects of the external world. An example is "London is unhappy" and "London is polluted"; another, perhaps better, is "*War and Peace* has run into numerous editions" and "*War and Peace* weighs three pounds".
(iv) If we explain the intuitive truth of "P(n)" and "Q(n)" on the basis of the standard account, we obtain from the preceding steps that the object of the external world denoted by "n" belongs to two disjoint sets of objects.
(v) No object can belong to two disjoint sets of objects; therefore the individual of the external world denoted by "n" does not exist. For instance, London does not exist.
(vi) This conflicts with the obvious fact that London exists.[7]

So, the premises of the argument entail a contradiction: some of them must be dropped. There are in principle several alternatives: we could say (i) that the standard account is incorrect; or that (ii) externalist semantics is incorrect; or that (iii)

[7] «London is not a fiction.» (Chomsky 2000, 37) Elsewhere, speaking about a similar argument concerning the name "bank", he writes: «which is not to deny, of course, that there are banks, or that we are talking about something (or even some thing) if we discuss the fate of the Earth (or the Earth's fate) and conclude that it is grim […].» (Chomsky 2000, 181).

"unhappy" and "polluted" do not denote disjoint sets of objects; or that (iv) the argument itself (i.e. the derivation of the contradiction) is not valid.[8] Chomsky opts for the second alternative, as the following passage clearly suggests:

> the properties of such words as 'house', 'door', 'London', 'water' and so on do not indicate that people have contradictory or otherwise perplexing beliefs. There is no temptation to draw any such conclusion, if we drop the empirical assumption that words pick out things [...]. (Chomsky, 2000: 129)

In order to appreciate the significance of Chomsky's choice it is worthwhile to consider in some detail another possible reaction to the argument. One might observe that the sentences "London exists", "London does not exist" have in fact different meanings in different contexts: the former, asserted in a conversation, expresses the common-sense truth that there is a town named "London"; but also the latter might be true: asserted by a physicist it would expresses the scientific truth that towns such as London are not objects of physics. So—the objector might continue—there is no real contradiction in saying that London exists and does not exist: it exists for common sense, and it does not exist for physics; no contradiction has been derived.

Well, Chomsky would certainly agree on the remark that there is a difference, even a dramatic difference, between the points of view of science and of common sense. But he would stress that this remark would not yield a solution to the problem. The problem arises from the fact that, on the one hand, model-theoretic semantics is intended to be a science, while, on the other hand, externalist semantics assumes that "London" denotes an entity of the external world, and this assumption is true only for common sense, not for any science whatsoever; as a consequence externalist semantics cannot be a science as it is intended to be.

A possible escape from this impasse would be to give up the very idea that externalist semantics is a scientific enterprise, and to base it on the intuitive or common-sense notion of reference. I see at least two objections to this idea:

(i) If semantics is not a scientific enterprise, what is it? Its standpoint does not seem to be the one of common sense, although its primitive notions are intended to be notions belonging to common sense; it is not because of its demand of *systematicness*, which is extraneous to common sense. The only alternative I can see is that it is proposed as a *philosophical* enterprise. With this move philosophy sets itself up as an autonomous

[8] P. Casalegno has argued against the validity of Chomsky's argument. In Casalegno (1997), 359, he observes that London may be unhappy because of its inhabitants, or polluted because of the air above it, but when we say that it is unhappy we do not identify London with its inhabitants nor, when we say that it is polluted, do we identify it with the air above it; but this is precisely what Chomsky fallaciously does in order to derive the conclusion of his argument. It seems to me that Casalegno confounds here an epistemological remark with an ontological one; on the epistemological side, it is surely correct that we may assert that London is unhappy because of its inhabitants, without identifying London with its inhabitants; but the standard account of the truth-conditions of 'London is unhappy' calls for the ontological side of the question: London is unhappy simply if it belongs to a set of objects. Therefore, when we state the truth-conditions of 'London is unhappy', we cannot avoid making a choice about the sort of object London is: the inhabitants of a certain region, or the buildings of that region, or the air above it, and so on; and, as soon as we choose one alternative, the contradiction follows.

intellectual enterprise, different from both common sense and science, whose point of view is dangerously similar to that «mysterious point» from which it is possible to judge that «our scientific knowledge somehow 'mirrors' an independently existing reality».

(ii) According to Chomsky,

> From the natural-language and common-sense concepts of reference and the like, we can extract no relevant "relation between our words and things in the world". (Chomsky, 2000: 150)

hence there is no common-sense notion of reference to base a semantics on. P. Casalegno, among others, has argued, with partially convincing arguments, against the claim that according to common sense there is no idea of a stable relation between names and things of the external world (Casalegno, 2006: 406–408); but I do not think his argument hits the bullseye: Chomsky's point is that there is no common-sense notion of reference *to base a semantics on*, i.e. a common-sense notion capable to play the role of a basic notion of a theory aiming to be, if not scientific, certainly systematic; and the wealth of "London"-like examples speaks for itself.

1.1.2.2 Predicates and Compositionality

Let us see now an example of the difficulties arising when predicates are assigned an externalist denotation. Consider the sentences.[9]

(9) The house is green
(10) The ink is green
(11) The banana is green
(12) The stoplight is green.

One of the fundamental ideas of externalist semantics is that predicates denote functions or, equivalently, sets of entities of the external world. The problem with (9)–(12) is that for each sentence we must assign to the predicate "is green" a different set: the set of things which are green on the outside (i.e. whose exterior surface reflects green light), in the case of (9) and (11); the set of things which, when applied to paper and allowed to dry, will be green, in the case of (10); the set of things which emit green light, in the case of (12).

As in the case of "London", a possible answer to this difficulty is to say that "is green" simply denotes the set of green things, and to add that there are several ways in which a thing of the external world may be green. But consider the following sentence:

(13) John is white-haired, he drinks white coffee and white wine;

should we say that being grey, being brown and being yellow are ways of being white? Moreover, the sets of green, or white, things of the external world, so understood, are not entities any science may admit within its ontology; therefore either

[9] See Stainton (2008), 22, and the bibliography therefrom.

externalist semantics gives up the ambition of being a scientific explanation of meaning, or such a scientific explanation cannot assign to predicates denotations of that sort.[10]

Let us compare the sentences[11]

(14) Cars have wheels
(15) Unicycles have wheels.

(14) is intuitively true iff each (normal) car has wheels; (15) is *not* intuitively true iff each (normal) unicycle has wheels, but iff each (normal) unicycle has a wheel; as a consequence (14) entails "Jim's car has wheels", but (15) does not entail "Jim's unicycle has wheels", since (15) is true but "Jim's unicycle has wheels" is not. Such examples show, according to Chomsky, that

> the syntactic structures are not a projection of the semantics, and that the relation between 'the world of ideas' and the syntactic system is fairly intricate. (Chomsky, 1977: 31)

In particular, it seems that the syntactic structures do not respect a fundamental principle of Fregean, and more generally model-theoretic, semantics: the principle of compositionality:

> We cannot simply assign a meaning to the subject and a meaning to the predicate [...], and then combine the two. Rather, the meaning assigned to each phrase depends on the form of the phrase with which it is paired. (*Ibid.*)

The simple plurals "cars" and "unicycles" express universal quantifications; but which set is to be assigned to "have wheels" in order to respect compositionality? Not the set of things with at least two wheels, for then (15) would come out as false, while it is intuitively true. Not the set of things with at least one wheel, for then "Jim's unicycle has wheels" would come out as true, while it is intuitively false.

1.1.2.3 Truth

Is it possible to find some objection to truth analogous to Chomsky's second objection to reference? One can be trivially extracted from Chomsky's objection to externalist reference, as I have reconstructed it: consider the two sentences "*War and Peace* has run into numerous editions" and "*War and Peace* weighs three pounds": intuitively they are both true, but if we explain their truth-conditions according to the standard account they cannot be *both* true, since no object of the external world can belong to two disjoint sets of objects. Hence, either we renounce the standard account (in particular compositionality), or we renounce the intuitive notion of truth. From this point of view, Chomsky's objection to (classical) truth is that it conflicts with compositionality. Arguments for this conclusion can be found in Chomsky (1977);

---

[10] Another problem involves vague predicates: clearly {x|x is bald} is *not* a set, in spite of notation, since "is bald" has not a precise extension; and vagueness is a phenomenon concerning the vast majority of natural language predicates. Cf. Pietroski (2005a), 58–66.

[11] For a discussion of these examples see Chomsky (1977), 30 ff., and Pietroski (2003), Sect. 2.

according to them, the conflict arises from the fact that the intuitive notion of truth results from an interaction between different cognitive components. This point is further elaborated by P. Pietroski, who concludes:

> the fact that a sentence has a certain truth condition, as used in a certain conversational situation, is a massive *interaction effect*. (Pietroski, 2003: 239)

1.2 Anti-realist Objections

I shall explain in Chap. 2 what is meant by "semantical anti-realism"; for the time being let us take it as the idea that realist truth is not a key notion for a semantic theory, and let us see some reasons for this denial. I shall illustrate two of them: the first is related to Dummett's so-called anti-realist argument, the second is Prawitz's objection to Tarskian definition of logical consequence as truth preservation.

1.2.1 An Argument Related to Dummett's 'Anti-realist Argument'

Although Dummett's ideas are not tied to an anti-realist choice—and he has been particularly careful to underline this—it is a matter of fact that in a number of his writings he has sketched an anti-realist argument. Even if this is indisputable, it is equally clear that whenever we try to pin down the individual passages of the argument we find ourselves in an awkward position; and we realize almost right away that, in the best case, we can arrive at a 'rational reconstruction' of the argument. For this reason the argument needs to be analyzed extremely carefully. I have done this elsewhere (Usberti, 1995: Chap. 4); here I shall give a brief sketch of an improved version of it proposed by Prawitz (Prawitz, 1977), then I shall introduce an objection to it, in my opinion very convincing, and I shall show how the same conclusion can be reached without being exposed to the objection.

Here are the essential steps of Dummett-Prawitz's argument, accompanied by a brief explanation/justification:

(16) A theory of meaning is a theory of understanding.

EXPLANATION. With this slogan I summarize the idea that what a theory of meaning is expected to explain is, rather than what the meaning of a sentence α is in itself, what *knowledge* of the meaning of α by a subject consists in.

(17) Knowledge of meaning is partially implicit.

EXPLANATION. Knowledge of the meaning of α can be *explicit* or *implicit*. The former is verbalizable knowledge, that is the capacity to explain the meaning of α by using a synonymous sentence, or else by formulating rules for the use of α.

Knowledge of the meaning of α cannot consist entirely of explicit knowledge, because the possibility of verbalizing our knowledge presupposes that the language in which it is verbalized is already understood, i.e. that the meanings of its sentences are already known; so, to avoid an infinite regress, we must conclude that knowledge of meaning is at least partially implicit.

(18) An adequate theory $\mathcal{T}$ satisfies the a *condition of observability*, according to which the following conditional can be derived in $\mathcal{T}$:

$$\forall x \forall \alpha \exists b \exists S(KM(x,\alpha) \rightarrow (b \in \mathcal{B}_\alpha \wedge E(x,b,S))),$$

where
$\mathcal{B}_\alpha$ is the class of (behavioral) tokens of knowledge of the meaning of α,
KM(x,α) abbreviates "x knows the meaning of α",
E(x,b,S) abbreviates "x shows b in the situation S".

EXPLANATION. Wittgenstein's slogan «Meaning is use», which neo-verificationists are sympathetic with, would require that knowledge of meaning is a necessary *and sufficient* condition of knowledge of use. However,

> one can be sceptical about the existence of any such situation S or of any finite set of situations correlated with appropriate behaviour such that one can definitely decide whether a person knows the meaning of a sentence. [...] But since we demand empirical import of theoretical terms, an assumption about the possession of knowledge must have some observable consequences [...]. (Prawitz, 1977: 12)

(19) *Assumption*: Knowledge of the meaning of α consists in knowing the conditions under which α is classically true.

EXPLANATION. The essential characteristic of the classical notion of truth is its bivalence: every sentence is true or false, independently of our ability to recognize it as such.[12]

(20) Classical truth-conditions are such that sometimes it is impossible to recognize that they obtain.

EXPLANATION. There are, both in ordinary language and in many formalized languages, sentences which Dummett calls (with an unhappy choice of terminology) "undecidable", meaning by this sentences that we are not (now) able to judge true or false. Dummett distinguishes three fundamental categories of this type:

(a) Sentences about inaccessible regions of time–space (e.g. «Plato had a mole under his left armpit»);
(b) Sentences involving unlimited quantifications over infinite totalities (e.g. Goldbach's conjecture: «Every even number greater than 2 is the sum of two primes»);
(c) Conditional sentences in the subjunctive mood (e.g. «If John took this medicine he would die»).

[12] This point will be explained in Chap. 2.

(21) It is possible that a subject $\mathcal{S}$ knows the meaning of a sentence α without being in a position to show any of the forms of behavior that are considered signs of the knowledge of the meaning of α.

EXPLANATION. (21) follows from (19) and (20). For, let us consider an 'undecidable' sentence α and suppose that the subject $\mathcal{S}$ knows its meaning; on the basis of (19) he knows its (classical) truth-conditions; but by (20) it is impossible for him to recognize that such truth-conditions obtain; thus if α is true (false), $\mathcal{S}$ cannot show any of the forms of behavior that are considered signs of the knowledge of the fact that α is true (false); hence $\mathcal{S}$ cannot show any of the forms of behavior that are considered signs of the knowledge of the truth-conditions of α, i.e. of the knowledge of its meaning.

(22) Knowledge of the meaning of a sentence *does not* consist in knowing the conditions under which it is classically true.

EXPLANATION. (21) contradicts (18). Thus either (19) or (20) is false. Since (20) is obviously true, being the description of a fact, the content of the assumption (19) is false. With this the argument is concluded.

Chomsky has raised an important objection to the condition of observability as an adequacy condition for a theory of semantic competence:

> Imagine a person who knows English and suffers cerebral damage that does not affect the language centers at all but prevents their use in speech, comprehension, or let us suppose, even in thought. [...] Suppose that the effects of the injury recede and with no further experience or exposure the person recovers the original capacity to use the language. In the intervening period, he had no capacity to speak or understand English, even in thought, though the mental (ultimately physical) structures that underlie that capacity were undamaged. Did the person know English during the intervening period?
>
> [...] The cognitive property that concerns me holds of the person who possesses the mental structure, thus of the aphasic throughout, as we learn from the fact of his recovery. In this case, the fact of his recovery provides evidence that he had knowledge of English, though none of his behavior (even his thought) at the time provided any evidence for possession of this knowledge.
>
> Suppose that there is a second aphasic like the first, but because of some other and irrelevant problem (say, a circulatory disorder) he never recovers speech. Should we say in this case that the knowledge of English was lost? That would seem perverse. The first aphasic recovered because he had retained a certain mental (ultimately physical) state, a certain state of knowledge, namely, knowledge of English. His recovery provides evidence for the fact. One can imagine all sorts of evidence that might indicate that the aphasic who did not recover was in exactly the same (relevant) state; say, electrical activity of the brain or evidence from autopsy. The conclusion that the second aphasic retained his knowledge of English would have to be based on evidence, of course, but not necessarily evidence from behavior. To deny that the aphasic who did not recover had knowledge of his language would seem as odd a move as to deny that the one who did recover knew his language when he was unable to use this knowledge. (Chomsky, 1980: 51–52)

I agree with this objection; moreover, I agree with its core presupposition, which is mentalism, or realism about the mental: the gist of the objection is that the subject occupies a definite mental (ultimately physical) state, which is *not* a capacity, and whose existence can therefore be revealed by non-behavioural evidence. However, I hold that the conclusion of the argument can be obtained even if we give up condition (18) of adequacy, i.e. the condition that incorporates precisely the anti-mentalistic presupposition of Dummett-Prawitz's argument. Here is how.

Let us come back to assumption (19): what does it mean, exactly, knowing the conditions under which α is classically true? Which is the relation between knowing the classical truth-conditions of α and knowing its truth-value? The following answer seems to be uncontroversial:

(23) A subject $\mathcal{S}$ knows the classical truth-conditions of α iff, whenever $\mathcal{S}$ has access to all information specified by the truth-condition of α, $\mathcal{S}$ knows the truth-value of α.

EXPLANATION. We can quote a passage from Heim & Kratzer's handbook of semantics; commenting upon the derivation of the truth-conditions of "Ann smokes" they write:

> If you check the proof again, you will see that we end up with the truth-conditions of "Ann smokes" because the lexicon defines the extensions of predicates by specifying a condition. Had we defined the function denoted by "smoke" by displaying it in a table, for example, we would have obtained a mere truth-value. We didn't really have a choice, though, because displaying the function in a table would have required more world knowledge than we happen to have. We do not know of every existing individual whether or not (s)he smokes. And that's certainly not what we have to know in order to know the meaning of "smoke". (Heim & Kratzer, 1998: 20–21)

Now let us replace adequacy condition (18) with the following one, congenial to realist assumptions about the mental implicit in Chomsky's objection:

(24) An adequate theory $\mathcal{T}$ satisfies the a *condition of specifiability*, according to which the following conditional can be derived in $\mathcal{T}$:

$$\forall x \forall \alpha \exists \sigma (\mathrm{KM}(x,\alpha) \rightarrow (\sigma \in \Sigma_\alpha \wedge \mathrm{O}(x,\sigma))),$$

where
Σ_α is the class of mental states in which the meaning of α is known,
KM(x, α) abbreviates "x knows the meaning of α",
O(x, σ) abbreviates "x occupies the mental state σ".

EXPLANATION. Knowledge is in general a mental state; knowledge of the meaning of α is a mental state somehow corresponding to the meaning of α.

It seems to me that from (16)–(17), (19)–(20) and (23)–(24) conclusion (22) can be inferred. Let us see how.

(25) It is possible that a subject $\mathcal{S}$ knows the meaning of α without occupying the mental state corresponding to knowing the meaning of α.

EXPLANATION. Consider an 'undecidable' sentence of the form $\forall x\beta$ whose quantifier ranges over an infinite domain (Goldbach's Conjecture is a good example), and suppose that the subject $\mathcal{S}$ knows its classical truth-condition, i.e. the truth-condition of $\beta[\underline{a}/x]$, for every individual a of the domain D; by (23), whenever $\mathcal{S}$ knows the truth-value of $\beta[\underline{a}/x]$, for every $a \in D$, $\mathcal{S}$ knows the truth-value of $\forall x\beta$. But a mental state in which $\mathcal{S}$ knows the truth-value of $\beta[\underline{a}/x]$, for every $a \in D$, does not exist: there are infinitely many of these truth-values, and no mental state can contain an actually infinite set of pieces of information. *Actually* infinite: in order to know the truth-value of $\forall x\beta$ it is not sufficient to inspect the potentially infinite sequence $\beta[\underline{a_1}/x]$, $\beta[\underline{a_2}/x]$, …; it is necessary to realize that such sequence is 'complete', hence to conceive it as actually given.

(26) Conclusion (22) holds.

EXPLANATION. (25) contradicts (24). Thus either (19) or (23) is false. Since (23) is obviously true, the content of the assumption (19) is false. With this the argument is concluded.

Someone might object that a subject $\mathcal{S}$ knows the infinitely many facts relevant to the truth of $\forall x\beta$ if $\mathcal{S}$ knows a proof of the sentence or a counterexample to it, i.e. an individual a and a proof that $\beta[\underline{a}/x]$ is false. However, this sort of explanation is not available to the supporter of a realist conception of truth, for whom it is perfectly possible that a sentence of the form $\forall x\beta$ is true even if no proof of it objectively exists. Consider Goldbach's Conjecture and suppose that it is true and no proof of it objectively exists. What does this mean? If the reason for which an even number is the sum of two primes were the same for all even numbers, it is plausible that a proof of Goldbach's Conjecture would exist; it might be very difficult to find it, but it should exist, because a proof (realistically understood) simply puts into evidence the uniform reason why infinitely many facts subsist. By the same reasoning, a proof of Goldbach's Conjecture would exist if there were a finite number of reasons for its truth, or even if such reasons were infinite in number, but in some way epistemically surveyable through some law, or rule, or algorithm. If, on the contrary, no proof of Goldbach's Conjecture objectively exists, this must be because the reasons for its truth are not only infinite in number, but also so heterogeneous as to be unsurveyable. In that case, the mental state of a subject $\mathcal{S}$ who has at his disposal the infinitely many pieces of information that are relevant in order to recognize the truth of Goldbach's Conjecture can only be characterized as a state in which $\mathcal{S}$, as a matter of pure fact, has access to an infinite list and to the fact that that list is completed; but this would be an impossible mental state, if we conceive mental states as computational states. Concluding, the appeal to realist truth in the explanation of meaning has the effect of rendering *impossible* an explanation along computational-representational lines of knowledge of meaning.

1.2.2 Prawitz's Objection to Tarskian Definition of Logical Consequence

A second reason of suspicion on realist truth as the key notion of a theory of meaning can be extracted from the following remarks by Dag Prawitz about inference:

> It is said that with the help of valid inferences, we justify our beliefs and acquire knowledge. The modal character of a valid inference is essential here, and is commonly articulated by saying that a valid inference *guarantees* the truth of the conclusion, given the truth of the premisses. It is because of this guarantee that a belief in the truth of the conclusion becomes justified when it has been inferred by the use of a valid inference from premisses known to be true. But if the validity of an inference is equated with [truth preservation (*g.u.*)] [...], then in order to know that the inference is valid, we must *already* know, it seems, that the conclusion is true in case the premisses are true. After all, according to this analysis, the validity of the inference just means that the conclusion is true in case the premisses are, and that the same relation holds for all inferences of the same logical form as the given one. Hence, on this view, we cannot really say that we infer the truth of the conclusion by the use of a valid inference. It is, rather, the other way around: we can conclude that the inference is valid after having established for all inferences of the same form that the conclusion is true in all cases where the premisses are. (Prawitz, 2005: 675)

Let me try to make the argument explicit:

(27) For every subject $\mathcal{S}$, a valid inference $\mathcal{I}$ from α to β is intuitively useful for $\mathcal{S}$ if, and only if, for every time t, $\mathcal{S}$ knows at t that β only if there is a time t' such that (i) $t' < t$; (ii) $\mathcal{S}$ knows at t' that $\mathcal{I}$ is valid; (iii) $\mathcal{S}$ knows at t' that α; (iii) $\mathcal{S}$ does not know at t' that β.

(28) A good explanation of the validity of an inference $\mathcal{I}$ must account for its utility as well, i.e. explain how $\mathcal{I}$ can be at the same time valid and useful.

(29) Assume that we define an inference as valid if, and only if, it preserves (realist) truth.

(30) From (29) and (27) it follows that, for every subject $\mathcal{S}$, a truth-preserving inference $\mathcal{I}$ is useful for $\mathcal{S}$ if, and only if, for every time t, $\mathcal{S}$ knows at t that β only if there is a time t' such that (i) $t' < t$; (ii) $\mathcal{S}$ knows at t' that $\mathcal{I}$ is truth-preserving; (iii) $\mathcal{S}$ knows at t' that α; (iv) $\mathcal{S}$ does not know at t' that β.

(31) Condition (30) (ii) means, by Tarski's definition of truth-preserving inference, that, for every model $\mathcal{M}$, either $\mathcal{S}$ knows at t' that α is false in $\mathcal{M}$ or $\mathcal{S}$ knows at t' that β is true in $\mathcal{M}$.

(32) Then there cannot be a time t' satisfying the conditions (i)–(iv) specified in (30); for, if (ii) holds, then, by (31), for every model $\mathcal{M}$, either $\mathcal{S}$ knows at t' that α is false in $\mathcal{M}$ or knows at t' that β is true in $\mathcal{M}$; by (iii), $\mathcal{S}$ does not know at t' that α is false in $\mathcal{M}$; hence $\mathcal{S}$ knows at t' that β is true in $\mathcal{M}$, in contradiction with (iv).

(33) Hence, if we equate the validity of an inference with its being truth preserving, a valid inference cannot be useful; by (28), the definition of validity as truth preservation does not account for its utility.[13]

1.3 Prospects for an Internalist Semantics

The preceding arguments, coming from very different traditions, suggest (i) that the relations between language and external world cannot be studied with the methods of science, and (ii) that the relations between language and mind *can* be studied with the methods of science. If someone has been convinced, or at least intrigued, by those arguments, the next question is whether there is room for a scientific *semantics*.

If—following Morris (1938)—we define semantics as the study of the relations between language and world, evidently there is no room: the very notion of an internalist semantics is contradictory. But there is another, well established sense of "semantics": the study of the *meanings* of linguistic expressions. The possibility of a scientific semantics in this sense is not ruled out by Chomsky's arguments; it is therefore interesting to consider a bit more closely how it might be conceived.

First, Chomsky stresses that, although such a study deserves the name of semantics because of its connection with the intuitive notion of meaning, «[t]his work could be considered syntax in the technical sense; it deals with the properties and arrangements of the symbolic objects» (Chomsky, 2000: 174); it can therefore take the form of a computational-representational theory: «The study of C-R systems, including 'internalist semantics', appears to be, for now, the most promising form of naturalistic inquiry» (Chomsky, 2000: 47).

Second, if Chomsky's critiques of the realist-externalist notion of reference are accepted, two options are open: either to give up appealing to reference and truth in the explanation of linguistic competence, or to define notions that are internalistically acceptable and simultaneously capable of playing the *roles* of reference and truth. I shall call "thin semantics" the former alternative, "thick semantics" the latter.

1.3.1 Thin Semantics and Thick Semantics

Let us try to understand what is at stake. In "Questions of Form and Interpretation" Chomsky states the alternative in the following terms:

> Is there a system of language-independent representation in terms of which we can characterize speech-act potential, role in inference, and so on, and which is related in some interesting way to other structures of language? Is there a "broader-minded logic" of the sort to which Jespersen alluded that enters into linguistic theory as a level of representation,

[13] If this reconstruction is correct, a supporter of Tarski's definition of logical consequence would presumably question (31); but he would still owe us an explanation of what knowing that an inference is truth preserving amounts to.

> playing a significant role in the use of expressions in thought and communication; or, alternatively, should a theory of speech-acts, inference, or truth be developed directly in terms of the categories of syntax themselves? (Chomsky, 1977: 35)

The allusion to the role of inference deserves to be emphasized: as, according to Chomsky, «it seems that the computational procedure [out of which an I-language is composed, together with a lexicon] is too austere to use these resources [i.e., numbers and *inference*]» (Chomsky, 2000: 121),[14] it is natural to ask which system is capable to use inference. A plausible idea is that it is the system Chomsky calls "conceptual-intentional"; under this hypothesis, the definition of internalistically acceptable notions of truth and denotation would at the same time permit the formulation of a computational-representational theory of the *cognitive preconditions* of linguistic use and of some conjectures about the internal structure of the conceptual-intentional system.

Thin semantics, on the contrary, is characterized by the idea that, beyond the domain of current syntactic studies, one directly encounters the unsurveyable variety of linguistic uses, of which a scientific theory cannot be given, but only a description from the point of view of common sense.

The crucial question underlying the alternative between thin and thick semantics concerns therefore the limits of syntax in a broad sense, i.e. of computational explanation, and correspondingly the possibility of conceiving the conceptual-intentional system as a part of the language faculty in a broad sense, in the sense that its processes can be conceived as computations ruled by programs, to be studied by the same methods and within the same internalist framework as syntax: thin semantics answers No, thick semantics answers Yes, and considers the definition of internalist notions of object, denotation, truth, entailment, synonymy, and so on as a first, indispensable step in this direction.

1.3.2 Some (Bad) Reasons for a Thin Semantics

Most people inspired by Chomsky seem to opt for thin semantics.[15] Their reasons can be made explicit in the following points, which, taken together, make up a sort of argument for the thesis that a computational-representational theory of linguistic use is impossible.

(34) Once the theories of meaning related to externalist semantics have been discarded, the totality of the theories of meaning capable of explaining what knowledge of meaning, and therefore linguistic competence, consist in is exhausted by two great families: translation theories and use theories. The

---

[14] I guess Chomsky is alluding, for example, to the fact that bare output conditions do not rule out the sentence "Colourless green ideas sleep furiously", from which contradictions may be inferred.

[15] A conspicuous exception is Jackendoff. Chomsky seems open to both possibilities, even though in fact he has suggested some reasons for a thick semantics, as we will see.

former equate the meaning of an expression E of a language $\mathcal{L}$ with the translation E′ of E into a system $\mathcal{L}'$ which can be identified with a language (mentalese, or the language of logical forms, of discourse representations, and so on). The latter equate it with the totality of possible uses of E. (Different theories will differ in characterizing the class of *possible* uses, but this need not concern us here.)

(35) The typical difficulty of translation theories is clearly stated by N. Hornstein:

> it is clear why translation theories are insufficient as full theories of meaning. They postulate an internal language – mentalese – into which natural languages are translated. They do not, however, explain what it is to understand a concept in mentalese – what subabilities and knowledge are required to explain the complex ability of competent speakers to appropriately use and comprehend an infinite number of novel sentences of mentalese. In effect, a translation theory of this sort presupposes what we want to have explained. [...] The demand for explanation is simply pushed one step back; instead of asking what it means to understand a word or a sentence, we ask what it means to grasp a concept. No progress has been made. (Hornstein, 1984: 123–4)

(36) Use theories, on the contrary, are particularly apt to explain knowledge of meaning: to know the meaning of E amounts to having a capacity—the capacity to use it appropriately or correctly. But Chomsky has raised an important objection to this idea, contained for instance in the long passage from Chomsky (1980) quoted in Sect. 1.2.1.

(37) Use theories have a fundamental defect: of the totality of uses of an expression it is not possible to give a computational-representational theory, i.e., according to Chomsky—a scientific theory; of use it is only possible to give a *description*, which will necessarily employ the conceptual categories of common sense.

(38) Moreover, such a description is not only unavoidable, but also welcome if we want in some way to reconcile a scientific syntax with common sense.[16]

(39) As a consequence, thin semantics proposes to integrate a computational-representational theory of syntax with pragmatics conceived as a description of use which adopts the categories of common sense, in particular its naive realism.

It may be useful to state the main reasons that can be adduced in favour of (37).

(40) Criticizing the Neo-Fregean strategy of solution of many problems Fregean semantics is confronted with consisting in relativizing truth to indices, N. Hornstein remarks that

> In the *Philosophical Investigations*, Wittgenstein tried to suggest that the idealization concerning language that logicians and semantical theorists make in order to develop their theories begs most of the important issues. He did this by describing the variety of language, its myriad uses, and, in particular, the crucial role that contextual parameters play in determining the "meaning" of a linguistic expression like a sentence. He noted that these contextual parameters do not themselves appear to be liable to precontextual elaboration. At the very least, Wittgenstein can be read as making the accurate observation that, as yet, there is no general theory about what must or will count as a significant or relevant context that is itself

[16] On this and the following point see McGilvray (1998).

> context insensitive. But without such account we have no general theory of interpretation or linguistic productivity. Given this view, the real trouble with semantical theories is that they *presuppose* that there is something fixed that is nonlinguistic and that acts as the backdrop against which rules are constructed. (Hornstein, 1984: 144–145)

Hornstein uses this Wittgensteinian idea to criticize externalist semantics; however, he holds that his criticism has a wider scope, in the sense that it may be viewed as an argument against the very possibility of a systematic approach to the theory of meaning:

> The problem lies in postulating a fixed background. Whether it is a mentalistic language of thought, or an external world, or fixed confirmation or verification procedures is secondary. (Hornstein, 1984: 146)

He concludes with the conjecture that «There may not be a general account of what it is to know the meaning of an arbitrary sentence» (Hornstein, 1984: 148), since the presupposition that «all sentences have a certain property in common that is crucial to their meaning and that a competent native speaker knows» (ibid.) may be false.

(41) Linguistic use is intentional: its description requires intentional terminology; and a computational-representational theory of intentional relations is impossible. «Naturalistic inquiry will always fall short of intentionality», Chomsky writes (Chomsky, 2000: 45).
(42) Linguistic use is ruled by norms: it is subject to criticism by 'experts', it is appropriate or not; and of normative notions it is not possible to give a computational-representational theory.
(43) Linguistic use is a form of action; human action is driven by goals and intentions, is free, is «influenced but not determined by internal state, appropriate to situations but not caused by them» (Chomsky, 2000: 17). And science has nothing to say about free actions:

> Rousseau, developing Cartesian ideas in an original way, remarked that "Nature commands every animal, and the beast obeys. Man feels the same impetus, but he realizes that he is free to acquiesce or resist; and it is above all in the consciousness of this freedom that the spirituality of his soul is shown." And "the power of willing, or rather of choosing," as well as "the sentiment of this power," lies beyond the bounds of physical explanation, he believed. (Chomsky, 1980: 7–8)

I have made these point explicit to make clear why, even if (34)–(43) are taken together as an argument, its conclusion cannot be (39): it cannot be because, if we accept (40)–(43), thin semantics is *not* the only possibility left open. Thick semantics, as I have characterized it, is a theory not of use but of the *cognitive preconditions* of use, i.e. of the kind of cognitive structures and information a subject must have in order to be capable of using correctly an expression, where "correctly" simply means in accord with their meaning. This is sufficient to answer (36): when meaning is explained in terms of cognitive preconditions of use, it is natural to conceive knowledge of meaning as a particular kind of mental state, to be explained in terms of a computational-representational theory. On the other hand, such a theory should not be conceived as simply a part of the theory of an I-language; as Chomsky writes,

> The I-language is a (narrowly described) property of the brain, a relatively stable element of transitory states of the language faculty. Each linguistic expression (SD) generated by the I-language includes instructions for performance systems in which the I-language is embedded. It is only by virtue of its integration into such performance systems that this brain state qualifies as a language. Some other organism might, in principle, have the same I-language (brain state) as Peter, but embedded in performance systems that use it for locomotion. (Chomsky, 2000: 27)

It is only by virtue of such an integration that the outputs of the I-language, the SEMs, qualify as *meanings*. In other terms, SEMs are meanings only for a system that is so structured as to use them for representing thoughts. As a consequence, a theory of meaning cannot avoid reference to a *selected* class of possible uses; reference to the use of names to refer to objects and of sentences to represent thoughts is just intended to functionally select conceptual-intentional systems as the ones to which an I-language is integrated.

Let me try to explain why, even granted that (40)–(43) are good reasons against a theory of use, they are not against a theory of the cognitive preconditions of use.

Ad (40).

One of the central theses of this book is that, if one gives up the realist assumption of an external world, a background for a systematic theory of meaning can be *constructed* (rather than presupposed). I will argue for this simply by showing how to construct the theoretical background; but it should be clear from the outset that Hornstein's argument is not obviously extendible to approaches different from the realist one, since his remark that the set of indices and rules for their use are themselves contextually sensitive applies to the realist semantics, but not obviously to others; for example, justification procedures might be context-sensitive in a way substantially different from truth-conditions.

Ad (41).

Chomsky's reasons for this opinion are numerous and, in my opinion, convincing. But their common presupposition—which Chomsky shares with the vast majority of contemporary philosophers—is that intentionality is a *relation* of *aboutness* between mental states and things or states of affairs of the external world.[17] However, this is only *one* notion of intentionality, the externalist one. There is another one, which I will call "internalist" or "I-intentionality": intentionality as the *property* of representations of having a (n immanent) *content*.[18] For example, Chomsky writes:

> The representations are postulated mental entities, to be understood in the manner of a mental image of a rotating cube, whether it is the consequence of tachistoscopic presentations of

---

[17] Here is, for example, how intentionality is characterized by J. McGilvray, explicitly echoing Chomsky:

> Intentional properties arise when an inner state or process is treated as a representation of something in the world, or as having referential or alethic properties. That 'John' is used by Mary on an occasion to refer to John is an intentional property of 'John' on that occasion. (McGilvray 1998: 231–232).

[18] Arguably this is Husserl's original notion of intentionality.

a real rotating cube, or stimulation of the retina in some other way; or imagined, for that matter. (Chomsky, 2000: 160)

What does he mean with “a mental image *of* a rotating cube”? Not an image standing in some relation with a rotating cube, of course, since it might not be the consequence of the presence of a real rotating cube. The only possibility I see is that he means a mental representation that is visually *experienced as* a rotating cube and this is precisely what I mean by saying that that image has a rotating cube as its immanent *content*. In this sense the study of visual perception, which—as Chomsky remarks—«keeps to visual experience» (Chomsky, 2000: 195, fn. 1) and is therefore completely internalist, does concern intentional content, if “intentionality” and “content” are internalistically understood:

The representational content is manifest to us subjectively, and is clearly detectable in discriminatory behavior. Representations with different contents are needed to explain the discriminative capacity, and this provides the top-down motivation for positing them. (Segal, 1989: 210)

Chomsky would refuse this reference to content (Chomsky, 2000: 23); but—I suspect—only because, again, he uses “content” in the technical sense of so much philosophical discourse, which is essentially externalist.[19] In any case, I will introduce a notion of I-content and will try to define it in the next chapters. I do not see any reason to renounce to such an internalist construal of intentionality and content, which is both legitimate and present within the intuitive notions. On the positive side, when intentionality is conceived in this way computational-representational theories of it *are* possible: in the case of perception, they are the actual work of computational theories of perception, whose object is just the relation between perceptual inputs and perceptual experiences as outputs; in the case of language, some suggestions will be given in Chap. 4.

It is usually held that the Twin-Earth thought experiments show that the content of some mental states has to be externistically individuated. Timothy Williamson, for instance, argues that the sentence

(44) One believes that there are tigers

expresses a broad condition[20] by using a Twin-Earth scenario (Williamson, 2000: 53). Here is the argument:

(45) Consider a world w like the actual world a except that instead of tigers there are schmigers, creatures similar in appearence to tigers but quite different in ancestry and internal constitution.

[19] See for instance Chomsky (2000): 159.

[20] A condition, in his terminology, is what is specified by a ‘that’-clause, and either obtains or fails to obtain in each case, i.e. in a possible total state of a system consisting of an agent at a time paired with an external environment; it is broad iff it is not narrow, narrow iff for all cases α and β, if α is internally like β, then C obtains in α iff C obtains in β (where α is internally like β iff the total internal physical state of the agent in α is exactly the same as the total internal physical state of the agent in β). A state S is narrow iff the condition that one is in S is narrow, broad otherwise. Externalism is the thesis that some mental states are broad.

(46) Assume that in a John believes truly that there are tigers, that Twin-John—John's *doppelgänger* in w—is in exactly the same internal physical state as John, and that both John and twin-John express their beliefs by the sentence "There are tigers".
(47) Twin-John believes truly.
(48) Since Twin-John's belief is true iff there are schmigers, he believes truly that there are schmigers, *not* that there are tigers.
(49) John and Twin-John are in exactly the same internal physical state, but in a the condition C specified by (44) obtains, while in w it does not obtain; therefore C is broad.

The argument seems to me unconvincing from several points of view. First, why should (47) be accepted? We might alternatively describe Twin-John's situation by saying that he believes falsely that there are tigers; moreover, while for (47) no justification is given by Williamson, the alternative description might be justified by the fact that Twin-John would assent to the sentence "There are tigers" and that that sentence is false in w. Second—and more to the point—the argument presupposes, at step (48), that the content of mental states has to be individuated in terms of their bivalent truth-conditions[21]; bivalence is the mark of realism,[22] and realism explains truth and denotation in terms of relations between linguistic entities and objects or facts of the external world; externalism about content[23] analogously explains the content of mental states in terms of a relation between the subject and an external environment. So the presupposition of step (48) is not legitimate in the context of an argument that should prove the truth of externalism about content. It seems therefore more accurate to draw from the Twin-Earth thought experiments the disjunctive conclusion that «Either mental contents are not individuated individualistically, or they are not individuated in terms of their [bivalent] truth conditions» (Boghossian, 1994: 34). In conclusion, the Twin-Earth experiments offer no argument against the possibility of individuating mental contents in terms of internalistically acceptable notions.

The adoption of an internalist notion of intentionality requires us to abandon the common sense point of view about the contents of mental states, since common sense is naturally realist-externalist. Since, as we have seen, there are independent reasons for abandoning the point of view of common sense, this necessity should be welcome, or at least not be seen as a difficulty.

Ad (42).

Even granted that linguistic use is governed by public norms, or at least by explicit norms speakers obey, it does not follow that cognitive preconditions of use must be conceived as norms of this kind: they may be seen as computational programs implemented by the mind, or as general conditions imposed onto such programs. I will come back to this point in Chap. 3.

[21] Cf. fn. 20: a condition either obtains or fails to obtain in each case.

[22] See Chap. 2.

[23] In the terminology introduced in fn. 20, externalism about content may be taken to be the thesis that some condition is broad.

Ad (43).

While use is a form of action, cognitive preconditions of use are not: they may be states of cognitive systems, of which a computational-representational theory can in principle be given.

1.3.3 Some (Good) Reasons for a Thick Semantics

The distinction between use and cognitive preconditions of use should have cleared the field of some possible misunderstandings about thick semantics: that it is normative, that it is concerned with phenomena involving freedom; and the characterization of it as the study of relations between language and (other components of) mind, as opposed to relations between language and external reality, should have made clear that it is to be conceived as «syntax in the technical sense» that «it deals with the properties and arrangements of the symbolic objects» (Chomsky, 2000: 174). The alternative between thin and thick semantics, therefore, does not concern the thesis that «the basic split is between syntax and pragmatics» (McGilvray, 1998: 274, fn. 35), but the question: how far can syntactical explanation be pushed forward before giving way to pragmatic description?

The question "Where are the boundaries of syntax?" seems to have no a priori answer; as a consequence, the methodologically best choice is to try to extend its limits as far as possible, since what is at stake is the possibility of simultaneously extending the limits of a computational-representational explanation, and of formulating conjectures about the structure of the cognitive systems connected to the language faculty.

The idea that the common-sense point of view is in some domain indispensable is in conflict with this methodological principle, since it implies that of that cognitive domain it is impossible to give a *theory*, but only a common-sense description. So, for instance, P. Pietroski holds that

> Referring to water is relevantly like seeing water. It can't be done without *some* kind of contact with at least some H_2O. (Pietroski, 2005b: 282)

But this is not the opinion of several people who try to develop a computational-representational theory of vision. Or, to make another example, Hornstein's conjecture quoted in (40) conflicts with the plausible idea of formulating hypotheses about the internal structure of the conceptual-intentional system, whereas the fact that in Hauser et al. (2002) the conceptual-intentional system is included into the Faculty of Language Broad[24] entails that it is possible—and interesting—to develop

[24] The faculty of language in the broad sense, including a sensory-motor system, a conceptual-intentional system, and the computational mechanisms for recursion; it is hypothesized to be distinct from the faculty of language in the narrow sense, only including recursion, which provides the capacity to generate an infinite range of expressions from a finite set of elements, and is the only uniquely human component of the faculty of language.

a computational-representational theory of it. Or, to make a third example, consider how J. McGilvray proposes to look at inference:

> Generally speaking, use values measure *appropriate* use, given the task to which a/the SEM is put. With this in mind, look at correctness of inference. I say Harriet is chasing George. If so, she is following him, and both following and chasing intentionally. I say Mort is painting his house; if so, he is painting the outside. These are 'analytic truths' (indeed, a priori analytic truths). Similarly, I say Mort has started a book. He has started to read a book, or he has started to write one. Anyone who has the relevant vocabulary items at all realizes this, and everyone relies on this kind of detailed (but implicit) knowledge of structure and texture while using these and thousands of other lexical items. Surely, then, there is a strong sense in which SEMs guide their use. It is guidance alone: they do not control it. (McGilvray, 1998: 277)

It seems to me that to reduce inference to a question of appropriate use of lexical items it to renounce from the outset to the possibility of giving a *theory* of the domain of inferential relations. Although narrow syntax is "too austere" to use inference, it does not follow that inference can only be pragmatically described, from the point of view of common sense. If the systems capable to use inference are, as I have conjectured above, the conceptual-intentional ones, the possibility is open that such systems, besides being performance systems in relation to narrow syntax, are also competence systems in relation to other components of the mind (perception, memory, etc.), and that they have an internal structure predating the emergence of language. This is suggested, for example, by R.C. Gallistel:

> The generalization from a jay's own behavior to the likely behavior of others seems to me of particular interest for the light it sheds on the relation between thought and natural human languages. I would suggest that the jay's generalization is most readily understood if one assumes that in the thought of a bird, as in, I believe, every language, the symbol for an action is independent of the agent and the direct and indirect objects (<I>**take**<your><food>, <you>**take**<my><food>). The inferences that birds draw from their own behavior to the possible behavior of others suggest to me that this way of representing actions predates by hundreds of millions of years the emergence of natural human languages. (Gallistel, 2011: 259)

The need of internalist notions of denotation and truth has to be justified. In other words, the question is whether denotation and truth have an irreducible role within an internalist framework. Here is an argument for an affirmative answer in the case of denotation.

There is a well known distinction between two kinds of ambiguity of a word: accidental, also called *homonymy*, and 'systematic', also called *polysemy*; "trunk" (case/proboscis) is an example of the first kind, "book" (work/object) of the second, where the systematic nature of the ambiguity has to do with the abstract/concrete opposition. The interesting fact is that the distinction is linguistically relevant, as Chomsky observed; for instance,

> We can say that
>
> 1. The bank burned down and then it moved across the street;
>
> 2. The bank, which had raised the interest rate, was destroyed by fire;

> [...] Referential dependence is preserved across the abstract/concrete divide. Thus (1) means that the building burned down and then the institution moved; similarly (2) [...]. But we cannot say that:
>
> 4. The bank burned down and then it eroded; or
>
> 5. The bank, which had raised the interest rate, was eroding fast;
>
> [...]. Sentence (4) does not mean that the savings bank burned down and then the river bank eroded.
>
> [...] In the case of "bank", the natural conclusion is that there are two LIs that happen to share the same I-sound (homonymy), and that one of them, "savings bank", is polysemous, like "book" [...]. (Chomsky, 2000: 180)

Why are 1. and 2. grammatical and 4. and 5. are not? The reason seems to be more or less the following: on the one hand syntax requires that, in all these sentences, the noun phrase and the pronoun are coindexed; on the other hand, in 4. and 5. "the bank" cannot be coindexed with the pronoun, while in 1. and 2. it can. At this point the question is: why is coindexing possible in 1. and 2., and not in 4. and 5.? Commenting on similar examples Chomsky (1977: 69) writes: «It seems that a general principle of syntax-semantics interaction is involved, and again it seems plausible to attribute it to universal grammar». If Chomsky's suggestion is correct, it seems plausible that such a principle should make reference to an internalist notion of denotation.[25]

Concerning truth, an important reason for defining an internalist notion of truth can be extracted from the following objection to internalism raised by Fodor:

> It looks like the inferential relations that Chomsky has in mind for semantic theories to explain are species of *entailment* [...]. But an entailment is a kind of necessary *truth*, and truth is a world/mind relation. [...] Chomsky's Internalism requires a notion of entailment that is somehow freed from such notions as reference and truth; but I don't think there is one. (Fodor, 2000: 4)

I do not agree on the second point made by Fodor, that truth is a world/mind relation; realist-externalist truth of course is, but there are other, epistemic, notions of truth, in terms of which entailement can be defined, and some among them are internalistically acceptable, as I shall argue in Chaps. 2, 6 and 9; hence «a notion of entailment that is somehow freed from such notions as [realist] reference and truth» does exist.[26]

But I agree on the first point made by Fodor: a definition of entailment requires *some* notion of truth. The phenomena belonging to Jespersen's 'broader-minded logic' were characterized by Chomsky (1977) in the following terms:

> The relations between "murder" and "assassinate", or "uncle" and "male", or "cheerful" and "unhappy", ought to be expressible in terms that are not drawn from the theory of syntactic forms and categories or the world of fact and belief. [...] Considerations of modality do not suffice to make the relevant distinctions. "I found a proof of the parallel postulate" and "I found a Euclidean triangle with angles adding up to 200°" have the same truth

---

[25] I shall elaborate on this point in Chap. 4.

[26] See also Chap. 4, fn. 78.

> value in all possible worlds, but are utterly different in meaning and correspondingly make different contributions to the truth value of sentences in which they are embedded (e.g., "John believes that...") Furthermore, it seems reasonable to suppose that the factors that enter into determining the necessary falsehood of these expressions are different in kind from those that assign the same property to "I found a female uncle". (Chomsky, 1977: 35)

Later, in Chomsky (2000) the examples multiply, including such cases as the entailment relationship between (6) and (7). However, (6) and (7) are atomic sentences, while intuitive data to be explained include also entailment relations between such sentences as

(50) John is a boy

and

(51) It is false that John is not a boy,

between (50) and

(52) John is boy or Mary is a girl,

and so on and so forth. On the one hand, knowledge of these entailments (and of a lot of others) seems to belong to linguistic competence, because it involves knowledge of the meaning of the logical constants; on the other, it seems not to belong to the faculty of language in a narrow sense, as we have seen above (Sect. 1.2.1), because it involves inference; it seems therefore plausible that it is a component of the faculty of language in a broad sense. It is at this point that the choice of a thick semantics appears methodologically promising. In order to characterize the entailment relations exemplified by (50)–(52) it is not sufficient to make reference to the set of lexical features representing the meaning of "John is a boy" and "Mary is a girl"; it is necessary to give a general characterization of the meaning of the logical constants, and this cannot be done without defining a computable, internalistically acceptable, notion of truth for sentences of arbitrary logical complexity; in return, such a general characterization will permit to formulate hypotheses about the inner structure of the conceptual-intentional system(s), in particular about the system of knowledge underlying our inferential capacities.

An analogous point could be made about synonymy relations. Some intuitions involving synonymy seem to be a significant part of our semantic competence. For example, if E is, for a subject S, synonymous with E', then the thought expressed by a sentence α containing E will be the same, for S, as the thought expressed by the sentence obtained from α by replacing E with E'. Again, in order to characterize the synonymy relation between expressions of any kind (including sentences), the definition of a computable, internalistically acceptable, notion of truth seems indispensable.

1.3.4 The Scope of a Thick Semantics

Let me sum up the preceding discussion. Chomskyan internalism follows from the methodological principle according to which the sole criterion of intelligibility of the natural world is represented by the methods of the empirical sciences, and from the factual remark, quoted above, that «For the present, the best-grounded naturalistic theories of language and its use are C[omputational]-R[epresentational] theories», for a computational-representational psychological theory never invokes, in explaining relevant data, properties of things 'out there'. An internalist semantics, conceived as a computational-representational theory of semantic competence, should be thick, since linguistically relevant aspects of semantic competence seem to require explanations in terms of some notion of denotation and truth; if it is conceived as a theory of the cognitive preconditions of use it can escape the difficulties to which theories of use are exposed, and it may permit formulating conjectures about the structure of the conceptual-intentional systems.

The characteristic claim of the advocate of a thick semantics is that it is *never* the time to reintroduce the conceptual framework of common sense into a psychological explanation: not in linguistics, but not even in the various theories of perception, of memory, of thought, and so on. Of course there is a limit to this methodological claim: when a scientific (i.e. computational-representational) explanation is *impossible*, a common-sense description is unavoidable. On the other hand, the central tenet of the anti-realist about the external world—as I propose to conceive anti-realism in this book—is that, if an anti-realist theory of meaning is adopted, then a computational-representational *explanation* of the central notions of a thick semantics, and of intentionality, becomes possible.

I have spoken of *some* notion of denotation and truth; by this I meant theoretical (as opposed to intuitive) notions capable of playing the essential roles of the intuitive notions of truth and denotation. More specifically, notions—let me call them "$\mathcal{C}$-truth" and "$\mathcal{C}$-denotation" ("$\mathcal{C}$" for computational)—that meet the following conditions: (i) are capable of playing the role of key notions of a theory of meaning; (ii) are internalistically acceptable; (iii) play a substantial role in the explanation of our linguistic competence; and (iv) account for the essential aspects of the corresponding intuitive notions.

Condition (i) requires that an internalist semantics be based on a theory of meaning; as explained above (in *Ad (41)*), this is essential if we want (as I said I want) to develop a theory of content and, in perspective, of propositional attitudes and in general of I-intentionality. Condition (ii) is an obvious consequence of the adoption of an internalist standpoint, and condition (iii) an equally obvious consequence of the acceptance of Chomsky's criticism of externalist reference and truth. As for condition (iv), one might even wonder whether it is legitimate to impose it on a definition. The notions of denotation and of referential object are theoretical notions, intended to serve the needs of a science, linguistics; why should they resemble the corresponding intuitive notions? Elementary particles, for example, are a kind of objects postulated by physics, but we do not require that they resemble in any way common-sense physical objects. I agree with this remark, but from it it does not follow that condition (iv) is not legitimate; the condition does not require that the theoretical notions *resemble* the intuitive ones, but that they *account for* them. One reason why they *should* account for them is that there is an important difference between physics, for example, and psychology, of which linguistics is conceived by Chomsky as a part. Psychology studies the mind, and the mind has representations, in particular representations of objects and of thoughts in the non-relational sense explained above, i.e. representations and beliefs having an immanent content; the theoretical notion of content to be defined is intended to contribute to explain the intuitive notion of immanent content, once it is conceived—internistically—as what the mind experiences. Moreover, the mind has representations and beliefs concerning other subjects' representations and beliefs: if it is possible to characterize them in computational terms, the domain of broad syntax, hence of computational-representational explanation, will be considerably enlarged.

1.3.5 Is Model-Theoretic Semantics Compatible with Internalism?

Chomsky explicitly admits the legitimacy of a thick semantics:

> [I]t could be that a technical notion of *reference* should be introduced in the study of the syntax of mental representations, much as relations among phonetic features are introduced into phonology. (Chomsky, 2000: 202, fn. 6)

> Within internalist semantics, there are explanatory theories of considerable interest that are developed in terms of a relation R (read "refer") that is postulated to hold between linguistic expressions and something else, entities drawn from some stipulated domain (perhaps semantic values). (Chomsky, 2000: 38–9)[27]

On the other hand, he is inclined to think that much work in model -theoretic semantics, and therefore in natural language semantics, can be reinterpreted in such a way as to be compatible with an internalist framework.[28] I see at least three difficulties.

First, a fundamental idea of model-theoretic semantics is that names denote *individuals* belonging to some domain, but examples of the kind of "London" seem to clash directly with this idea, as we have seen: an individual cannot belong to two disjoint sets.

Second, another fundamental idea of model-theoretic semantics is that predicates denote sets of individuals, and that P(a) is true if the individual denoted by a belongs to the set denoted by P; but this analysis of predication in terms of the set theoretic relation of membership is not applicable when the denotations of names are mental entities.

---

[27] See also Chomsky (2000), 204, fn. 11; Chomsky (1981), 324.

[28] Cf. for instance Chomsky (1981), 344, fn. 3; Chomsky (2000), 204, fn. 11; Chomsky (2012), 207.

Third, classical truth-conditions are such that sometimes it is impossible to recognize that they obtain (see (20) above), and, as we have seen above, in such cases knowledge of them does not meet the specifiability condition (24).

If these objections are accepted, model-theoretic semantics is *not* compatible with an internalist framework, and the basic notions of object, predication and truth must be defined in some essentially different way; as we will see in the next chapters, some valuable suggestions in this sense come from intuitionism.

1.4 Conclusion: The Rationale for a Convergence

Two main reasons for a convergence between Chomskyan internalism and meaning-theoretical anti-realism emerge from the preceding discussion.

(i) Dummett-Prawitz's anti-realist argument, once purged of its behaviourist presuppositions, brings to light the impossibility to characterize knowledge of the externalist meaning of α as a mental state; in this sense it highlights the impossibility of a computational-representational theory of knowledge of the externalist meaning of α stressed by Chomsky.

(ii) Chomsky requires a theory of competence meeting the standards of scientific explanation. If we agree, and we hold that our inferential capacities are at least partially connected to our semantic competence of the logical constants, we hold as well that it is necessary to develop a scientific approach to our semantic competence of the logical constants, hence a computational-representational theory of our inferential competence. A central aspect of this competence is that we know that α entails β *before* knowing that β is true; therefore we need a definition of entailment capable to account for this, i.e. such that our belief that β is *justified* by our (justified) beliefs that α and that α entails β. It is by no means necessary to be anti-realist to acknowledge this need; for example, J. Etchemendy writes[29]:

> A logically valid argument must, at the very least, be capable of justifying its conclusion. It must be possible to come to know that the conclusion is true on the basis of knowledge that the argument is valid and that its premises are true. This is a feature of logically valid arguments that even those most sceptical of modal notions recognize as essential. Now, if we equate logical validity with mere truth preservation, as suggested in the last section, we obviously miss the essential characteristic of validity. For in general, it will be impossible to know both that an argument is "valid" (in this sense) and that its premises are true, without antecendently knowing that the conclusion is true. (Etchemendy, 1990: 93)

In conclusion, a computational-representational theory of our semantic competence of the logical constants requires a theory of inference that does not reverse the intuitive priorities.

---

[29] As a matter of fact, Prawitz explicitly recognizes (Prawitz 2005: fn. 3) Etchemendy's paternity of his own remarks about inference quoted at the beginning of Sect. 1.2.2.

To sum up the preceding discussions, we can schematically characterize Chomskyan methodological internalism as a form of realism about the mental plus the requirement of computational-representational explanations in psychology; anti-realism about external world as the refusal of externalist accounts of meaning. I have argued for a convergence of the two views towards a computational-representational explanation of meaning, I-intentionality, and so on. An interesting question, at this point, is whether a philosophical position merging anti-realism about the external world and realism about the mental is exemplified. I hold that mathematical intuitionism, with its view of logic as a *logique du savoir* (as opposed to a *logique de l'être*) is such an example. Since the intuitionistic theory of the meaning of the logical constants is often likened to the neo-verificationist one (which is by no means realist about the mental), I shall devote the next chapter to disentangle the two.

References

Boghossian, P. (1994). The transparency of mental content. *Philosophical Perspectives, 8*, 33–50.

Casalegno, P. (1997). *Filosofia del linguaggio*. La Nuova Italia Scientifica.

Casalegno, P. (2006). Chomsky sul riferimento. In E. Ballo, M. Franchella (Eds.), *Logic and philosophy in Italy* (pp. 397–410). Polimetrica.

Cecchetto, C. (1998). La semantica fra internismo ed esternismo. Qualche riflessione sulla recente produzione filosofica di Chomsky. *Lingua e Stile, XXXIII*(1), 23–49.

Chomsky, N. (1977). *Essays on form and interpretation*. Elsevier North-Holland.

Chomsky, N. (1980). *Rules and representations*. Columbia University Press.

Chomsky, N. (1981). *Lectures on government and binding*. Foris Publications.

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge University Press.

Chomsky, N. (2012). *The science of language. Interviews with James McGilvray*. Cambridge University Press.

Etchemendy, J. (1990). *The concept of logical consequence*. Harvard University Press.

Fodor, J. (2000). *The mind doesn't work that way*. The MIT Press.

Friedman, M. (1993). Remarks on the history of science and the history of philosophy. In P. Horwich (Ed.), *World changes: Thomas Kuhn and the nature of science* (pp. 37–54). MIT Press.

Gallistel, C. R. (2011). Prelinguistic thought. *Language Learning and Development, 7*, 253–262.

Hauser, M., Chomsky, N., & Fitch, W. T. (2002). The language faculty: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Blackwell.

Hornstein, N. (1984). *Logic as grammar*. The MIT Press.

Jacob, P. (2002). *Chomsky, cognitive science, naturalism and internalism*. https://jeannicod.ccsd.cnrs.fr/ijn_00000027

Kant, I. (1783). *Prolegomena to any future metaphysics*. Engl. Tr. L. W. Beck. Bobbs-Merrill, 1950.

McGilvray, J. (1998). Meanings are syntactically individuated and found in the head. *Mind & Language, 13*(2), 225–280.

Morris, C. (1938). *Foundations of the theory of signs*. The University of Chicago Press.

Pietroski, P. (2003). The character of natural language semantics. In Barber (Ed.), *Epistemology of language* (pp. 217–256). Oxford University Press.

Pietroski, P. (2005a). *Events and semantic architecture*. Oxford University Press.

Pietroski, P. (2005b). Meaning before truth. In G. Preyer & G. Peter (Eds.), *Contextualism in philosophy* (pp. 255–302). Clarendon Press.

Prawitz, D. (1977). Meaning and proofs: On the conflict between classical and intuitionistic logic. *Theoria, 43*, 2–40.
Prawitz, D. (1985). Remarks on some approaches to the concept of logical consequence. *Synthese, 62*, 153–171.
Prawitz, D. (2005). Logical consequence from a constructivist point of view. In S. Shapiro (Ed.), *The Oxford handbook of philosophy of mathematics and logic* (pp. 671–695). Oxford University Press.
Prawitz, D. (2015). Explaining deductive inference. In H. Wansing (Ed.), *Dag Prawitz on proofs and meaning* (pp. 65–100). Springer.
Segal, G. (1989). Seeing what is not there. *The Philosophical Review, 98*(2), 189–214.
Stainton, R. J. (2008). Meaning and reference: Some chomskyan themes. In E. Lepore, & B. Smith (Eds.), *The Oxford handbook of philosophy of language*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199552238.003.0036
Usberti, G. (1995). *Significato e conoscenza*. Guerini.
Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

Chapter 2
Varieties of Semantical Anti-realism

Abstract In the preceding chapter I argued that Chomskyan motivations for an internalist theory of meaning converge with certain motivations for semantical anti-realism. In this chapter, after a detailed characterization of semantical realism and anti-realism, two varieties of semantical anti-realism are introduced—mathematical intuitionism and neo-verificationism—and the question is discussed of which variety is best suited to the motivations specified in Chap. 1. More precisely, in Sect. 2.1 Dummett's semantic characterization of the the traditional debates between realists and anti-realists (realism/anti-realism debates, for short) is expounded, his proposal of bivalence as the criterion of realism in semantics is justified, and some necessary refinements and qualifications of this basic idea are introduced. The basic ideas of Brouwer and Heyting—the pères fondateurs of intuitionism—about logic and the meaning of the logical constants are outlined in Sect. 2.2, while Sect. 2.3 presents the neo-verificationist program(s) for a theory of meaning elaborated by Dummett, Prawitz and Martin-Löf as the outgrowth of some basic results in proof-theory like normalization theorems for natural deduction systems and Curry-Howard isomorphism. In Sect. 2.4 a comparison is made between intuitionist and neo-verificationist views about some basic issues (*tonk*'s problem, the canonical/non-canonical distinction, epistemic transparency of proofs), from which the rationale is extracted for preferring intuitionism as a source of inspiration for an internalist semantic.

Keywords Realism/Anti-realism · Dummett · Intuitionism · Heyting · Brouwer · BHK-explanation · Neo-verificationism · Tonk · Canonical/Non-canonical · Prawitz · Martin-Löf · Transparency · Prior · Theory of Grounds

2.1 Dummett's Semantic Characterization of the Realism/Anti-realism Debates

While in the traditional formulations the realism/anti-realism debates concern the *ontological* question whether entities of a certain kind *exist* (as the realist claims) or not (as the anti-realist claims), Dummett proposes a *semantic* characterization, inspired by the principle that

G. Usberti, *Meaning and Justification. An Internalist Theory of Meaning*, Logic, Epistemology, and the Unity of Science 59,
https://doi.org/10.1007/978-3-031-24605-0_2

(1) «the problem is not the existence of mathematical objects, but the objectivity of mathematical statements» (Dummett, 1978: xxviii).[1]

The main tenets of the semantic characterization can be summed up in the following points:

(2) The realism/anti-realism debates concern classes of *statements* rather than of *entities*.
(3) What is in discussion is the best way of characterizing the *meaning* of such statements.
(4) Assuming that meaning is to be characterized in terms of truth-conditions, what is in discussion is the notion of truth used in the characterization[2];
(5) the realist employs a notion satisfying the *principle of bivalence*, according to which
    Every proposition is either true or false;
    the anti-realist employs a non-bivalent notion.

Some refinements and qualifications are necessary. First of all, bivalent truth is seen by Dummett as a *necessary* condition of realism, not as a sufficient one. For a philosophical view to be called realist it must also make an essential appeal—in explaining how the truth-value of statements is determined—to the semantic, in the sense of relational, notion of reference, i.e. to a relation between singular terms and a domain of entities in general independent of the language. We can therefore schematically present Dummett's position through two theses:

(6) realism $\Rightarrow$ bivalent truth
(7) (bivalent truth) and (semantic reference) $\Rightarrow$ realism.

It should be clearly stressed that these are not to be seen as logical implications; realism, anti-realism and conceptions of truth are considered by Dummett as historically determinate philosophical positions, not easily nor unquestionably reducible to well defined classes of theses; (6) and (7) therefore do not express conceptual relations between classes of philosophical theses, but (hopefully illuminating) connexions between historical philosophical views.

According to (6) any non-bivalent semantics is a semantics based on an anti-realist theory of meaning. A plausible objection is that there are several semantics that admit, with different motivations, the existence of truth-value gaps, i.e. of sentences neither true nor false; but it seems inadequate to label the theories of meaning on which they are based as anti-realist. A typical example are non-bivalent semantics in which truth-value gaps are induced by the presence in the language of non-denoting terms[3]—a possibility acknowledged, with different motivations, both by Frege and by Strawson,

[1] Dummett ascribes this principle to Kreisel (in Dummett, 1975 he calls it "Kreisel's dictum"), but Sundholm has convincingly argued, in Sundholm (2020), that the principle was not endorsed by Kreisel, and that it should be ascribed to Dummett himself.

[2] The reason why I have stated thesis (4) in conditional form will be explained below.

[3] Another example are non-bivalent semantics employed to treat the phenomenon of vagueness. Dummett discusses them in Dummett (1995).

who cannot certainly be considered as anti-realists. To such an objection Dummett is sensitive, to the point of taking into consideration a possible weakening of thesis (6).[4] Let us call *principle of valence* the principle according to which

(8) Every statement is either true or not true,

and *objectivistic truth* any notion of truth that satisfies (8)[5]; the thesis

(9) realism ⇒ objectivistic truth

would be a weakening of (6), in the sense that it would yield a much more inclusive charaterisation of realism. For in a semantics with truth-value gaps the principle of valence is obviously valid, so that that semantics would be counted as realist; analogously, many-valued and possible worlds semantics would be classified as available choices for the realist.

However, Dummett decides *not* to choose objectivistic truth as a necessary condition of realism and to keep to bivalence, therefore continuing to classify the semantics just quoted as anti-realist. The rationale for this is essentially that Frege, and Russell in 1905, *were* in fact anti-realists *about possible objects*, against Meinong.[6] A consequence of Dummett's decision is that anti-realism appears now as a variegated multiplicity of positions. The principle of valence seems to be used by Dummett as a criterion do distinguish between objectivistic and non-objectivistic anti-realism; but the mere refusal of the principle of valence gives us no information as to how the notion of truth might be characterized. In Sect. 2.2 we will see a specific case of non-objectivistic semantics, which is moreover considered by Dummett as a particularly interesting prototype of anti-realistic semantics: the intuitionistic explanation of the meaning of the logical constants given by Heyting. In the Conclusion I shall come back to the question of how to characterize semantic objectivism.

2.2 Intuitionism

In this section and the following I will introduce two explanations of the meaning of the logical constants when applied to mathematical sentences which, according to Dummett's criterion of realism illustrated in Sect. 2.1, are two varieties of semantical anti-realism: mathematical intuitionism and neo-verificationism. Intuitionism is of course much more than an explanation of the meaning of the logical constants; but here I will restrict my attention to the intuitionists' views of logic and meaning.

---

[4] See for instance Dummett (1992).

[5] The reason why Dummett labels "objectivistic" such a notion of truth is that, according to it, a statement may be true even though we have no means of recognizing it as such.

The acceptance of (8) as a criterion of objectivism for a semantic theory is proposed for example in Dummett (1982), 242.

[6] Russell indeed avoided being a realist about possible objects without abandoning bivalence, by not accepting definite descriptions as genuine singular terms.

2.2.1 Brouwer

According to Brouwer there is a mathematical science «which consists of the *mathematical consideration of mathematics* or *of the language of mathematics.*» (Brouwer, 1907: 61, fn. 1) Presumably this 'second order' mathematics makes assertions which are true in virtue of the presence in the mathematician's mind of constructions performed at the level of intuitive mathematics. But it is a fact that Brouwer never minded to clarify such a notion of truth nor to develop a theory of meaning of the logical constants. On the other hand, Brouwer often discusses the validity of logical principles, and to do it he must assign them a precise meaning. It is therefore possible to extract from his writings the rudiments of an implicit theory of meaning. Let us see, in particular, how does he interpret implication. In Brouwer (1907), in the context of a discussion about the relations between mathematics and logic in which he manifests the conviction that hypothetical reasoning and the notion of contradiction should be explained away from the conceptual framework of intuitionistic mathematics, since their admission would mean to accept an at least partial dependence of mathematics on logic, he writes:

> In one particular case the chain of syllogisms [...] seems to come nearer to the usual logical figures and [...] actually seems to presuppose the hypothetical judgment from logic. Here it seems that the construction is *supposed* to be effected, and that starting from this hypothesis a chain of hypothetical judgements is deduced. [...] But this is no more than apparent; what actually happens is the following: one starts by setting up a structure which fulfills part of the required relations, thereupon one tries to deduce from these relations, by means of tautologies, other relations, in such a way that these new relations, combined with those that have not yet been used, yield a system of conditions, suitable as a starting-point for the construction of the required structure. Only by this construction will it be proved that the original conditions can be fulfilled. (Brouwer, 1907: 72–73)

Mark van Atten has proposed a convincing interpretation of the first part of this passage, according to which

> In order to establish A→B, one has to conceive of A and B as conditions on constructions, and to show that from the conditions specified by A one obtains the conditions specified by B, according to transformations whose composition preserves mathematical constructibility. (van Atten, 2009: 128)

On this reading any use of hypothetical constructions (hence of logic) is avoided by considering conditions on constructions instead of constructions themselves: «Instead of a 'chain of hypothetical judgements' that one seems to make, one is really making a chain of transformations in which from required relations (i.e., given conditions) further relations are derived» (van Atten, 2017: 14). It may be interesting to consider a specific example. There is a passage of Heyting (1960) in which Heyting gives a detailed account of how negation should be analyzed from an intuitionistic standpoint; it deserves to be quoted extensively in order to compare what Heyting had in mind with Brouwer's explanation of implication (on van Atten's reading). Heyting is arguing for the necessity of the (primitive) notion of hypothetical construction:

> As with the interpretation of general statements, the notion of a hypothetical construction is also essential for understanding the meaning of negation. For example, to prove that 2+2=5 is false, we make the following constructions: I and II: repeated construction of the number 2; III: construction of 2+2; IV: construction of 5; V: hypothetical construction of a one-to-one correspondence between the results of III and IV; VI: general method for deducing a contradiction from V. In the case of 2+2=5, the construction V is obtained by matching one by one the entities of which 4 is composed to entities of which 5 is composed; we see (VI) that there is always one entity of 5 that remains without a correspondent. Since construction V is hypothetical, the method applies to any correspondence between 4 and 5, so that we are entitled to assert the negation of 2+2=5. Brouwer (1907: 72) has made a useful remark on this subject. As we can see in the simple example considered here, the hypothetical construction V does not fulfill all the conditions of the problem (which would be impossible). Indeed, we construct a one-to-one correspondence between 4 and a subspecies of 5; then we find that this subspecies never exhausts 5. More generally, in demonstrating the negation of a proposition P, we describe a construction that satisfies some of the conditions contained in P, and we find that it violates one of the other conditions.$^{(1)}$
>
> ———
>
> (1) In the place quoted, Brouwer disputes the necessity of hypothetical constructions. His argument is valid in that one does not have to imagine a one-to-one correspondence between 4 and 5 to see that it cannot exist. But he forgets that it must be shown that *every* one-to-one correspondence between 4 and a subset of 5 misses an entity of 5; this general statement can only be proved by talking of a hypothetical correspondence. (Heyting, 1960: 179)[7]

In order to grasp the difference between Heyting and Brouwer it may be useful to distinguish two attitudes about hypothetical constructions[8]: an ontological one, according to which the relation between hypothetical and actual constructions is similar to the genus/species relation; and an epistemological attitude, according to

---

[7] «Comme pour l'interprétation des énoncés généraux, la notion d'une construction hypothétique est aussi indispensable pour comprendre le sens de la négation. Par example, pour démontrer que $2+2=5$ est faux, on fait les constructions suivantes: I e II: construction répétée du nombre 2; III: construction de $2+2$; IV: construction de 5; V: construction hypothétique d'une correspondance bi-univoque entre les résultats de III et IV; VI: méthode générale pour déduire de V une contradiction. Dans le cas de $2+2=5$, la construction V s'obtient en faisant correspondre une à une les entités dont se compose 4 à des entités dont se compose 5; on constate (VI) qu'il y a toujours une entité de 5 qui reste sans correspondant. Comme la construction V est hypothétique, la méthode s'applique à toute correspondance entre 4 et 5, de sort qu'on a le droit d'affirmer la négation de $2+2=5$. Brouwer (1907: 72) a fait à ce sujet une remarque bien utile. Comme on le voit dans l'exemple simple considéré ici, la construction hypothétique V ne remplit pas toutes les conditions du problème (ce qui d'ailleurs serait impossible). En effet, nous construison une correspondance biunivoque entre 4 et une sous-espèce de 5; ensuite nous constaton que cette sous-espèce n'épuise jamais 5. Plus généralement, en démontrant la négation d'une proposition P, nous décrivons une construction satisfaisant une partie des conditions contenues en P, et nous constatons qu'elle viole l'une des autres conditions.$^{(1)}$»

———

(1) A l'endroit cité, Brouwer conteste la nécessité de constructions hypothétiques. Son argument est valable en ce qu'on ne doit pas imaginer une correspondance biunivoque entre 4 et 5 pour constater qu'elle ne peut pas exister. Mais il oublie qu'il faut démontrer que *toute* correspondance biunivoque entre 4 et un sous-ensemble de 5 laisse échapper une entité de 5; on ne peut démontrer cette affirmation générale qu'en parlant d'une correspondance hypothétique.

[8] See Usberti (1995: 26 and 28).

which there is only one kind of entities, constructions, towards which two kinds of cognitive acts can be distinguished: the act of *performing* a construction, and the act of *conceiving* a construction. Heyting's passage encourages the ontological attitude: when he writes «in demonstrating the negation of a proposition P, we describe a construction that satisfies some of the conditions contained in P», he is saying that what we are assuming to have is a construction c satisfying only some of the conditions contained in "$2 + 2 = 5$", i.e. that c is, strictly speaking, an *approximation* to a construction for the proposition in question. As a consequence, a general explanation of the meaning of negation would involve the further, obscure and problematic, notion of approximation to a construction: the ontological attitude seems misleading.

Let us apply Brouwer's explanation of implication (on van Atten's reading) to Heyting's example. At step V, instead of *constructing* an approximation to a one-to-one correspondence between $2 + 2$ and 5, we *conceive* an arbitrary construction c onto which the condition C is imposed of being a one-to-one correspondence between $2 + 2$ and 5; at step VI we tranform C into the condition C' of being a one-to-one correspondence between 1 and 2 (through approximately the following steps: represent $2 + 2$ as $<x, w, y, z>$, 5 as $<x', w', y', z', e>$; hence associate x' with x, …, z' with z, e with z; then the 'final segment' of c is a one-to-one correspondence between 1 and 2); at step VII we observe that c is a construction satisfying C', i.e. a construction "of an incompatibility" or of a known falsehood. As it can be seen, it is condition C that has been transformed, not contruction c, which has simply been conceived: the fact that it satisfies C' shows that it does not exist, i.e. that it cannot be constructed.

It is important to stress that condition C' exists, i.e. is well defined, even if c does not exist. This is essential for the possibility of transforming a construction for α into one for β when it is unknown whether a construction for α does exist: what is decisive is the possibility of transforming a *condition* imposed onto a construction for α into a *condition* imposed onto a construction for β. Such a possibility is not guaranteed in general; for example, it does not seem possible to transform a construction satisfying the condition of being a proof of "$1 = 2$" into one satisfying the condition of being a proof of "There is a greatest natural number". As a consequence—van Atten concludes (van Atten, 2009: 130)—the *ex falso sequitur quodlibet* is not valid in general.

2.2.2 Heyting

In 1930 Heyting presented in three papers intuitionistically acceptable formal systems for propositional and predicate logic, arithmetic, the theory of spreads, species and choice sequences. Since then the problem arose of giving a semantical interpretation of the logical constants—a particularly compelling problem for someone who didn't believe that an axiomatic system can express axiomatic thought by itself, independently of the meaning of the symbols occurring in it. As a matter of fact, as early as 1930 Heyting «had at least implicitly a clear grasp of the intuitionistic meaning of the logical operators» (Troelstra, 1981: 16), as can be inferred from the fact that he arrived at his formalization by closely examining the axioms and theorems of *Principia Mathematica* and by making a system out of the acceptable ones.[9]

The attention Heyting paid to the linguistic expression of the mathematical activity entailed for him the need of giving a characterization of mathematics from which its intrinsic communicability followed.

The goal of every science is, according to Heyting, «to organize domains of experience which so far seemed quite apart from each other in a wider structure which embraces them together.» (Heyting, 1974: 90) From this point of view there is no essential difference between the activity of the reason in science and in ordinary life: in both cases it is a spontaneous activity, although in science the search for regularities is pursued consciously and often through the collaboration of several individuals.

From an abstract viewpoint a collection of objects among which there are certain relations is a structure; hence the goal of every science is to isolate (or to introduce) structures in the experiential continuum. On the other hand, modern mathematics is no more reducible to the science of number and measure; its methods are so pervasive that it can be adequately described just as the general theory of structures. From this it follows that all sciences, from physics to the humanities, aren't but applied mathematics. They differ from each other only in the manner in which they gather their material, not in the methods by which they order it.

But what is mathematics? Heyting conceives it as a form of free activity performed almost continually by our mind, consisting in focusing our attention on a perception, i.e. in isolating a perception within the continuous and originally unstructured flux of the perceptual *erlebnis*. With a typical shift Heyting considers less committing to speak, instead of isolating a perception or an object, of *creating an entity*. (Heyting, 1974: 80) The fundamental activity of our mind when we are awake is therefore creating entities. This act can be repeated in such a way that the mind, after having created an entity, creates another one keeping the first into memory. This is the genesis of the process of counting, by which we construct mentally the natural numbers.

The concept of natural number (or better, of iterable creation of an entity) is the first of the fundamental notions postulated by Heyting in order to explain mathematical activity. Other such notions are *hypothetical construction* and *general method* of construction. For example, if I wish to prove that every natural number has a greater prime number, I must take an arbitrary natural number n, calculate $n! + 1$ and factorize this number; each of its prime factors will be greater than n. Hence the correct interpretation of an universally quantified proposition is hypothetical, and the proof of such a hypothetical proposition is a general method of construction which, applied to a hypothetical construction of n, yields the construction of a prime number greater than n. Other fundamental notions are those of *contradiction*, necessary in order to explain negation, and of a *choice sequence* (as an element of a spread), needed to account for the intuitionistic conception of the continuum.

---

[9] Letter to O. Becker of 23.VII.1933, quoted in Troelstra (1981: 16).

In the above list of fundamental concepts Heyting mixes mathematical *entities*, such as natural numbers and choice sequences, with abstract notions needed to explain the meaning of mathematical sentences. In other words, when he speaks about mathematics he makes reference to an activity which, although mental, grows and articulates through language, in the sense that it must be conceived as consisting essentially in proving mathematical propositions. And such an activity must be essentially communicable. Of course, the basic difference from Hilbert remains: mathematical assertions are not to be understood as formulas, i.e. as mere arrays of signs to be investigated metamathetically, but as sentences endowed with sense.

Against Frege's view of logic as the science of the laws of truth, Heyting raises the following objection. We have seen that an essential feature of the (classical) notion of truth is that it satisfies the principle of bivalence (5); this amounts to asserting that truth is a notion transcending our cognitive capacities, in the sense that a sentence can be true (or false) even if it is not possible for us to recognize it as such, solely because 'that's how things are'. But the existence of an external reality which renders each proposition true or false is an ontological hypothesis Heyting holds cannot be accepted as a method to prove mathematical assertions:

> The intuitionist mathematician, as a mathematician, will not oppose a philosophy which holds that the mind, in its creative activity, reproduces beings from a transcendent world, but he will consider this doctrine too speculative to serve as a foundation for pure mathematics. (Heyting, 1939: 73)[10]

Notice how even on this point, on which he agrees completely with the father of intuitionism, Heyting's tone is peculiarly different from Brouwer's: instead of refusing the classical notion of truth in the name of an idealistic view of the foundations of mathematics, he appeals to the neutrality of mathematics, and more generally of science, with respect to philosophical options.

Moreover, adopting classical truth as the key concept of a meaning theory of the logical constants commits us to a quite unsatisfying explanation of the meaning of implication. Heyting alludes to the criticisms against material implication as an *explicans* of the intuitive notion of implication and quotes explicitly C. I. Lewis (in Heyting, 1956a, 1956b), who just started from an analysis of the so-called paradoxes of material implication and of the relations between implication and deducibility to develop the first systems of modal logic. He isn't anyway sympathetic with Lewis' modal approach, which he charges for explaining *obscura per obscuriora.* The main objection he raises against the classical construal of implication is that, if the truth-value of a proposition is conceived as determined independently of our knowledge, it cannot be understood how the truth-value of a proposition may depend on the one of another.

The notion of truth ought to be replaced, according to Heyting, with that of *knowledge.* It is therefore a *logique du savoir* Heyting opposes to classical logic conceived

[10] «Le mathématicien intuitionniste, en mathématicien, ne s'opposera pas à une philosophie qui soutiendra que l'esprit, dans son activité créatrice, reproduit des êtres d'un monde transcendant, mais il considérera cette doctrine comme trop spéculative pour servir de fondement aux mathématiques pures.»

by him as a *logique de l'être*. But in this way—it might be objected—intuitionistic logic presupposes, instead of an ontology, a theory of knowledge. Heyting answers that

> Some starting point must be chosen; what matters is that the basic notions be as immediate as possible. Now, at least for man, knowledge is more immediate than being, which manifests itself for him only through an analysis of knowledge. (Heyting, 1956b: 228)[11]

The explanation of implication no longer gets into the above difficulty, since it is quite natural that *knowledge* of a proposition may depend on *knowledge* of another.

Someone might object that knowledge is nothing but knowledge of truth, so that a logic of knowledge cannot but rely ultimately on the notion of truth. This would be a serious misinterpretation of Heyting; according to him knowledge *replaces* truth in the role of basic concept of intuitionistic logic: «The notion of truth makes no sense [...] in intuitionistic mathematics» (Heyting, 1958: 279); as a consequence, there is no truth preexisting to our knowledge: if we want to go on using the word "truth", we could say that, vice versa, it is our knowing (or judging, in Frege's terminology) a proposition that makes it a truth. This was, by the way, Brouwer's attitude towards truth: «there are no non-experienced truths» (Brouwer, 1949: 488).

As for the notion of proposition, Heyting writes:

> A mathematical proposition expresses a certain expectation. For example, the proposition, "Euler's constant *C* is rational" expresses the expectation that we could find two integers *a* and *b* such that $C=a/b$. Perhaps the word "intention", coined by the phenomenologists, expresses even better what is meant here. We also use the word "proposition" for the intention which is linguistically expressed by the proposition. (Heyting, 1931: 58–59)

Heyting uses therefore the term "proposition" in a double sense: both for a (meaningful) sentence α and for what is expressed by α; in order to avoid confusion I shall use "(meaningful) sentence" for the first sense, and will reserve "proposition" for the second sense. I will therefore say that, according to Heyting,

(10) A mathematical (meaningful) sentence expresses a certain expectation or intention.
(11) A proposition is the expectation or intention expressed by a mathematical (meaningful) sentence.

A natural question arises: A mathematical sentence expresses an expectation *of what*? An explicit answer is contained in the following passage:

> [E]ach proposition means [...] the intention of a mathematical construction which must satisfy certain conditions. A proof of a proposition consists in the realisation of the construction required in it. (Heyting, 1934: 14)[12]

---

[11] «Il faut choisir quelque point de départ; ce qui importe c'est que les notions de base soient aussi immédiates que possible. Or, du moins pour l'homme, le savoir est plus immédiat que l'être, qui ne se manifeste pour lui que par une analyse du savoir.»

[12] «[J]ede Aussage steht [...] für die Intention auf eine mathematische Konstruktion, die bestimmten Bedingungen genügen soll. Ein Beweis für eine Aussage besteht in der Verwirklichung der in ihr geforderten Konstruktion.»

Summing up:

(12) The expectation or intention expressed by a mathematical sentence α is the expectation of a proof of α.

(13) A proof of α is the realization of the expectation/intention expressed by α.[13]

While (12) tells us what the expectation/intention expressed by α *is*, it does not yet tell us what does *grasping* it consist in. If we keep present Frege's analysis of grasping the thought expressed by α, according to which it amounts to knowlege of the truth-conditions of α, and we remark that proofs of α are what makes α evident to us, we can surmise that, according to Heyting, grasping the expectation/intention expressed by α amounts to knowlege of the evidence-conditions of α; and while it seems difficult to further analyze knowledge of (classical) truth-conditions, knowledge of evidence-conditions is convincingly explained as being in a position to recognize what counts as a proof of α.[14] In this way we arrive at the following suggestion:

(14) Grasping the expectation of a proof of α amounts to being in a position to recognize what counts as a proof of α.[15]

An immediate consequence of (10)–(14) is

(15) Understanding a mathematical sentence (grasping the proposition it expresses) amounts to being in a position to recognize what counts as a proof of it;

and an immediate consequence of (15) is

(16) A subject $\mathcal{S}$ who understands a mathematical sentence α is in a position to recognize what counts as a proof of α:

if $\mathcal{S}$ were not in such a position, (s)he would not understand α, contrary to the assumption that (s)he understands α.[16] I will call "epistemic transparency" the property of proofs expressed by (16).

---

[13] The close agreement between (12)–(13) and Kolmogorov's explanation in Kolmogorov (1932) is clearly put into evidence in Sundholm (1983).

[14] For a similar suggestion see for example Dummett (1981: 507).

[15] Throughout this book I use "in a position" in the sense defined by Williamson (2000: 95) for being in a position to know:

> To be in a position to know p, it is neither necessary to know p nor sufficient to be physically and psychologically capable of knowing p. No obstacle must block one's path to knowing p. If one is in a position to know p, and one has done what one is in a position to do to decide whether p is true, then one does know p.

[16] I have not found, in my (partial) exploration of Heyting's writings, an explicit statement of (16). Explicit attributions to the intuitionists of the view expressed by (16) can be found in Kreisel («the basic intuitionistic idealization that we can recognize a proof when we see one», Kreisel, 1962: 202) and in Dummett:

> mathematical objects themselves are mental constructions […] in the sense that, for them, *esse est concipi*. They exist only in virtue of our mathematical activity, which consists in mental operations, and have only those properties which they can be recognized by us as having. (Dummett, 2000: 5)

A (meaningful) sentence for which we have a proof may be *asserted*:

(17) Having a proof of α is a necessary and sufficient condition under which α can be asserted.

This is shown by the fact that in Heyting (1956a, 1956b) the inductive definition of the notion of proof of α (see below) is replaced by an inductive specification of «necessary and sufficient condition under which a complex expression can be asserted» (Heyting, 1956a, 1956b: 101). It seems legitimate to say that, according to Heyting, (17) is a consequence of the epistemic character of proofs, viz. of their property of conferring *evidence* to mathematical sentences; for example, the passage quoted above from Heyting (1931) continues as follows: «The intention [...] refers not only to a state of affairs thought to exist independently of us but also to an experience thought to be possible [...].»

Like many other words of the same kind, "assertion" is ambiguous between the act of affirming and the result of such an act. Although it is not possible (nor even advisable, under several respects) to remove this ambiguity, Heyting clearly favors the construal of assertion as an act:

(18) «The affirmation of a proposition is not itself a proposition; it is the determination of an empirical fact, viz., the fulfillment of the intention expressed by the proposition.» (Heyting, 1931: 59)[17]

It seems therefore clear that, according to Heyting, there is only one kind of entities: propositions, but they can be the (intentional) objects of different kinds of acts, among which there is assertion. Strictly speaking assertion does not belong to mathematics since it is the determination of an empirical fact[18]; mathematical theorems, as affirmations *about* mathematical activity, are not themselves part of that activity.

On the contrary, what gives evidence to propositions, namely proofs, *are* mathematical objects, as Heyting explicitly says:

(19) «A proof of a proposition is a mathematical construction which can itself be treated mathematically.» (Heyting, 1931: 60)

The points (10)–(19), which synthesize Heyting's views on the meaning of mathematical sentences, show that the key concept of his theory of meaning is the concept of proof, but they do not explain *what* a proof of α is; this is the job of his definition of this notion, by induction on the logical complexity of α. As the definition lacks in fact an (explicit) base clause, it is to be understood as an explanation of the (intuitionistic) meaning of the logical constants.

The logical constants denote logical functions. «A logical function is a process for forming another proposition from a given proposition.» (Heyting, 1931: 59) While

[17] According to Heyting, «An assertion [*Satz*] is the affirmation [*Behauptung*] of a proposition [*Aussage*].» (Heyting, 1931: 58).

[18] See also Heyting (1960: 178).

for Frege, who conceived logic as the theory of truth, the sense of a logical function is to be explained in terms of truth-conditions, for Heyting, who interprets logic as the theory of evidence, it is to be explained in terms of proofs, since it is proofs which confer evidence to propositions. But for the two of them the explanation must be inductive, since the sense of a complex sentence is a function of the senses either of its subsentences or of sentences of lesser complexity.

Heyting gave several versions of his explanation of the intuitionistic meaning of the logical constants. I choose the last one (from Heyting, 1979: 851), as, presumably, the most satisfying from his point of view. The leading principle of this explanation is the *principle of positivity*, according to which every mathematical theorem must express the result of a mathematical construction (see the quotation above from Brouwer, 1954).

In consequence of (10)–(19), Heyting's meaning explanation consists in fact in a definition of the notion of proof of α, by induction on the logical complexity of α:

(20) **Heyting's Explanation**

(H∧) A proof of α∧β consists of a proof of α and a proof of β.
(H∨) A proof of α∨β consists of a proof of α or a proof of β.
(H¬) A proof of ¬α consists of a general method which transforms any hypothetical proof of α into a contradiction.[19]
(H→) A proof of α→β consists of a general method which transforms any given proof of α into a proof of β.
(H∀) A proof of ∀xα, where the domain of x is the species S, consists of a method M such that, if C denotes the construction of an entity d together with the proof that d∈S, then M transforms C into a proof of α[d/x].
(H∃) A proof of ∃xα, where the domain of x is the species S, consists of the construction of an entity d, of a proof that d∈S, and of a proof of α[d/x].

It is important to understand exactly the status of this definition of the concept of proof of α, which Heyting proposes as, simultaneously, an explanation of the intuitionistic meaning of the logical constants. Let us start from the remark that, at first sight, there is something strange in the very idea of inductively defining the notion of proof of α, since it seems that there is no need of such a definition: a proof of α—it might be suggested—is any argument which makes evident that certain objective facts subsist; for example, a proof of ∀xα is any argument which makes evident that the facts $\alpha[d_1/x]$, $\alpha[d_2/x]$, ... subsist. A suggestion of this kind is just what an intuitionist cannot accept: it makes reference to the notion of objective fact, and this notion is unacceptable within the framework of the intuitionist view of mathematics. In the absence of objective facts to prove, what do intuitionistic proofs prove? Heyting's explanation is just an answer to this question: defining what a proof of a complex sentence is in terms of what a proof of each of its components is, he explains what a proof of α is without saying which objective facts it must prove; or,

[19] Heyting explains that «The notion of contradiction has to be considered as fundamental: $1 = 2$ may be taken as a paradigm.»

in other terms: a fact is nothing but a true proposition, and a true proposition is a proved proposition.[20]

This means that all notions occurring in the inductive definition must be assumed as primitive, and the 'definition' itself must be understood not as a formal definition, but as an *elucidation* in the sense Frege gave to this word in a famous passage from "Logic in Mathematics" (Frege, 1979: 207). On this point Heyting is explicit; in Heyting (1960: 179–180), for instance, he writes:

> The main fundamental notions of constructive mathematics, encountered so far, are I: construction of a natural number, II: hypothetical construction, III: general method of construction, IV: contradiction. It is impossible to give a definition of these notions that satisfies the mathematical conditions of precision. Moreover, they are not primitive notions in the sense given to this word in formal mathematics. We have just found them a posteriori through an analysis of mathematical constructions, but they are not posited a priori at the basis of these constructions.[21]

The grasp of this fundamental point seems to be the gist of the philosophical "sea change" stressed by van Atten (2015: 198 ff.) between Gödel's objection to Heyting's explanation when he writes in Gödel (1941: 190) that the clause for implication requires that

> the notion of derivation or of proof must be taken in its intuitive meaning as something directly given by intuition, without any further explanation being necessary. This notion of an intuitionistically correct proof or constructive proof lacks the desirable precision [,]

and Gödel's decision, in Gödel (1958), to take the notion of computable functional as primitive, thereby accepting Heyting's explanation in two strictly related footnotes of that paper, when he writes

> As is well known, A. M. Turing used the concept of a computer in order to give a definition of the concept of a computable function of first order. But *if this latter concept was not intelligible, then the question whether Turing's definition is adequate would be meaningless* [;]. (Gödel, 1958, 139, fn. 8; *emphasis added*)

and, immediately before:

> One may doubt whether we have a clear enough characterisation of this concept [of computable functional of finite type], but not wheter it satisfies the axioms given [in this paper]. We meet the same apparent paradox with a concept which lies at the base of intuitionistic logic, viz. the concept of a materially correct proof. (*Ibid.*, fn. 7)

[20] I will come back to this point in Sect. 6.3.2.2 of Chap. 6.

[21] «Les principales notions fondamentales des mathématiques constructives, rencontrées jusqu'ici, sont I: construction d'un nombre naturel, II: construction hypothétique, III: méthode générale de construction, IV: contradiction. Il est impossible de donner de ces notions une définition satisfaisant aux conditions mathématiques de précision. D'ailleur, elle ne sont pas des notions primitives au sens donné à ce mot dans les mathématiques formelles. Nous venons de les trouver a posteriori par une analyse des constructions mathématiques, mais elles ne sont pas posées a priori à la base de ces constructions.»

An important consequence of the primitive nature of the notions used in the inductive definition (20) is that the whole metalanguage in which it is formulated is intuitionistic, in the sense that the very logical constants occurring in it must be understood intuitionistically; Heyting's explanation, like Frege's one, is not reductive.

Let us consider some specific aspects of this explanation.

IMPLICATION

1. Although the notion of general method is primitive, we can try to elucidate it further. A proposal in this sense has been to conceive a general method as an open proof, i.e. as an argument with undischarged assumptions.[22] On this interpretation, the method of transforming any proof of α into a proof of β consists in merely appending the proof of β from α to the proof of α. A problem with this[23] is that such an 'appending' is a *uniform* operation, in the sense that it does not depend upon the structure of the proof of α; as a consequence it cannot account for cases in which we recognize some operation which involves internal transformation of any given proof of α as nevertheless always yielding a proof of β. For example, the inference from (i) $\forall x(\alpha \rightarrow \beta)$ to (ii) $\exists x\alpha \rightarrow \exists x\beta$ is intuitionistically valid, but if we have an open proof of (i) and a proof of $\exists x\alpha$, we cannot obtain a proof of $\exists x\beta$ by merely appending something to the open proof: we need to know for which particular number *n* the proof of $\exists x\alpha$ yielded a proof of $\alpha(\underline{n})$.

A different idea is proposed in Heyting (1934), according to which $\alpha \rightarrow \beta$ means «the intention of a construction which leads from every proof of α to a proof of β» (Heyting, 1934: 14); to prove $\alpha \rightarrow \beta$ we need therefore an operation *on* proofs of α, not a proof of β from α as assumption: we need a *function* from proofs of α to proofs of β.

I will therefore understand Heyting's clause for implication in this functional sense:

(H′→) A proof of $\alpha \rightarrow \beta$ is a (computable) function *f* such that, for any x, if x is a proof of α, then f(x) is a proof of β.[24]

The functional interpretation of implication is incorporated into Troelstra's and van Dalen's formulation of Heyting's clauses, called by them "The BHK-interpretation" (BHK for "Brouwer-Heyting-Kreisel", or "Brouwer-Heyting-Kolmogorov"); van Atten (van Atten, 2017: 6–7) prefers "The Proof Interpretation" or, better, "The Proof Explanation". I record here the whole Proof Explanation for future reference (from Troelstra and van Dalen, 1988: 9):

---

[22] The proposal is inspired by Gentzen's rule of implication introduction (Gentzen, 1935); it will be discussed in Sect. 2.3.2.

[23] Cp. Dummett (2000: 9).

[24] See also Prawitz (2015: 85):

> *c* is construction of A → B iff *c* is an effective operation that applied to any construction *c*′ of A yields as value a construction *c*(*c*′) of B.

(21) **The Proof Explanation**

(PE∧) A proof of $\alpha \wedge \beta$ is given by presenting a proof of α and a proof of β.

(PE∨) A proof of $\alpha \vee \beta$ is given by presenting either a proof of α or a proof of β (plus the stipulation that we want to regard the proof presented as evidence for $\alpha \vee \beta$).

(PE→) A proof of $\alpha \rightarrow \beta$ is a construction which permits us to transform any proof of α into a proof of β.

(PE¬) Absurdity $\perp$ (contradiction) has no proof; a proof of $\neg\alpha$ is a construction which transforms any hypothetical proof of α into a proof of a contradiction.

(PE∀) A proof of $\forall x\alpha$ is a construction which transforms a proof of $d \in D$ (D the intended range of the variable x) into a proof of $\alpha[d/x]$.

(PE∃) A proof of $\exists x\alpha$ is given by providing $d \in D$, and a proof of $\alpha[d/x]$.

2. An important objection to the functional interpretation of implication is that clause (H′→) is impredicative. Here is how van Atten states the objection:[25]

> How is the domain of such a function *f* to be understood?
>
> It cannot be given by an inductive definition of all proofs of A, as there can be no such definition. In particular, proofs of A may themselves contain the implication A→B, for example in this way:
>
> $$\frac{(A \rightarrow B) \rightarrow A \quad A \rightarrow B}{A} \rightarrow E$$
>
> Thus a prior explanation of A → B would be required, rendering the definition circular.
>
> But if the domain is to be all proofs of A in an absolute sense that can not be given by an inductive definition, then, so the claim goes, the definition of such a function *f* will take a form that renders it impredicative:
>
> f is a function such that, for any x in the totality of all intuitionistic proofs, if x is a proof of A, then f(x) is a proof of B.
>
> Any specific definition of this form will define an individual proof of A→B by referring to a totality to which it belongs, and thus be impredicative. (van Atten, 2018: 3)

The assumption which the whole argument depends upon is that the domain of *f* «cannot be given by an inductive definition of all proofs of A, as there can be no such definition.» But this is not true: Heyting's explanation is just such an inductive definition. How to explain, then, the assumption? In my opinion it is due to a failure to keep apart two notions of proof that should be accurately distinguished: *evidential* and *inferential* proofs of α.[26] The former are defined by induction on the logical

[25] For a similar formulation see Dummett (2000), 269–270.

[26] A different position about the objection of impredicativity is expressed in Martino and Usberti (1988), where a clear distinction between evidential and inferential proofs was not recognized. More precisely, I still agree that «a proof of $\alpha \rightarrow \beta$ should be a method applicable only to proofs of α not involving proofs of more complex propositions» (Martino & Usberti, 1988: 150), but I hold that this is just what Heyting does, since he is defining evidential, not inferential, proofs. On the other hand, I still agree that there is a reason for the impredicativity of Heyting's inductive definition itself, based on the fact the standard of evidence may change through time and is therefore indefinitely extensible (Martino & Usberti, 1988: 149–150).

complexity of α, the latter by induction on the number of inferential steps employed to obtain α as conclusion; this difference corresponds to an intuitive one: an evidential proof of α is a way of making α evident, while an inferential proof of α is a sequence of steps through which α is obtained as a conclusion from certain premisses. In a sense the difference corresponds to the one between closed and open proofs in a natural deduction system, as having an evidential proof of α entitles one to assert α, while having an inferential proof of α entitles one to assert α only if one is entitled to assert the assumptions α depends on; but the correspondence is misleading, since what is constitutive of an evidential proof of α is what makes α evident (a certain pair if α is a conjunction, a certain function if α is an implication, etc.), not the sequence of steps through which the evidence for the premises is transformed into evidence for α.

The charge of impredicativity concerns the notion of inferential proof, as is evident from the example used by van Atten to show the circularity of the definition. Notice that, in the example, the validity of $\rightarrow -$ E is presupposed; this is of course legitimate only if proofs are seen as sequences of valid inferential steps, and applications of $\rightarrow -$ E are recognized as valid inferential steps. But this is not the case when Heyting defines the notion of proof of α by induction on α: the validity of $\rightarrow -$E is a *consequence* of the definition, not a part of it; hence Heyting is defining a different notion of proof: not a sequence of inferential steps but a way of obtaining evidence. The tacit inductive hypothesis of (H→) is that what has been explained are the evidential proofs of α, and the complexity of α is less than the one of $\alpha \rightarrow \beta$: we cannot presuppose that the totality of the evidential proofs of α we assume to know contains $\alpha \rightarrow \beta$ as an inferential step.[27]

The importance of the notion of evidential proof as distinct from the one inferential proof is clearly appreciated by Prawitz:

> The realization of the required construction is called a proof by Heyting. Although the term proof is here used in its usual epistemic sense in so far as the realization of the intended construction is the requirement for asserting the sentence, the explanation of the term does not presuppose the notion of inference. Therefore we have again (as in [Proof-Theoretic Semantics]) a candidate for how to account for legitimate inferences. (Prawitz, 2015: 84)

3. Kreisel has proposed that Heyting's clause for implication should be supplemented with a 'second clause' requiring a proof that the method in question yields, for every d, a proof of α(d/x).[28] Dummett has proposed a rephrasing of Heyting's clause according to which the operation which a proof of $\alpha \rightarrow \beta$ consists of is such that *we can recognize* that, applied to a proof of α, it yields a proof of β (Dummett, 2000: 8).

The rationale for these proposals is the necessity to guarantee the epistemic transparency of proofs required by Heyting's explanation (see point (16) above). However, both the second clause and Dummett's rephrasing are motivated by the

[27] I will come back to the, to me fundamental, distinction between evidential and inferential proofs in Chap. 5.

[28] See for example Kreisel (1962: 205).

false presupposition, induced by a classical understanding of the metalanguage in which Heyting's definition is formulated, that it may happen that a general method yields, as a matter of fact, a proof of β for every proof of α, even if no mathematician is in a position to know that it does; on the contrary, from an intuitionistic standpoint, there are no knowledge-independent facts, so that there is no room for distinguishing between the mere fact that the method works and a proof of this fact. Hence both proposals are both pleonastic and misleading.[29]

NEGATION

1. In order to meet the principle of positivity, Heyting adopts the hypothetical interpretation of negation, according to which negation becomes a particular case of implication: if ⊥ denotes the Contradiction, ¬α can be defined as an abbreviation of α → ⊥. Note, however, that, if ¬α is defined in this way, (H¬) in (20) should be slightly modified: a proof of ¬α should be required to transform any hypothetical proof of α not into a contradiction, but into *a proof* of a contradiction. This is exactly clause (PE¬) of the Proof Explanation (21).[30] We will see some consequences of the adoption of (PE¬) in Chap. 5.
2. An unclear point of Heyting's explanation is his justification of *ex falso sequitur quodlibet* (EFQ):

> Axiom [¬p→(p→q)] may not seem intuitively clear. As a matter of fact, it adds to the precision of the definition of implication. You remember that p→q can be asserted if and only if we possess a construction which, joined to the construction p, would prove q. Now suppose that ⊢¬p, that is, we have deduced a contradiction from the supposition that p were carried out. Then, in a sense, this can be considered as a construction, which, joined to a proof of p (which cannot exist) leads to a proof of q. (Heyting, 1956a, 1956b: 102)

Mark van Atten has convincingly criticized this justification:

> One easily recognizes Heyting's effort to explain *Ex Falso* as much as possible along the same lines as other implications, namely, by providing a concrete construction that leads from the antecedent to the consequent. [...] But it does not fit into Heyting's original interpretation of logic in terms of intentions directed at constructions and the fulfillment of such intentions either. For to fulfill an intention directed toward a particular construction we will have to exhibit that construction; we will have to exhibit a construction that transforms any proof of p into one of q. But how can a construction that from the assumption p arrives at a contradiction, and therefore generally speaking not at q, lead to q? It will not do to say that such a construction exists "in a sense". A construction that is a construction "in a sense", as Heyting helps himself to here, is no construction. (van Atten, 2017: 63–64)

---

[29] This may be the reason why the requirement, contained in Troelstra (1977: 977), that a proof of α → β includes the insight that the construction has the required property has been dropped in the Proof Interpretation of Troelstra and van Dalen (1988). It seems to me that, for the same reason, the requirement, in (20)(H∀), that M applies to C ought to be replaced by the requirement that M applies to d.

[30] It should also be noted that Heyting's clause for negation has "any hypothetical proof of α" where the clause for implication has "any given proof of α". (I thank an anonymous referee for this remark.) One might wonder whether "hypothetical proof" and "given proof" mean the same; I would answer yes, in view of the fact in other papers [for instance Heyting (1974)] Heyting explicitly says that general methods apply to hypothetical contructions.

I shall come back to this point in Chap. 5.

From this presentation it is easy to see why the intuitionist refuses bivalence; α is true iff there is a proof of α, and it is false iff there is a proof of its negation; but it does not hold that, for every α, either there is a proof of α or there is a proof of its negation; therefore it does not hold that, for every α, either α is true or α is false.

2.3 Neo-Verificationism

The distinctive feature of Neo-Verificationism is the establishment of a strict parallelism between some important results in general proof theory (normalization theorems, Curry-Howard isomorphism) and the architecture of an intuitionistic theory of meaning of the logical constants. In this section I shall describe this parallelism, starting with a description of the relevant results in proof theory, and then outlining the architecture of neo-verificationist theories of meaning.

2.3.1 Some Results in Proof Theory

In Gentzen (1935) a "calculus of natural deduction" is set up, i.e. «a formalism that reflects as accurately as possible the actual logical reasoning involved in mathematical proofs» (Gentzen, 1935: 74). To this purpose he analyzed each argument into atomic steps and stated the rules for such steps. His analysis was of a highly systematic character in that to each logical constant **C** rules of two kinds are associated: the former—***C****-introductions*—establish the premisses under which we are entitled to infer a formula having **C** as its principal operator, thereby introducing it into an argument; the latter—***C****-eliminations*—determine the conclusion we are authorized to infer from a formula having **C** as its principal operator (maybe together with other premisses), thereby eliminating it from an argument. Taken together, introductions and eliminations can therefore be seen as the *rules for the use* of the logical constants in deductive practice.

Although Gentzen's main purpose was to give a formalization of logic, when he explained the rules of the natural deduction calculus he suggested an idea of fundamental importance both for proof theory and for the neo-vericationist theories of meaning. Here is how he states it:

> The introductions represent, as it were, the 'definitions' of the symbols concerned, and the eliminations are no more, in the final analysis, than the consequences of these definitions. This fact can be expressed as follows: In eliminating a symbol, we may use the formula with whose terminal symbol we are dealing only 'in the sense afforded it by the introduction of that symbol'. An example may clarify what is meant: We were able to introduce the formula A→B when there existed a derivation of B from the assumption formula A. If we then wished to use that formula by eliminating the →-symbol [...], we could do this precisely by inferring B directly, once A has been proved, for what A→B attests is just the existence of a derivation of B from A. [...]

> By making these ideas more precise it should be possible to display the E-inferences as unique functions of their corresponding I-inferences, on the basis of certain requirements. (Gentzen, 1935: 80–81)

The exact relation between introductions and eliminations in Gentzen's natural deduction system is characterized by Prawitz in the following terms:

> What makes Gentzen's systems especially interesting is the discovery of a certain symmetry between the atomic inferences, which may be indicated by saying that the corresponding introductions and eliminations are *inverses* of each other. The sense in which an elimination, say, is the inverse of the corresponding introduction is roughly this: the conclusion obtained by an elimination does not state anything more than what must have already been obtained if the major premiss of the elimination was inferred by an introduction. For instance, if the premiss of an $\wedge$E was inferred by introduction, then the conclusion of the $\wedge$E must already occur as one of the premisses of this introduction. Similarly, if the major premiss A$\rightarrow$B of an $\rightarrow$E was inferred by introduction, then a proof of the conclusion B of the $\rightarrow$E is obtained from the proof of the major premiss of the $\rightarrow$E by simply replacing its assumption A by the proof of the minor premiss. [...]
>
> In other words, a proof of the conclusion of an elimination is already 'contained' in the proofs of the premisses when the major premiss is inferred by introduction. We shall refer to this by saying that the pairs of corresponding introductions and eliminations satisfy the *inversion principle*. (Prawitz, 1971: 246–247)

The reason why the inversion principle is interesting for proof theory is that its validity in a formal system permits to *reduce* its proofs to a *normal form* by eliminating from them any sort of 'redundancy'. For instance, the proof on the left of (22), containing two consecutive applications of the rule of introduction of $\wedge$ ($\wedge$I) and of one rule of elimination of $\wedge$ ($\wedge E_1$), reduces to the proof on the right:

(22) $$\dfrac{\dfrac{\begin{array}{cc}\pi_1 & \pi_2\\ \alpha & \beta\end{array}}{\alpha\wedge\beta}\wedge I}{\alpha}\wedge E_1 \qquad\Longmapsto\qquad \dfrac{\pi_1}{\alpha};$$

analogously in the case of implication:

(23) $$\dfrac{\dfrac{\dfrac{\dfrac{\alpha}{\pi_1}}{\beta}}{\alpha\rightarrow\beta}\rightarrow I \qquad \dfrac{\pi_2}{\alpha}}{\beta}\rightarrow E \qquad\Longmapsto\qquad \dfrac{\dfrac{\dfrac{\pi_2}{\alpha}}{\pi_1}}{\beta}.$$

A priori it is possible that the elimination of a redundancy creates another one; as a matter of fact it turns out, as Prawitz proved in Prawitz (1965), not only that for every provable formula α of any one of the quoted systems *there is* a derivation of α *in normal form*, i.e. completely free from redundancies (*theorem of normal form*), but also that every derivation can be transformed, by means of an effective *reduction* *procedure*, into another one in normal form (*normalization theorem*). Two even stronger results can indeed be proved: the *strong normalization theorem*, according to which the sequence of reduction steps effectively transforming a derivation into a normal one is finite; and the *uniqueness theorem*, according to which different applications of the reduction procedure yield a unique normal form.

Another relevant result in (general) proof theory is the discovery of an important correspondence—*Curry-Howard isomorphism*—between the proofs of the natural deduction calculus formalizing intuitionistic logic **Int** and the terms of a typed λ-calculus **K**. The terms of such a calculus are assigned to types and denote objects of the corresponding types. For example, if u and v are terms of types U and T denoting the objects $\underline{u}$ and $\underline{v}$, respectively, then <u, v> (the ordered pair of u and v) is a term of type U × V and denotes the ordered pair <$\underline{u}$, $\underline{v}$>; conversely, if t is a term of type U × V denoting <$\underline{u}$, $\underline{v}$>, then p_1t and p_2t are terms of types U and T,[31] denoting $\underline{u}$ and $\underline{v}$, respectively. To make another example, if v is a term of type V denoting $\underline{v}$ and x is a variable of type U, then λ xv is a term of type U → V denoting the function that to any object $\underline{u}$ of type U associates the object $\underline{v}[\underline{u}/x]$[32]; conversely, if t is a term of type U → V denoting the function $\underline{t}$ and u is a term of type U denoting $\underline{u}$, then t(u) is a term of type V denoting the result of the application of $\underline{t}$ to $\underline{u}$. The use of terms of the λ-calculus is governed by *conversion rules*. The rules for the examples of terms just given are the following: p_1<u, v> converts into u; p_2<u, v> converts into v; (λ×v)u converts into v[u/x]. A term t *reduces* to a term u if u can be obtained from t by finitely many applications of the conversion rules. A term is in *normal form* when it has no reducible subterm.

Curry-Howard isomorphism is a one-to-one correspondence between the proofs of the natural deduction formalization of **Int** and the terms of the λ-calculus such that the structure imposed onto the class of terms of the λ-calculus by the relation of reducibility is the same as the structure imposed onto the class of proofs of the natural deduction system by the reduction operations that permit their normalization. For example, the fact that the proof on the left of (22) reduces to the proof on the right corresponds to the fact that the term p_1 <u, v> converts into u; the fact that the proof on the left of (23) reduces to the proof on the right corresponds to the fact that the term (λ×v)u converts into v[u/x]. As a consequence, the process of reduction of a proof of the natural deduction system to its normal form can be conceived as the process of computation of a term of the λ-calculus.

[31] Usually p_1t and p_{21}t are are called "left projection" and "right projection" of t, respectively.

[32] $\underline{v}[\underline{u}/x]$ is the denotation of the term v[u/x], i.e. of the term obtained from v by replacing x with u (with the usual restrictions).

2.3.2 *Theory of Meaning*

The keystone of the parallelism between proof theory and neo-verificationist theory of meaning of the logical constants is the inversion principle. In this section I shall illustrate its role in the theory of meaning.

We have seen above that natural deduction rules can be seen as the rules for the use of the logical constants in deductive practice. It is not difficult to see a deep analogy between (the second) Wittgenstein's slogan that meaning is use,[33] Gentzen's passage from Gentzen (1935) quoted above and an interesting passage from the almost contemporaneous Carnap (1934: XV). Leaving some important differences aside, the analogy consists in the idea that it is legitimate and interesting to give up «the old superstitious view that an expression must have some independently determined meaning before we can discover whether inferences involving it are valid or invalid»,[34] and to look at the issue from the opposite side: state any inference rule for an expression, and it will determine a meaning, in accordance with Wittgenstein's slogan that meaning is use.

There is however a crucial difference between Wittgenstein and Carnap, on the one side, and Gentzen on the other: according to the former *all* rules are meaning determining, according to the latter only introductions are, while eliminations are 'consequences' of the introductory definitions. While Carnap's approach may be seen as one of the roots of inferential-role semantics, it is no exaggeration to say that Gentzen's idea is the main source of much of contemporary neo-verificationism and of proof-theoretic semantics.[35]

What is at stake emerges perhaps in the best way when one takes into consideration a basic difficulty the Wittgensteininan idea of meaning as use seems to run into from the outset. According to it the rules for the use of an expression, in particular of a logical constant, are *constitutive* of the meaning of that expression; for instance, the meaning of $\wedge$ is constituted by the rules $\wedge I$, $\wedge E_1$ and $\wedge E_2$ mentioned above. In his paper "The runabout inference ticket" Arthur Prior criticizes this idea by introducing into our language the constant *tonk* and defining its meaning by the following rules:

(24) $\texttt{tonk-I:}\ \dfrac{\alpha}{\alpha \texttt{tonk} \beta} \qquad \texttt{tonk-E:}\ \dfrac{\alpha \texttt{tonk} \beta}{\beta}$

By the transitivity of the relation $\vdash$ of deducibility it follows that $\alpha \vdash \beta$, for every α and β—a rather uncomfortable outcome against which we cannot legitimately oppose that such a constant doesn't exist, if we have abandoned «the old superstitious view» mentioned above.

---

[33] «For a *large* class of cases—though not for all—in which we employ the word "meaning" it can be defined thus: the meaning of a word is its use in the language.» (Wittgenstein, 1958: Sect. 43).

[34] I am quoting from the famous paper (Prior, 1960: 38), which I shall discuss in detail in a moment.

[35] See for instance Dummett (1991b), Prawitz (1971), Martin-Löf (1985). For a concise but illuminating evaluation of inferential-role and proof-theoretic semantics see Prawitz (2015), Sect. 3.4.

However, according to many Prior's objection is not destructive; N. Belnap has proposed an analysis of it that suggests a way out:

> We are not defining our connectives *ab initio*, but rather in terms of an *antecedently given context of deducibility*, concerning which we have some definite notions. By that I mean that before arriving at the problem of characterizing connectives, we have already made some assumptions about the nature of deducibility. That this is so can be seen immediately by observing Prior's use of the transitivity of deducibility in order to secure his ingenious result. But if we note that we already *have* some assumptions about the context of deducibility within which we are operating, it becomes apparent that by a too careless use of definitions, it is possible to create a situation in which we are forced to say things inconsistent with those assumptions. (Belnap, 1962: 131)

From this point of view it is natural to require that the definition of a new logical constant is consistent with the properties deducibility had before its introduction. I shall not state Belnap's formulation of the consistency requirement[36]; I remark instead that Gentzen's passage quoted above contains an idea which offers, *in nuce*, a solution to Prior's problem different from Belnap's one, and much more articulated.[37] I will describe it through different steps.

(25) Introductions and eliminations, which taken together can be equated to rules for the correct use of the logical constants in an argument, play two essentially different roles: the former are rules for the assertion of sentences, the latter rules for drawing consequences from the assertion of those sentence.

(26) The meaning of a sentence is equated (not with its use conditions in general, as in Wittgenstein's slogan, but) with its *assertibility* conditions.

(27) As a consequence of (25) and (26), introduction rules are meaning-giving; it is in this sense that they can be seen—according to Gentzen's words—as «the 'definitions' of the symbols concerned».

The key idea of the neo-verificationist theory of meaning emerges at this point: to 'generalize' the inversion principle, in the sense of formulating it as the following requirement to be imposed onto any system of rules for the use of the logical constants:

(28) «the conclusions that can be rightly inferred from a sentence must be only such ones as are guaranteed to hold when the conditions for asserting the sentence are satisfied.» (Prawitz, 1980: 6)

In other terms, the idea is that the same relation existing between introductions and eliminations in the closed natural deduction system formalizing **Int** (i.e. in a system with a fixed set of introduction and elimination rules) is now required to hold between introductions and an open set of other rules used to draw consequences from the assertion of sentences (i.e. in an *open* system of natural deduction rules). From this point of view, (28), often called the requirement of *harmony* between introduction

---

[36] But see Footnote 67.

[37] It should be stressed that Gentzen formulated his idea much before Prior stated his problem, and that the relevance of the former to the latter was firstly noticed only much later by Dummett in Dummett (1973) and (1975).

rules and other rules, determines the structure of the *justification* of the rules to draw consequences from sentences: a rule with principal premiss α and conclusion β is justified if its use to infer β can be dispensed with under the assumption that α has been inferred by means of an introduction rule. It is important to stress the role of this last assumption in the architecture of the neo-verificationist theory of meaning; Dummett calls it, in Dummett (1991a, 1991b), the *Fundamental Assumption*:

(29) If a sentence can be asserted, it can be inferred by means of an introduction rule.

We can now see that *tonk*'s elimination rule blatantly violates the harmony requirement: if the premiss α*tonk*β of a *tonk*-E is obtained by a *tonk*-I, the premiss of the *tonk*-I was α, which in general has no relation to the conclusion β of *tonk*-E. On the one hand, this is a way out of Prior's objection, since a natural deduction system that meets (28) is normalizable, as we have seen, and a normalizable system is consistent[38]; hence a system meeting (28) meets Belnap's consistency requirement as well. On the other hand, the harmony requirement does much more than warranting consistency of a system of rules: it *justifies* each rule, in the sense described.

It is not difficult to extract from these ideas, mixed with some intuitionistic ideas, a *neo-verificationist program* for a theory of meaning, remarkably different both from the verificationist theory of meaning of the neo-positivists[39] and from the intuitionistic one.

Some components of this program are borrowed from what might be called "the Fregean paradigm" (Casalegno 1992), and are independent of the realism/anti-realism debate and from the meaning-as-use idea. Here is a list of some of them, together with some brief commentary.

(30) *Articulation*

a theory of meaning should be articulated into a theory of reference, a theory of sense, a theory of force.[40]

(31) *Molecularity:* The smallest expressions which can be understood independently of others are sentences.[41]

(32) *Compositionality:* The meaning of a sentence is determined by its structure and the meanings of its constituents.[42]

The principle of compositionality is meant in the sense that also the grasp of the meaning of a sentence is determined by the grasp of its structure and the grasp of the meanings of its constituents.

[38] See Prawitz (1965: 44).

[39] For a clear exposition of the main differences see Cozzo (1994), Chap. 3, Sect. IV.

[40] A precise characterization of these articulations and a detailed justification of their necessity is contained in Dummett (1976).

[41] Molecularim goes in hand with Frege's *Context Principle*: «Only in the context of a sentence do the words mean anything.» (Frege, 1884: §62).

[42] For different versions of this principle, and for motivations undelying it, see Szabó (2020).

What (31) and (32) together exclude is that the grasp of the meaning of a sentence α depends on the grasp of the meaning sentences of which α is a constituent, and more generally on the grasp of the meanings of sentences logically more complex than α. If this were not the case we might have *holistic* consequences, according to which it is possible in principle that the grasp of the meaning of a sentence α depends on the knowledge of the whole language α belongs to, with the consequent impossibility of developing a systematic theory of meaning for such a language.

However, (31) and (32) together do *not* exclude is that the grasp of the meanings of constituents of a sentence α depends in turn on the grasp of the meanings of other sentences $\alpha_1, \alpha_2, \ldots$ provided that $\alpha_1, \alpha_2, \ldots$ are not logically more complex than α. For example, the grasp of the meaning of a predicate occurring in an atomic sentence α may depend on the grasp of the meanings of other sentences $\alpha_1, \alpha_2, \ldots$, provided that $\alpha_1, \alpha_2, \ldots$ are atomic.

Two important consequences of the insertion, by the neo-verificationists, of Gentzen's ideas about meaning into an intuitionistic framework are, first, a new way of conceiving intuitionistic proofs, which are now seen as sequences of inferential steps consisting in the application of introductions, eliminations and possibily other rules, hence as *inferential* proofs, also called *arguments*; second, not every argument will be equated with a proof: as the point (28) above makes clear, only justified or *valid* arguments will be, provided that valid arguments can be defined in general. In this sense, the a definition of the notion of validity for arguments is the core of a neo-verificationist theory of meaning.

Before giving this definition a further point must be clarified. As have seen in Sect. 2.1.2, the proofs defined by Heyting are *evidential* proofs; but now we see that the neo-verificationists look at intuitionistic proofs as *inferential* proofs. From this standpoint Heyting's definition becomes unsatisfying. According to Heyting, a proof of $\alpha \vee \beta$ is either a proof of α or a proof of β. However, even in the mathematical domain this clause is too restrictive if it is understood as concerning inferential proofs: a primality test for n,[43] where n is some very large number, is an example of a(n inferential) proof of Prime(n)∨¬(Prime(n),[44] hence of a sentence of the form $\alpha \vee \neg\alpha$, which is neither a proof of α nor a proof of ¬α. An analogous objection can be raised against Heyting's clause for ∃.[45] Another objection to the same clauses is that

> a proof of a disjunction A(t)∨B(t) may very well proceed even intuitionistically by first proving ∀x(A(x)∨B(x)) and then applying universal instantiation to infer A(t)∨B(t). Given

---

[43] I.e. an algorithm for determining whether n is prime.

[44] Such a test should not be confused with a general method M consisting in applying to every number x a test for x; M is a proof of ∀x(Prime(x) ∨¬Prime(x)).

[45] See for example Dummett (1975: 239–241) and Dummett (2000: 271). Dummett formulates at that place also an objection to Heyting's clause for →, considered by him as more serious than the objections concerning ∨ and ∃. I shall introduce and discuss it in Sect. 2.4.2. It should be retorted to these objections, and to the subsequent one, that Heyting was defining not the notion of inferential proof, but of evidential proof; presumably, the possibility of understanding Heyting's proofs as inferential proofs, offered by the adoption of Gentzen's ideas, was so attractive to neo-verificationists to induce them to ignore that distinction.

> such a proof, we do not know which of the two disjuncts holds. Hence, it is not correct to say that a proof of a disjunct needs to consist of a proof of one of the disjuncts together with indication of which disjunct is proved. In intuitionistic meaning explanations of the kind exemplified above, proof must thus be meant in a quite restrictive way. (Prawitz, 2005: 684)

It is therefore natural, from the neo-verificationist point of view, to introduce a fundamental distinction between *canonical* proofs and proofs *tout court*, based on the idea that the former are given by arguments whose last inference is the application of an introduction rule and such that the arguments for the premisses of the last inference are (presentations of) proofs of the premisses; the latter must be conceivable as effective methods for costructing canonical proofs. The central notion is therefore the former, since the latter is intended as defined in terms of it. Since the premisses of the last, introductory, inference are always subformulas of the conclusion, the molecularity requirement is met, and the explanation of the meaning of the logical constants in terms of canonical proofs can consist in an inductive definition.

However, even after the introduction of the distinction between canonical and non-canonical proofs a difficulty remains concerning the molecularity requirement. It has to do with implication: according to the idea just sketched, a canonical proof of $\alpha \rightarrow \beta$ is given by an argument whose last inference is of the form

(33) $$\dfrac{\begin{matrix}[\alpha]\\ \pi\\ \beta\end{matrix}}{\alpha\rightarrow\beta}\ ,$$

where π is a proof of β from α as assumption: it is not required (nor could it be required) that π is in turn canonical; as a consequence it is possible that it contains formulas that are not subformulas of $\alpha \rightarrow \beta$, of an arbitrary complexity; the explanation of the meaning of $\rightarrow$ risks therefore not to be inductive.[46]

The difficulty had already been stressed by Gentzen:

> In interpreting A→B in this way, I have presupposed that the available proof of B from the assumption A contains merely inferences *already recognized as permissible*. On the other hand, such a proof may itself contain other →-inferences and then our interpretation *breaks down*. For, it is *circular* to justify the →-inferences on the basis of an →-interpretation which itself already involves the presupposition of the admissibility of *the same* form of inference. (Gentzen, 1936: 167)

As a matter of fact, what is presupposed in the statement of the premiss of a $\rightarrow$ -introduction is only that a proof of β from the assumption α is available; that such a proof contains merely inferences already recognized as permissible is also presupposed only if it is assumed that what constitutes a proof of β from α is to be explained by making reference to the permissibility of each inference occurring in the argument. Precisely this assumption, made both by Gentzen and Dummett, Prawitz is prepared to give up in order to meet the difficulty.

---

[46] Under the assumption that proofs are to be conceived as inferential, i.e. as sequences of inferentisl steps, this problem is essentially the same as the one, discussed above, of the impredicativity of the clause for implication in Heyting's Explanation. If that assumption is given up and proofs are conceived as evidential (as Heyting did, in my opinion), the problem vanishes.

2.3.3 Prawitz's Notion of Valid Argument

Proofs are given by valid arguments. An argument for α is a collection of inferences in tree form with α as endformula and with a specification of which formulas are (free) assumptions and at which places assumptions and (free) parameters are *bound* by some inferences. If all assumptions and inferences in an argument are bound the argument is *closed*, otherwise it is *open*. An open argument is to be thought as a schema for obtaining closed arguments (its *instances*) by replacing free parameters with closed terms and free assumptions with closed arguments.[47]

There is an obvious analogy between arguments and derivations in a natural deduction system; but I remind the reader that the two notions are of a different nature: derivation is a formal notion, argument (hence also proof in this inferential sense) an intuitive or informal one.[48] In particular, it is not assumed that a specific set of inference rules is given; all that is required is that the meaning of the sentences occurring in an argument is *defined* by canonical arguments for them, and that non-canonical inferences occurring in a non-canonical argument are *justified* by certain procedures, which will be illustrate after the definition.

Prawitz's strategy consists essentially in: (i) defining the notions of valid argument (v.a.) and canonical argument (c.a.) by *simultaneous* induction, and (ii) relativizing both notions to a set J of *justifying procedures*. I give only the clauses for ∧, → and ∀.

(34) **Definition.**[49]

1. A J-c.a. for $\alpha\wedge\beta$ is an argument whose last inference is of the form

$$\frac{\begin{array}{cc}\pi_1 & \pi_2\\ \alpha & \beta\end{array}}{\alpha\wedge\beta},$$

where π_1 and π_2 are J-v.a.'s for α and β, respectively.

2. A J-c.a. for $\alpha \rightarrow \beta$ is an argument whose last inference is of the form

$$\frac{\begin{array}{c}[\alpha]\\ \pi\\ \beta\end{array}}{\alpha\rightarrow\beta},$$

where π is an open J-v.a. for β from α as assumption.

[47] Note that it is *informal* arguments—as opposed to formal arguments or *derivations* - Prawitz is here speaking of, hence also *informal* proofs will be characterized through the notion of valid argument. About the relations between formal and informal notions see Prawitz (1985), §3.3.

[48] The difference is substantive, owing (also) to Gödel's first theorem: of an undecidable sentence G there is no derivation in **PA**, for instance, but there is a valid argument with G as end formula.

[49] Prawitz never gave explicitly this definition; it seems however to conform to the spirit of his approach, allowing to expound it very succinctly.

3. A J-c.a. for $\forall x\alpha$ is an argument whose last inference is of the form

$$\frac{\begin{matrix}\pi[a]\\ \alpha(a)\end{matrix}}{\forall x\alpha},$$

where $\pi[a]$ is an open J-v.a. for $\alpha(a)$.

4. A closed argument for α is J-valid iff the composition of the procedures in J yields an effective method for finding a J-c.a. for α.
5. An open argument for α is J-valid iff all its closed instances obtained by replacing free parameters with closed terms and free assumptions with closed J-v.a.'s for the assumptions are J′-valid, for every extension J′ of J.

Some remarks about this definition.

Concerning Gentzen's difficulty, notice that the immediate subargument of a canonical argument for $\alpha \rightarrow \beta$ is not required to be canonical; as a consequence, the open argument for β from α as assumption may contain inferences involving sentences of greater complexity than both α and β. But this is no longer a problem as far as the composition of the justifying procedures associated to each step of the argument can be seen as a method for transforming every valid argument for α into a valid argument for β, since such a method is defined exclusively with reference to arguments for subformulas of $\alpha \rightarrow \beta$. The key idea for meeting the difficulty is that the validity of an argument is not to be defined stepwise, by reference to the validity of each individual inference constituting it, but by the global property of the procedures in J of providing a method for finding a canonical proof of the conclusion. The order of conceptual priority is reversed with respect to what might seem the most natural strategy: instead of defining as valid an argument when it is composed from valid inferences, «the validity of an inference rule is explained in terms of validity of arguments» (Prawitz, 1985: 169).[50] Let us see how.

A closed canonical argument is valid as such. Since canonical arguments are of introductory form, introduction rules are self-justifying; and this is plausible on account of the fact that, according to Gentzen's suggestion, introduction rules are meaning-giving. The other rules have to be justified. How? As we have seen above, Prawitz's idea (adopted by Dummett, 1991a, 1991b) is that a rule is justified when, through certain justifying operations, he show how its conclusion might be obtained (under the fundamental assumption) *without using the rule*. In the case of Gentzen's elimination rules, the justifying operations are the *reductions* used by Prawitz (1965) in the proof of the normalization theorem.[51] Consider for instance implication: its elimination rule is justified, under the (fundamental) assumption that its premisses are obtained by introductions, by showing how its conclusion might be obtained without using the rule, i.e. by showing that

---

[50] It is precisely the 'unnaturalness' of this inversion of conceptual priorities that will be at the origin of Prawitz's more recent Theory of Grounds. See Sect. 2.4.3.1.

[51] In fact, justifying procedures aren't but a generalization of reduction procedures.

(35)
$$\dfrac{\dfrac{\dfrac{\dfrac{\alpha}{\pi_1}}{\beta}}{\alpha\to\beta} \quad \dfrac{\pi_2}{\alpha}}{\beta} \quad \texttt{reduces to} \quad \dfrac{\dfrac{\dfrac{\pi_2}{\alpha}}{\pi_1}}{\beta}$$

Clause 5. of the definition makes reference to the extensions of the set J. The reason for this is that otherwise there might be an (open) argument for β from α, valid relative to a set J of justifying procedures but not to some extension J′ of J: it would be sufficient that J′ contained new justifying procedures from which methods could be constructed for finding valid arguments for α not capable to be transformed (by the procedures in J) into methods for finding valid arguments for β. Notice however that, although Prawitz's manoeuvre solves this difficulty, it renders the definition highly impredicative in character.

2.3.4 Martin-Löf's Theory of Meaning

A systematic exploitation of Curry-Howard's isomorphism is represented by Martin-Löf's intuitionistic theory of types, whose most evident peculiarity is that its language is different from usual first order languages.[52] I shall not present the theory in detail; an idea of its language can be got from the language of the λ-calculus **K**. It contains terms for *objects* a, b, c,… (including, in particular, proofs), and for the *types* α, β, γ,… of those objects; among the types there are propositions, in accordance with the propositions-as-types idea,[53] equating a proposition with the set of its proofs.[54] The sentences of the language express what Martin-Löf calls *judgements*.[55] A distinctive feature of judgements, as opposed to propositions, is that of them it is senseless to say that they are false, but only that they are true or derivable from some assumptions; as a consequence they cannot be combined by means of the logical constants. Fundamental forms of judgement are the following:

(36) α:CAT (α is a category)[56]
(37) α = β:CAT (α e β are the same category)
(38) a:α (a is an object of the category α)

[52] In the presentation of the theory I will make reference to its version contained in Martin-Löf (1987).

[53] The point (41) below.

[54] Beware: the set of *its* proofs, not of the sentence expressing it: propositions are not expressed by sentences, but only denoted by terms.

[55] The distinction between propositions and judgements plays an essential role in the theory of meaning of Martin-Löf. To a critique of the rationale for it is devoted the paper Martino and Usberti (1991).

[56] A variant is α: TYPE (α is a type).

(39) a = b:α (a and b are the same α)

To obtain specific judgements it is sufficient to replace "α" and "β" with names of specific categories; for instance, the judgement

(40) SET:CAT.

says that sets are a category.

A category α is known, i.e. a judgement of the form (36) may be asserted, when (i) we know an *application criterion* of the category, i.e. we know what an object of α is, and (ii) we know an *identity criterion* of the objects of α, i.e. we know what it means for two objects of α to be the same. We know that two categories are the same, i.e. a judgement of the form (37) can be asserted, if every object of the former is an object of the latter, and identical objects of the former are identical objects of the latter. As a consequence, if we know a category α and we are given an object a (or two objects a and b), we are capable to establish whether a judgement of the form (36) [or (37)] is assertible or not. In this sense the category (the universal) precedes the object.

The application criterion will be of course different for different categories. In the case of the category of *sets* we must explain how the elements of a set are generated by means of 'introduction' rules, hence what its *canonical* elements are, and explain when two canonical elements of the set are equal elements.[57] The identity criterion is that α and β are identical sets when the canonical elements of the one are the canonical elements of the other, and equal canonical elements of the one are equal canonical elements of the other.

Applying the propositions-as-types idea, we can define the propositions as follows:

(41) PROP = SET:CAT,

namely the category of propositions is identical to the category of sets: every proposition is seen as a set (the set of its constructions), and every set is seen as a proposition (the proposition declaring that the set has an element).

The logical constants operate on propositions to yield new propositions.[58] Their meaning is explained through rules of four kinds: formation, introduction, elimination, and definitional identity. As an example, here are the rules for implication[59] followed, when necessary, by a short explanation:

---

[57] Notice that if we can characterize the canonical elements of a set we can also characterize all its elements, because a non-canonical element of a set α can be defined in general as a *method* (a primitive notion) that, whenever it is applied, yields a canonical element of α.

[58] As a matter of fact the logical constants are not primitive in the theory of types, but are defined in terms of more basic operations of a set theoretic nature. An explanation of this fundamental aspect of the theory of types is beyond the limits of the present exposition.

[59] For reasons of intuitive legibility I have somewhat simplified, and partially modified, the formulation of the rules contained in Martin-Löf (1987: 127–147); the actual formulation still permits to understand the essential aspects of Martin-Löf's theory of meaning.

(42) $\rightarrow$F $\dfrac{\alpha\texttt{:PROP} \quad \beta\texttt{:PROP}}{\alpha\rightarrow\beta\texttt{:PROP}}$ $\dfrac{\alpha=\gamma\texttt{:PROP} \quad \beta=\delta\texttt{:PROP}}{\alpha\rightarrow\beta=\gamma\rightarrow\delta\texttt{:PROP}}$

(43) $\rightarrow$I $\dfrac{\begin{matrix}[\texttt{x}:\alpha]\\ \texttt{b}:\beta\end{matrix}}{\lambda\texttt{xb}:\alpha\rightarrow\beta}$

Explanation: the rule defines the meaning of → by explaining how canonical elements, or *verifications*, of α → β are formed: given a function b from constructions for α to constructions for β, a verification of α → β is the result of abstraction on the variable x occurring in b.[60]

(44) $\rightarrow$E $\dfrac{\texttt{c}:\alpha\rightarrow\beta \quad \texttt{a}:\alpha}{\texttt{Ap(c,a)}:\beta}$

Explanation: if we have a proof c of α → β (i.e. a function from proofs of α to proofs of β) and we combine it, by means of the operation denoted by "Ap", with a proof of α, the result is a proof of β. Notice that elimination rules can be conceived as formation rules of propositions formed by means of the corresponding 'eliminative constants' (Ap, in the present case), whose meaning is explained by rules of the last kind.[61]

(45) $\rightarrow$DI $\dfrac{\texttt{a}:\alpha \quad \texttt{b}:(\alpha)\beta}{\texttt{Ap}(\lambda\texttt{xb,a})=\texttt{b(a)}:\beta}$

Explanation: if we have a function from proposition α to proposition β and we apply it to a construction for α, the result is the same construction for β we obtain by combining, by means of the operation denoted by "Ap", the corresponding verification of α → β with a. On the one hand this corresponds to the reduction operation used by Prawitz to prove the normalization theorem, on the other hand it means that the meaning of the 'eliminative constant' Ap is just functional application.

The reader will have noticed that introduction and elimination rules are different from Gentzen's ones; however, it is easy to obtain (a version of) Gentzen's rules by means of a 'proof suppression' mechanism consisting in replacing every judgement of the form a:α with the judgement α TRUE:

(46) $\rightarrow\texttt{I}_\texttt{G}$ $\dfrac{\begin{matrix}[\alpha\ \texttt{TRUE}]\\ \beta\ \texttt{TRUE}\end{matrix}}{\alpha\rightarrow\beta\ \texttt{TRUE}}$

(47) $\rightarrow\texttt{E}_\texttt{G}$ $\dfrac{\alpha\rightarrow\beta\ \texttt{TRUE} \quad \alpha\ \texttt{TRUE}}{\beta\ \texttt{TRUE}}$

60 In the original formulation Martin-Löf does not use λ but an ad hoc 'introductory constant'.

61 In the original formulation Martin-Löf uses an ad hoc 'eliminative constant'; here I use "Ap" because its meaning is the same.

Intuitively, the judgement form α TRUE means, in accordance with the verificationist interpretation of the notion of proof, that a construction for α exists. Martin-Löf remarks that the notion of existence invoked here is not the one usually expressed by means of the existential quantifier, and characterizes the judgement α TRUE as an *incomplete* judgement, or as an *abbreviated* way of saying that a certain object is a construction for α; therefore it is not to be understood as a judgement of a new kind.

All the judgements of the fundamental kinds are seen by Martin-Löf as examples of *analytic judgements*, where «to prove an analytic judgement it suffices to analyse the meanings of the terms which enter into it.» (Martin-Löf, 1987: 187). On the contrary, the judgement α TRUE (which, as we know, is not fundamental) is a typical example (the only one, in the theory of types) of *synthetic judgement*, since in order to prove it «you must construct the proof object which has been suppressed in it.» (*Ibid.*) It is therefore the proof suppression mechanism that is responsible of the appearance of synthetic judgements.

One of the merits of the theory of types is that the logic of analytic judgements is complete and decidable. In virtue of completeness it cannot happen, for instance, that α is a proposition, or that a is a contruction for α, and that the corresponging judgements are not assertible; in virtue of decidability we are capable to decide, for instance, whether a given object is or not a proposition, and whether, given an object a and a proposition α, a is or is not a construction for α. This last point, the decidability of the relation "c is a construction for α", is of crucial importance for the two questions to be discussed in the next section.

2.3.5 Knowledge of Meaning, Knowledge of Proofs

We have seen that Heyting's notion of evidential proof is intended to meet the requirement (16) of epistemic transparency, because if it did not, then the fundamental idea of Heyting's theory of meaning—that knowing the meaning of α amounts to being in a position to recognize a proof of α—should be abandoned. In this section I shall describe the attitude of three neo-verificationists—Dummett, Prawitz and Martin-Löf—about two questions of decisive importance from the point of view of transparency: what does knowledge of meaning consist in? And what does knowledge of a valid argument amount to?

As I noted in Chap. 1, during several years Dummett has elaborated a sort of 'anti-realistic argument' whose core idea is that if knowlege of the meaning of α is equated with knowledge of the classical truth-conditions of α, then in some cases such knowledge cannot be observable. As a matter of fact, Dummett's original objection was that in some cases knowledge of meaning cannot be *manifestable*, where the condition of manifestability is like the condition of observability stated at point (12) of Chap. 1, except that the conditional $\rightarrow$ is replaced by the biconditional $\longleftrightarrow$.[62]

[62] The weaker condition of observability was suggested by Prawitz (1977, 1978, 1980).

Leaving the plausibility of Dummett's requirement aside,[63] it is clear that, according to him, an adequate theory of meaning equating knowledge of the meaning of α to knowledge of what a proof of α is should enable us to decide whether a given construction is a proof of α, for only on this condition would the manifestability requirement be satisfied.

Since, in a neo-verificationist theory of meaning, mathematical truth is equated to existence of a proof, and proofs are given by valid arguments, it is obvious that knowing a valid argument cannot amount to knowing an argument that simply is *in fact* valid: it is also necessary that it is *known to be* valid.

But is it plausible to require that the notion of valid argument for α is decidable? To this question I have not found an explicit answer in Dummett, except for his sympathetic attitude towards the intuitionists' idea that proofs are epistemically transparent (see point (15) above).

Prawitz has repeatedly argued that the notion "π is a valid argument for α" cannot be plausibly proposed as decidable. Consider for instance a J-canonical argument for $\alpha \rightarrow \beta$. There are two distinct reasons that make highly implausible the assumption that the J-canonicity of π is decidable. Firstly, π is J-valid if and only if all its closed instances are J′-valid, for all the extensions J′ of J; and it is not clear how the totality of the extensions of J could be regimented. One might think about modifying the explanation of implication by requiring, as a second clause, that a J-canonical argument for $\alpha \rightarrow \beta$ contains also a J-valid argument π' for the statement asserting that π is J-valid, as suggested by Kreisel. But the decidability of the J-validity of π' could not be warranted without starting an infinite regress.

Secondly,[64] consider a closed instance of π: it is valid if and only if the composition of the procedures in J yields an effective method to transform every J-canonical argument for α into a J-canonical argument for β; but we know from Gödel's theorem that the totality of effective methods cannot be generated by any formal system, so the assumption that an effective method with the required property is decidable seems hardly plausible.

The moral Prawitz (1977) draws from these remarks is that the requirement that the relation "c is a proof of α" is decidable cannot be met, and that knowledge of the meaning of α should therefore be equated not with satisfying it, but with knowing the assertibility conditions of α. However, a necessary condition for knowing the assertibility conditions of α is knowing that α is assertible when a proof of α is known; and about the question of what knowledge of a proof of α is (in particular how it is different from the capacity to decide whether something is a proof of α) no suggestion is given.

[63] For a discussion of this requirement, an exposition and an evaluation of Dummett's 'anti-realistic argument' see Usberti (1995), Chap. 4.

[64] Cp. Prawitz (1977), Sect. III.3.

A different explanation of the validity of inference has been developed by Prawitz since the second half of 2000s with his Theory of Grounds (ToG)[65]; I shall discuss its impact on the question of knowledge of proofs, and in particular of decidability of the relation "c is a construction for α", in the following Sect. 2.4.3.1.

Martin-Löf's theory of types seems capable to suggest a way out of the difficulties pointed out by Prawitz concerning the decidability of the relation "c is a proof of α". The language of the theory allows to express that relation, but "c is a proof of α" is not a proposition, as we have seen: it is a judgement. While proofs of propositions are mathematical *objects*, which make propositions *true*, proofs of judgements are *acts*: acts of knowledge through which judgements become *evident*.[66] What enables us to recognize that something is a proof(-object) of α is a proof-act, not a proof-object, which would require a new proof of the fact that it has such property, giving rise to the infinite regress described above. This solves the first of the two difficulties pointed out by Prawitz.

Concerning the second, we have seen that, according to Martin-Löf, Gödel's incompleteness theorem concerns the logic of synthetic judgements, not the logic of analytic ones, which is the object of the theory of types. The completeness and decidability of the theory of types constitute therefore an answer to the second difficulty; when two symbols a and α of the theory are given, it is sufficient to analyze their meaning in order to establish whether a is proof of α, where the intuitive notion of analysis of meaning corresponds, within the theory, to a rigorously definable procedure of *type checking*, i.e. of control of wellformedness of the judgement a: α.

2.4 An Assessment

In this section I shall compare intuitionist and neo-verificationist views on some basic issues, and through this comparison I shall try to explain why I hold that the intuitionistic explanation of the logical constants is the view of meaning most congenial to the internalist theory of meaning I am trying to develop.

2.4.1 Tonk's Problem Revisited

The first issue is Prior's problem. As we have seen, the neo-verificationists accept Belnap's diagnosis of the problem, although their solution is different and more articulated. So let us consider this diagnosis more closely. At the beginning of his paper Belnap writes:

---

[65] For a systematic exposition and discussion of ToG see Piccolomini d'Aragona (2019).

[66] The dichotomy between proof-acts and proof-objects, central in the theory of types, has been made explicit and clear in its implications by G. Sundholm in Sundholm (1983).

> A possible moral to be drawn is that connectives cannot be defined in terms of deducibility at all; that, for instance, it is illegitimate to define *and* as that connective such that (1) A-*and*-B⊢A, (2) A-*and*-B⊢B, and (3) A,B⊢A-*and*-B. We must first, so the moral goes, have a notion of what *and* means, independently of the role it plays as premise and as conclusion. Truth tables are one way of specifying this antecedent meaning; this seems to be the moral drawn by J.T. Stevenson. There are good reasons, however, for defending the legitimacy of defining connectives in terms of the roles they play in deductions. (Belnap, 1962: 130)

And he concludes the paper with these words:

> one *can* define connectives in terms of deducibility, but one bears the onus of proving at least consistency (existence) [...]. But it is not necessary to have an antecedent idea of the independent meaning of the connective. (Belnap, 1962: 134)

Belnap seems therefore to share with Stevenson and many subsequent commentators the opinion that there is a basic difference between 'inferential' and 'tabular' definitions from the point of view which concerns us here: while the former are prone to Prior's disease, the latter are sheltered from it insofar as they are mere ways of specifying the 'antecedent' meaning of connectives. Second, Belnap thinks that also inferential definitions are legitimate, and we have seen above his argument for this claim. I think that both these claims are unjustified.

First, it would surely be incorrect to ascribe to Prior himself the opinion that tabular definitions are immune from *tonk*'s disease. On the one hand, he explicitly says: «Unlike Stevenson, I can see no difference in principle between these devices ['inferential definitions' and truth tables]» (Prior, 1964: 192). On the other hand, while in Prior (1960) he had given an inferential definition of *tonk*, in Prior (1964) he gives a tabular definition: P-*tonk*-Q «is true if P is true and false if Q is false (and therefore, of course, both true and false if P is true and Q false)» (Prior, 1964: 191). He did not mean to contrast two styles of definition at all; on the contrary, both of them are exposed to the same objections. One might disagree with Prior's objection to tabular definitions: a truth-table is required to describe a function, but a truth-table according to which the same sentence is both true and false does not describe a function. However, as we will see in a moment, in that article Prior raises other objections, concerning 'uniqueness', and these apply to both styles of definition.

Second, both inferential and tabular definitions of the meanings of logical connectives are not definitions in the strict sense; an actual definition must be non-creative and eliminable[67]; but the definition of *and*, for example, is non-eliminable: there is not a formula of L which is logically equivalent to α *and* β; the same holds for all

[67] See Suppes (1957: 153) and Tarski (1983: 307, fn. 3). If $\mathcal{T}$ is a set of axioms in the language $\mathcal{L}$ and R is a symbol not in $\mathcal{L}$, a definition of R over $\mathcal{T}$ is a sentence α[R] which contains the symbol R, and whose other symbols all come from $\mathcal{L}$. That α[R] is *non-creative* means that if β is a sentence of $\mathcal{L}$ such that $\mathcal{T} \cup \{\alpha[R]\} \vDash \beta$, then $\mathcal{T} \vDash \beta$; that α[R] is *eliminable* means that if $\gamma[x_1, \ldots, x_n]$ is any formula whose symbols apart from R all come from $\mathcal{L}$, then there is a formula $\gamma'[x_1, \ldots, x_n]$ of $\mathcal{L}$ such that $\mathcal{T} \cup \{\alpha[R]\} \vDash \forall x_1 \ldots \forall x_n (\gamma \leftrightarrow \gamma')$. Notice that α[R] is non-creative iff the extension $\mathcal{T} \cup \{\alpha[R]\}$ of $\mathcal{T}$ is conservative. Belnap's requirement of "consistency with antecedent assumptions" of a definition is precisely the requirement that the definitional extension is conservative.

connectives which, just for this reason, are sometimes called "primitives". Inferential and tabular definitions have the status of axioms, of which we may say that they 'implicitly define' the meaning of the primitives. Prior's argument concerns therefore the possibility of an axiomatic characterization of the meaning of the logical constants.

Let us consider the uniqueness objections, contained in Prior (1964). Belnap had put forward, in addition to the consistency requirement, a *uniqueness* requirement, according to which, if the logical constant **C** is given a characterization formally identical to that of **C′**, then any formula α having **C** as its principal logical constant must be logically equivalent to the formula obtained from α by replacing **C** with **C′**. Whenever this requirement is met we can say, according to Belnap, that for any propositions α, β there is a *unique* proposition α**C**β, or that **C** is unique. Prior objects that, while Belnap's uniqueness requirement may be a necessary condition of synonymy, it cannot be a sufficient condition. He gives two counterexamples.

First, in a language rich enough for the formulation of the propositional calculus there is actually an infinite number of intuitively non-synonymous connectives with the same truth table. Let us suppose we have defined (by truth tables or by inferential rules) the symbols $\wedge$ and $\neg$; then we can define the following infinite sequence of 'conjunction-forming connectives':

$$\begin{aligned} \alpha \wedge_0 \beta \quad &= \quad \alpha \wedge \beta \\ \alpha \wedge_{n+1} \beta \quad &= \quad \neg(\alpha \wedge_n \neg\beta) \wedge \neg(\neg\alpha \wedge_n \beta) \wedge \neg(\neg\alpha \wedge_n \neg\beta) \end{aligned}$$

All these connectives clearly have the same truth table, but they can hardly be synonymous: «I cannot see—Prior remarks—how the sense of a sentence can ever be identical with a logical complication of itself.» (Prior, 1964: 193).

Second, consider the connective *and*; it has the same truth table and the same inference rules as the connective & which is defined as follows:

$\alpha \& \beta =_{\text{def}} (\alpha \text{ and } \beta) \text{ or } \bot$,

where $\bot$ is the constant for The False. Shall we say that *and* and & are synonymous, so that "Today it's raining and the wind is blowing", for example, is synonymous of "Today it's raining and the wind is blowing or John is a married bachelor"?

A possible answer to these counterexamples is that the intended meaning of $\alpha \wedge \beta$, and of $\alpha\&\beta$, is the *simplest* sense satisfying the specified (inferential or tabular) conditions. But this will not do. Consider for instance the connective *, defined as follows:

$$\alpha^*\beta = \text{def}\neg\neg((\alpha \vee \beta) \wedge \neg(\alpha \wedge \beta));$$

it has the same truth table of the connective

$$\alpha@\beta = \text{def}(\neg\alpha \vee \neg\beta) \wedge (\beta \to \neg\alpha);$$

since their *definientia* have the same logical complexity, the simplest sense satisfying their tabular conditions simply does not exist.[68]

After these clarifications, the logical structure of Prior's argument should emerge more clearly. The thesis he is arguing for is that neither tabular nor inferential axiomatic characterizations of logical connectives can yield their meaning, because each definition

> implies that the sentence formed by placing a [connective] between two other sentences already has a meaning. For only what already has a meaning can be true or false (according as what it means is or is not the case), and only what already has a meaning can be inferred from anything, or have anything inferred from it. (Prior, 1964: 191)

The argument is essentially a *reductio ad absurdum* of the contrary assumption that

(48) Any tabular or inferential axiomatic characterization of any logical connective is sufficient to determine its meaning.

The argument goes on as follows[69]:

(49) Then *tonk* inference rules determine the meaning of a logical constant.
(50) A logic containing *tonk* and $\rightarrow$ is inconsistent.[70]
(51) There is no logical argument against inconsistent logics.

That this is a *reductio* depends on whether (51) is considered absurd or not. I shall not discuss this point here[71]; I only remark that, if (51) is considered absurd (and the argument is valid, as I presently assume) then (49) is false, hence also (48) is false. This seem to be Belnap's position, since he holds that a consistency requirement ought to be imposed onto any logical sytem. Hence, if my reconstruction of Prior's argument is correct, Belnap's position should be described as giving up the very idea

---

[68] See also Prior (1964), 194. Another possible answer has been suggested by G. Harman: the meaning of the logical constants might be characterized by the inferences *immediately* valid for them, where

> Immediate implication is a psychological notion. An immediate implication is one that is immediately obvious, one that can be immediately recognized. (Harman 1986: 131)

As a consequence, $\alpha \wedge \beta$ would not have the same meaning as $\neg(\neg\alpha \vee \neg\beta)$, since the former, but not the latter, immediately implies α. I see at least two objections to this suggestion. First, take $\alpha \wedge \beta$ and $(\alpha \wedge \beta) \wedge \alpha$: arguably *both* immediately imply α, but they have not the same meaning, if we accept Prior's idea that the sense of a sentence cannot be identical with a logical complication of itself. Second, for an implication to be (immediately or mediately) obvious the meanings of the sentences, and of the logical constants, involved must be known, hence such an implication cannot be conceived as *constitutive* of the meaning(s) of the logical constant(s) involved.

[69] The argument is not explicit; the following is a reconstruction of it I submit as plausible.

[70] Here is a justification of (50): Consider a language $\mathcal{L}$ containing implication $\rightarrow$ and *tonk*, and the theory $\mathcal{T}$ with the natural deduction rules associated to them. Then from the axiom $\alpha \vdash \alpha$ we obtain $\vdash \alpha \rightarrow \alpha$; let's abbreviate $\alpha \rightarrow \alpha$ with $\top$; by *tonk*-I we obtain $\vdash \top$ *tonk* β, for any β; by *tonk*-E we conclude that $\vdash \beta$, for any β: $\mathcal{T}$ is inconsistent.

[71] For an argument supporting the absurdity of (51) see Chap. 5.

that an axiomatic characterization of a logical connective is sufficient to determine its meaning, in particular that inference rules are meaning-giving: according to him, they are meaning-giving only if the system obtained by adding them is consistent with prior assumptions.

What has been underestimated is that consistency with prior assumptions, exactly like consistency *tout court*, must be proved in the metatheory, and that the metatheory will contain logical constants, whose meaning must be specified.[72] This question was clearly present to Hilbert, whose views about the existence of mathematical entities were significantly similar to Belnap's position about the existence of meanings of the logical connectives. It may therefore be instructive to put the discussions about it on the background of the debate about the significance of the axiomatic method in mathematics.

Some words about the philosophical background. In his letter to Hilbert of 12.27.1899 Frege expresses himself in a way amazingly similar to Prior's when he writes about elucidatory propositions: «I would not want to count them as part of mathematics itself but refer them to the antechamber, the propaedeutics.» (Frege, 1980: 36)[73] Concerning axioms, in the same letter he writes:

> I call axioms propositions that are true but are not proved because our knowledge of them flows from a source very different from the logical source, a source which might be called spatial intuition. From the truth of the axioms it follows that they do not contradict one another. (Frege, 1980: 37)

It is well known that Hilbert's position on this point was diametrically opposite; concerning the last sentence in the quotation above he remarks:

> I found it very interesting to read this very sentence in your letter, for as long as I have been thinking, writing and lecturing on these things, I have been saying the exact reverse: if the arbitrarily given axioms do not contradict one another with all their consequences, then they are true and the things defined by the axioms exist. This is for me the criterion of truth and existence. (Frege, 1980: 39–40)

Concerning axioms he explains further on: «Every axiom contributes something to the definition, and hence every new axiom changes the concept.» (Frege, 1980: 40).

Intuitionists as well took part in the debate about the axiomatic method. In particular Heyting, who was very sensitive to the merits of the method he himself used repeatedly, took a position very different from Hilbert's concerning the significance of axioms:

> From the intuitionistic point of view [...] the method cannot be used in its creative function. As a mathematical object is only considered to exist after its construction, it cannot be brought into being by a system of axioms. On the other hand, the descriptive function of a system of axioms is as important intuitionistically as it is classically. (Heyting 1961: 239)

[72] This underestimation pops up in Belnap's conclusion that «It is not necessary to have an antecedent idea of the independent meaning of the connective» (Belnap, 1962: 134).

[73] Compare with Prior's view of the 'inferential definition' and the truth-table of a connective as a mere aid, an «indirect and informal way of fixing [its] sense» (Prior, 1964: 192), «a piece of informal pedagogy» (Prior, 1964: 194).

The opinions of Frege, Hilbert and Heyting I have briefly sketched concerned the characterization of mathematical entities and notions. Let us now see whether those authors had analogous opinions about the meaning of the logical constants. I think that the answer is yes in the cases of Frege and Heyting, but no in the case of Hilbert.

As for Frege, he seems to put the senses of the logical constants precisely among the notions which cannot be defined *sensu stricto* but only elucidated, and which are therefore presupposed by the elucidation. In particular, in his paper "Compound thoughts" he repeatedly makes a remark that cannot be explained otherwise. After having characterized the sense of "A and B" and having explained the truth-conditions of this kind of compound thought, he remarks:

> That "B and A" has the same sense as "A and B" we may see without proof by merely being aware of the sense. (Frege, 1984: 393)

and from the context one can evince that he conceives the identity of sense between "A" and "A and A" in the same way. Further on (Frege, 1984: 394), speaking about the interchangeability of A and B, he insists that it should not be regarded as a theorem. Therefore, the sense of "A and B" cannot be *determined* by the truth table for "and", but is presupposed by it. In the same place Frege gives an example of a valid argument: from "A is true" and "B is true" it is legitimate to infer "(A and B) is true". And the validity of such an argument is presented by him as a consequence of the sense of "and" rather than as constitutive of it.

Heyting's position is analogous: the explanations of the meanings of the logical constants he gave in several papers are to be understood as instances of Fregean elucidations, not as formal, 'creative' definitions, nor as a system of axioms.

Hilbert's view is more difficult to grasp, since he never stated explicitly a meaning theory of the logical constants. Nevertheless, several remarks concerning this point are scattered in his writings, and in the handbook Hilbert & Bernays (1934) there is a "Logical characterization of the finitary position" which as a matter of fact is the sketch of a theory. Now, it is interesting that Hilbert does not even take into consideration an axiomatic characterization of the logical constants' finitary meaning; on the contrary, all the remarks he makes concerning this point are of the same kind as Frege's elucidations. Why? For a very good reason, I think: because, even if an axiomatic meaning theory were formulated, its consistency should be proved, and in the metatheory in which such a proof were given we should at any rate need (some) logical constants, whose meanings should be assumed to have been previously elucidated by means of informal explanations.[74]

It seems therefore that Prior's conclusion—that to believe that tabular or inferential axiomatic characterizations of logical connectives «can take us beyond the symbols to their meaning, is to believe in magic» (Prior, 1964: 191)—is justified. In particular, introduction rules cannot constitute the meaning of the logical constants, even if the consistency requirement or the harmony requirement is satisfied. As we have seen,

[74] More precisely, it is possible that the metatheory does not contain the logical constants but some other primitive notion, such as the notion of a computable functional of finite type used by Gödel in his *Dialectica* interpretation (Gödel, 1958). But such a notion (and consequently the logical constants, which can be defined in terms of it) cannot again be explained in any way but informally.

Prawitz traces his idea that introductions are meaning-giving, hence self-justifiying, back to Gentzen (1935: 80). I cannot avoid the impression that this passage has been taken too literally. On the one hand, Gentzen's 'definitions' could after all be construed as Fregean elucidations rather than as creative definitions. On the other hand, Gentzen himself later followed Hilbert's steps: in Gentzen (1936), when a specification of the finitary meaning of the logical constants is needed, he gives it in the style of Frege's elucidations, not by means of axioms or rules.[75] More precisely, for each logical constant he gives an informal explanation of its meaning, and then shows how the corresponding rules (*both* introductions *and* eliminations) are justified by that meaning, just as Frege had done in "Compound thoughts". The idea that introductions constitute the meaning is completely absent.[76]

Prior's remarks about uniqueness can be seen as an implicit reductio of assumption (48), independent from the *tonk* one, along the following lines:

(48) Any tabular or inferential axiomatic characterization of any logical connective is sufficient to determine its meaning.
(52) Then two connectives **C** and **C′** with the same tabular or inferential characterization have the same meaning, i.e. are synonymous.[77]
(53) There are cases in which $\alpha\mathbf{C}\beta$ and $\alpha\mathbf{C}'\beta$ have the same tabular or inferential characterization but are not intuitively synonymous.
(54) In those cases the tabular or inferential axiomatic characterization of a logical connective is not sufficient to determine its meaning.

It is interesting to observe that Heyting's identification of the meaning of α with its proof conditions (as opposed to either truth or assertibility conditions) is not exposed to this objection. What a(n evidential) proof of α is depends on the logical structure of α; since it is possible that α and β are logically equivalent/equiassertible without having the same logical structure, it may happen that α and β are logically equivalent/equiassertible but that a proof of α is not a proof of β. Hence, if meaning is equated with proof conditions, an account can be given of the possibility that two sentences with different logical structures are intuitively non-synonymous although they are logically equivalent.

To conclude, let me quote Prawitz's analysis of *tonk* problem in Prawitz (2005):

[75] See for instance Sect. 10.

[76] At this point Prior's position turns out to be entirely consonant with the following remark of Kleene:

> The axioms of an axiomatic theory are sometimes said to constitute an implicit definition of the system of the objects of the theory; but this can only mean that that the axioms determine to which systems, defined from outside the theory, the theory applies. (Kleene, 1952: 27)

[77] Notice however that the step from (48) to (52) is not valid in Martin-Löf's type theory, where it is possible that two connectives **C** and **C′** have the same introduction and elimination rules but are not synonymous because it is not the case that $\alpha\mathbf{C}\beta = \alpha\mathbf{C}'\beta$: PROP. Thanks to an anonymous referee for this remark.

> We cannot find a meaning for "tonk" that accords with such inference rules [...]. One cannot, therefore, say that sentences mean whatever the rules that govern them make them mean [...].
>
> [...] we need a notion of proof such that a proof of a sentence can reasonably be said to constitute evidence in view of the meaning of the sentence. Within intuitionism one has often tried to explain the meaning of the logical constants by resorting to a notion of proof. (Prawitz, 2005: 683)

It seems to me that this is the correct analysis of the problem, and the analysis Prior himself suggests: for certain rules to be acceptable there must be a meaning that accords with them: meaning comes first, and it must be explained in terms of evidence, hence of the notion of proof. But this approach to meaning is incompatible with the idea that rules are *constitutive* of meaning, even that *some* rules are, unless we «believe in magic».

2.4.2 The Canonical/Non-canonical Distinction

If introduction rules are not meaning-giving, they cannot be self-justifying; if they are not self-justifying, it is the whole neo-verificationist architecture of justification as reduction to canonical form that falls down. It is therefore worthwile to ask whether the neo-verificationist distinction between canonical and non-canonical proofs is really necessary. The distinction is absent from both Heyting's and Kolmogorov's writings; this is usually seen by neo-verificationists as an oversight, or an omission, in any case as something 'misleading',[78] since according to them there are cogent reasons to distinguish them. I shall argue that the distinction is not necessary within the intuitionistic conceptual framework.

The crucial rationale for the distinction, according to Dummett, has do with the meaning of $\rightarrow$ and $\forall$ (I shall consider only $\rightarrow$, since to $\forall$ analogous considerations apply). Here is how Dummett puts it:

> Now, for a construction to be a proof of [...] A→B [...] we are required to recognize it as operating [...] on any proof of A to yield [...] a proof [...] of B [...]. If we were to understand 'proof' here as meaning any ordinary informal proof, then this stipulation would place no restriction whatever on what we were to acknowledge as constituting a proof of [A→B]. [...] Whatever we chose to accept as being a proof of A→B, it would, provided that it itself conformed to the canons of ordinary informal proof, supply us with an effective means of transforming any proof of A into a proof of B, namely by annexing to the proof of A the given proof of A→B and then appending a single application of modus ponens. (Dummett, 2000: 271)

[78] Cp. for example Prawitz (1987), 139:

> The usual intuitionistic attempt to explain the logical constants in terms of what counts as proofs of sentences of different logical forms is quite misleading in that respect.

This passage, like other similar ones, is usually read, correctly in my opinion, as an argument for the thesis that «the intuitionistic explanations of the logical constants require a distinction between canonical proofs and demonstrations» (Prawitz, 1987: 140).[79] Understood in this way, it seems to me unconvincing, for it is clear that, in Heyting's inductive clause, "proof" could *not* be understood as meaning ordinary informal proof. With the expression "ordinary informal proof" Dummett is referring to such linguistic entities as arguments, the sort of things one finds in a logic handbook (as is made clear by his mention of Modus Ponens), or to the mental correlates of those arguments, i.e. to what I have called above *inferential* proofs. Heyting, on the contrary, is defining *evidential* proofs. Hence the validity of Modus Ponens is not *constitutive* of our capacity to grasp the notion of method: it is a *consequence* of such a capacity. Therefore, it is not true that the clause «A proof of A → B is a method for transforming every proof of A into a proof of B» imposes no restriction.

In conclusion, Dummett is surely right when he writes that, when the explanation of the logical constants is given,

> we are not appealing to an already understood notion of proof, of which notion the validity of the elimination rules is partially constitutive, but laying down what is to count as a proof in such a way that the validity of those rules follows as a consequence. (Dummett, 2000: 272)

But he is wrong when he goes on writing:

> in recognizing a construction as a proof of A→B [...] we are supposed to see it as transforming any proof of A into a proof of B [...] without appeal to the fact that we have, in it, a general construction that will do this for every proof of A [.]. (*Ibid.*)

In fact there is no need of the strange restriction imposed by Dummett (how could we see that a construction transforms any proof of α into a proof of β and refrain from appealing to the fact that it has such a property?), since no threat of vicious circularity hangs over Heyting's clause for implication, when proofs are understood as evidential proofs: Heyting is inductively *defining* the notion of evidential proof of α, not isolating (as Dummett seems to suggest) the canonical proofs of α within a previously given domain of inferential proofs.

There is, according to Dummett, another reason to introduce the canonical/non-canonical distinction; it has to do with the meaning of ∨ and ∃ (again, I shall consider only ∨, since to ∃ analogous considerations apply). Heyting defines a proof of α∨β as a proof of α or a proof of β; Dummett remarks:

> In an ordinary informal proof, however, a statement A∨B might appear as a line of that proof, asserted not because a proof had been given of one or other disjunct but because we have an effective means of obtaining such a proof [...]. We thus appear to be forced to acknowledge a distinction between a proof, in the strict sense of the word, and a mere demostration, the latter being related to the former by the fact that a demostration supplies an effective means of constructing an actual proof. (Dummett, 2000: 270–271)

[79] A different construal would take it as an argument for the thesis that a *neo-verificationist* explanation of the logical constants requires such a distinction, where neo-verificationism is characterized by an integration of intuitionistic and Gentzenian ideas. But Dummett's overall position encourages the former construal; see for instance Dummett (1975), 31.

If I interpret correctly the first sentence of this passage, Dummett is alluding to the fact that even in the mathematical domain Heyting's definition seems to be too restrictive; for example, a primality test for n, i.e. an algorithm for determining whether n is prime, would be accepted even by intuitionists as a proof of "Prime(n)∨¬(Prime(n)": if n is some very large number, we have an example of a proof of a statement of the form α∨¬α which is neither a proof of α nor a proof of ¬α. In the second sentence Dummett suggests the neo-verificationist diagnosis of this inadequacy: what Heyting is defining, as a matter of fact, is the notion of *canonical* proof, for which clause (H∨) is perfectly adequate.[80]

However, a consequence of this diagnosis is that the primality test, and, in general, decision procedures should be seen as cases of non-canonical proofs, i.e. of proofs requiring a justification.[81] This seems to me untenable. A non-canonical inference must in general be justified, as we have seen, and the justification procedure consists in showing how one could avoid the use of the elimination rule if its main premiss is derived by introduction—more briefly: in showing how the non-canonical proof reduces to a canonical one. Which canonical proof does the primality test for n reduce to?

Since the typical non-canonical proof of "C(n)" (where "C(n)" abbreviates "Prime(n)∨¬Prime(n)") is by mathematical induction,[82] its normal form is essentially a sequence of steps whose conclusions are C(0), C(1), …, C(n). The problem emerges at this point: adopting the justification procedure just described amounts to considering the assertibility of "C(n)" as conceptually depending upon the assertibity of "C(n-1)", "C(n-2)",…, "C(0)", which is not the case: when n is very large, "C(n)" is assertible today even if neither "Prime(n)" nor "¬Prime(n)" are assertible today, simply because the primality test has not been applied to n. The conclusion to draw

[80] A different diagnosis is possible, as Dummett himself concedes: that Heyting is not defining, adequately, the notion of canonical proof, but, inadequately, the notion of proof (*tout court*), and that a better clause for ∨ is the following:

> instead of requiring, for a construction to be a proof of A ∨ B, that it actually be a proof of A or of B, we could require merely that it constitute what we can recognize as being an effective method of finding a proof either of A or of B. (Dummett, 2000: 271)

I shall adopt (a modification of) this amendment of Heyting's clause (and an analogous one for ∃) in Chap. 5.

[81] Dummett writes for example:

> What makes this [the assertion of a disjunction without being able to say which alternative held good] legitimate […] is that we have a method which is in principle effective for for deciding which of the two alternative is correct: if we were to take the trouble to apply this method, the appeal to an argument by cases could be dispensed with. (Dummett, 1975: 239)

[82] In a natural deduction system, mathematical induction is the elimination rule corresponding to the two introduction rules for the natural numbers (corresponding to the first two Peano axioms):

$$\frac{}{N0}NI_1, \qquad \frac{Nn}{NSn}NI_2, \qquad \frac{N(t) \quad \alpha(0) \quad \begin{matrix}[\alpha(x)]\\ [\alpha(Sx)]\end{matrix}}{\alpha(t)}NE$$

seems to be that the primality test is not to be seen as a non-canonical proof of $\alpha \vee \neg\alpha$; what justifies arguments by cases is not their dispensability, but the very meaning of $\vee$.

I have argued that the reasons adduced by neo-verificationists for the distinction between canonical and non-canonical proofs are not cogent, from an intuitionistic point of view. More precisely, I have argued that there is no reason for the distinction if what is meant by "proofs" are evidential proofs, and that Heyting was just meaning evidential proofs in his explanation of the meaning of the logical constants. On the contrary, when the neo-verificationists speak of intuitionistic proofs, by "proofs" they mean usually inferential proofs, as is evident from Dummett's passages quoted above and form the following slightly different justification of the distinction by Prawitz, having again to do with the meaning of $\vee$:

> A problem with [Heyting's] explanation is that the notion of proof used here cannot stand for whatever establishes the truth of sentences in a normal intuitive sense. For instance, a proof of a disjunction $A(t) \vee B(t)$ may very well proceed even intuitionistically by first proving $\forall x(A(x) \vee B(x))$ and then applying universal instantiation to infer $A(t) \vee B(t)$. Given such a proof, we do not know which of the two disjuncts holds. Hence, it is not correct to say that a proof of a disjunct needs to consist of a proof of one of the disjuncts together with indication of which disjunct is proved. In intuitionistic meaning explanations of the kind exemplified above, proof must thus be meant in a quite restrictive way. (Prawitz, 2005: 684)

Here the (false, in my opinion) presupposition is clearly that Heyting is speaking of inferential, not of evidential, proofs; and the reason of this presupposition, I surmise, is that Prawitz is looking at intuitionistic proofs through the glasses of Gentzen's natural deduction rules, i.e. as sequences of inferential steps.[83]

The canonical/non-canonical distinction is extended by neo-verificationists to atomic sentences, in which case it is often conflated with the distinction between direct and indirect proofs. I shall discuss both this conflation and the extension of the canonical/non-canonical distinction to atomic sentences in Chap. 4.

2.4.3 Transparency

In this section I conclude my comparative assessment of intuitionist and neo-verificationist theories of meaning by introducing and discussing Prawitz's and Martin-Löf's views about the epistemic transparency of proofs; as for Prawitz, I shall examine his views after the introduction of the ToG.

[83] The same holds for Dummett; cf. his discussion of the intuitionistic notion of proof in Dummett (2000), §7.2.

2.4.3.1 Transparency and the ToG

In his 2015 paper "Explaining Deductive Inference" Prawitz introduces his ToG by stressing the risk to get entangled into circles when trying to explain what evidence is: on the one side, «for logically compound sentences there seems to be no alternative to saying that evidence comes from inference» (Prawitz, 2015: 77); on the other, an inference must be legitimate, and there seems to be no way to characterize a legitimate inference without making reference to the notion of evidence: a generic inference is legitimate if a subject who makes it and has evidence for the premisses thereby gets evidence for the conclusion. The only strategy allowing to escape this circularity—Prawitz suggests—is to accept the idea

> that it is in the nature of the meaning of some types of sentences that evidence for them can only be explained in terms of certain kinds of inference. The legitimacy of these inferences is then a datum that has to be accepted as somehow constitutive for the meaning of these sentences.[84] (Prawitz, 2015: 77–78)

But in order to implement this strategy it is necessary to account also for the legitimacy of inferences that are not meaning-giving. Here Prawitz distinguishes two alternatives, corresponding to two opposite ways of conceiving proofs. According to the former, a proof is a chain of legitimate inferences, hence the notion to be defined is the one of legitimate inference; according to the latter, a proof is not such a chain, but something to be defined on independent grounds, and this opens the possibility of defining a legitimate inference as an inference that gives rise to a proof when attached to a proof. This last possibility is the one adopted by Proof-Theoretic Semantics (PTS) and by Prawitz with the definition of valid argument stated above; in Prawitz (2015) (and in other recent papers) he expresses the opinion that it «turn[s] the usual conceptual order between inferences and proofs upside down» (Prawitz, 2015: 78), and this seems to be the main reason why he adopts now the former alternative. This brings to the fore the notion of legitimate deductive inference, thereby raising the fundamental problem of explaining why do certain inferences have the power to transmit to the conclusion the evidence we have for the premisses.[85]

The notions of ground and of ground-building operation are introduced just to define legitimate inference, and having evidence for α is programmatically equated to being in possession of a ground for α; within the framework of ToG a proof can be defined as a chain of valid inferences, i.e. inferences applying grounds for the premisses into a ground for the conclusion.

[84] For reasons explained above I do not think that Heyting would agree that the legitimacy of introductory inferences is constitutive of the meaning of their conclusions.

[85] In Prawitz (2011: 389) the problem is introduced and stated in the following terms:

> If we think of a proof as a chain of inferences, the crucial problem is thus what it is that gives certain inferences the power to justify their conclusions.

Let us now consider the question of transparency. Epistemic transparency is primarily a property of the *possession* of intuitive evidence: the possession of evidence E for a sentence α is *epistemically transparent* if, and only if, it cannot happen that one is in possession of E without being in a position to know that one is. When we are looking for a theoretical notion "X for α" to propose as an *explicans* of the intuitive notion of evidence for α, it may be useful to take into consideration also the question of the epistemic transparency of the *notion* "X for α" itself, i.e. the question whether it may happen that one is presented with an X for α without being in a position to know that it is an X for α.[86]

According to PTS, the *explicans* of the notion of evidence for α is the notion of valid argument for α; according to ToG, the *explicans* is the notion of ground for (the assertion of) α. Now, a common feature of PTS and of ToG is that neither the possession of a valid argument for α nor the possession of a ground for α is epistemically transparent. Let us see why.

In the case of PTS, having evidence for α amounts to having constructed a valid argument for α. Valid arguments may be either canonical (i.e. arguments whose final step is the application of an introduction rule) or non-canonical. According to Prawitz (2015: 83), we may assume that when we have constructed a valid *canonical* argument we know that it is valid; presumably the reason is the following: the last inference is an introduction, and introduction rules are meaning-giving, in PTS, so that in order to know that the argument π for α is valid it is sufficient to understand α and to know that the immediate sub-arguments of π are valid; this can be assumed by induction hypothesis, and the basic clause of the definition of validity says that the arguments for atomic sentences are valid and canonical, so that in order to know that they are valid it is sufficient to understand their conclusions [and «It can be assumed to be a part of what it is to make an inference that the agent knows the meanings of the involved sentences» (Prawitz, 2015: 96)]. On the contrary, when we have constructed a valid *non-canonical* argument, the only way we have to know that it is valid is to try to reduce it to canonical form: we know *how* to do it and, if the argument is valid, the reduction procedure will terminate, as a matter of fact, but in general we do not know *that* it will terminate.[87] Therefore we cannot say in general that, when we have constructed a valid argument for α, we know that we have.

Let us consider ToG. According to it, «it is convenient—in Prawitz's words—to think of evidence states as states where the subject is in possession of certain objects» (Prawitz, 2015: 88), called *grounds*. As grounds are objects,

> we get to know them via descriptions. To form a ground for an assertion is thus to form a term that denotes the ground, and it is in this way that one comes in possession of the ground. (Prawitz, 2015: 89)

[86] The question of the epistemic transparency of the notion of *intuitive* evidence for α will be discussed in a moment.

[87] The reason is strictly connected to the non-decidability of the existence of a normal form of a given formula. Cf. Church (1936).

In my opinion, there are two distinct reasons for the non-transparency of ground possession.

The first has to with grounds for atomic sentencs. According to Prawitz, we may assume that, when we have formed a term *t* denoting a ground for an atomic sentence α, we know that *t* denotes a ground for α. The reason for this assumption is explained in the following way:

> It can be assumed to be a part of what it is to make an inference that the agent knows the meanings of the involved sentences. Since the meanings of closed atomic sentences are given by what counts as grounds for asserting them, she should thus know that *T* denotes a ground for asserting an atomic sentence *A* when this is how the meaning of *A* is given. (Prawitz, 2015: 96)

This seems to me an excellent formulation of the reason why the agent *should* know that *T* denotes a ground for α, *not* a reason why the agent *does* know that *T* denotes a ground for α. In other terms: what Prawitz is stating is a *condition of material adequacy* to be imposed onto any definition of the notion of ground: the definition must be such that one knows that one possesses a ground for an atomic sentence α whenever one does possess such a ground. I utterly agree that this is the right condition of adequacy; actually, as I shall argue in a moment, I hold that it should be extended to possession of grounds for sentences of whatever logical complexity. But the question Prawitz should answer here is quite different, namely: is the condition of material adequacy actually satisfied by the definition he gives of ground for an atomic sentence? And the answer, I think, is negative, in the case of atomic empirical sentences. Take an atomic sentence like "It is raining"; according to ToG, a ground for asserting it «would be got by making an adequate observation» (Prawitz, 2012: 893). What is an *adequate* observation? More specifically: suppose the subject has a visual hallucination according to which (s)he has the experience of presently falling rain, and that in fact it is not raining; is the subject's observation that it is raining adequate or not? In a paper of 2002 Prawitz writes:

> That there are conclusive verifications of [observation sentences] does not mean that we cannot make mistakes by erroneously thinking that we have verified them. In mathematics we operate with the notion of conclusive proofs although of course we may also make mistakes there, in which case we say that what we thought was a proof was not really a proof. Similarly I would say that we may think to have seen something but that it later turns out that we did not see it. (Prawitz, 2002: 90–1)[88]

From this standpoint, seeing that it is raining is a conclusive verification of the sentence "It is raining"; since presumably "conclusive" means here factive, we can conclude that our subject's hallucinatory experience is *not* a case of seeing, i.e. it is not an adequate observation that it is raining; as a consequence the subject believes to have a ground for asserting that it is raining, but in fact (s)he is making a mistake, because (s)he has not such a ground.

[88] For a discussion of the analogy Prawitz postulates here between the domains of mathematics and perception see Chap. 3, Sect. 3.1.2.

This latter answer seems to be the official answer of ToG, since in one of the first papers devoted to the theory Prawitz writes:

> Outside of mathematics, we may consider observation statements, and for them, I suggest, we take relevant verifying observations to constitute grounds. For instance, a ground for a proposition 'it is raining' is taken to consist in seeing that it rains; taking "seeing" in a veridical sense, it constitutes a conclusive ground. (Prawitz, 2009: 186)

Now, if a ground for "It is raining" consists in seeing that it is raining, and seeing is taken in a veridical sense, i.e. is a factive way of having warrant to assert "It is raining", then it is illegitimate to assume that having such a ground is epistemically transparent, since, in general, the evidence actually available to a subject is not sufficient for her/him to discriminate between seeing (veridically) that it rains and merely having the impression to see that it rains: the experience of the subject is the same, whether (s)he is seeing or merely having the impression to see; it may therefore happen that the subject is in possession of a ground for "It is raining" without being in a position to know that (s)he is; in other words, the possession of a ground for "It is raining" is *not* epistemically transparent.

The second reason of the non-transparency of ground possession has to do with grounds for logically complex sentences, and is stated by Prawitz himself (2015: 96–97). The terms denoting them may be either canonical (i.e. having as first symbol one of the primitive operations) or non-canonical. When we are in possession of a canonical term t denoting a ground for α, we may assume that we know *that* it denotes a ground for α, since the meaning of α is given in terms of primitive operations, and consequently in order to know that t denotes a ground for α it is sufficient to understand α and to know that its sub-terms denote grounds for the subformulas of α—what may be assumed by induction hypothesis. On the contrary, when we are in possession of a non-canonical term t, the only way we have to know that it denotes a ground for α is to try to bring it to canonical form: again, we know *how* to do it, and, if it actually denotes a ground for α, the reduction procedure will terminate, as a matter of fact, but in general we do not know *that* it will terminate.

Concluding his discussion of PTS, Prawitz observes that

> If the latter [namely knowing *that* applying the reduction rules will produce an argument in canonical form for A (*g.u.*)] is required to have evidence for A, then it is not enough to know or to be in possession of a valid argument for A, one must also know that what one is in possession of is a valid argument. (Prawitz, 2015: 83–84)

In other terms, *if* intuitive evidence is assumed to be transparent, then having a valid argument for α cannot be taken to be a good *explicans* of having evidence for α. The question to consider is therefore whether possession of *intuitive* evidence is transparent.[89]

[89] Prawitz does not discuss this very question. He discusses a related one, namely whether possession of an *intuitionistic construction* of α is epistemically transparent (see below in the text); but intuitionistic constructions are theoretical *explicantia* of evidence, not intuitive evidence.

Maybe there are different intuitive notions of evidence, but if we want to build a theory of meaning based on evidence, we must select one notion, and our choiche will depend on the needs of the theory we are going to build. It seems to me that only an intuitive notion of evidence whose possession is transparent is capable to play the role Prawitz assigns to evidence in his explanation of inference.

Suppose the possession of intuitive evidence is not transparent; this means that there are situations in which a subject $\mathcal{S}$ is in some way in possession of evidence E for a transition from α to β (I assume for brevity that the premiss is only one), but is not in a position to know that (s)he is. If $\mathcal{S}$ is not in a position to know that (s)he is in possession of evidence E, there is some obstacle blocking her/his path to knowing that (s)he is, and one such obstacle may well be the fact that (s)he sees no reason to infer β from α, and therefore (s)he does not believe to be in possession of E. In such a case, why should $\mathcal{S}$ apply her/his 'transitional' evidence E to the evidence (s)he has for α? (S)he does not believe that the application of E will give her/him evidence E′ for the conclusion β, so (s)he has no reason to infer β from α. To be warranted *in believing* that β one must not only have evidence E′ for β, but also base one's belief on E′; and to base one's belief on E′, one must know *that* E′ is evidence for β; analogously, to be warranted *in believing* that β follows from α one must not only have evidence E for β from α, but also base one's belief on E; and to base one's belief on E, one must know *that* E is evidence for β from α. The same holds for acts like assertion and inference. In conclusion, if evidence, and more generally 'transitional' evidence, is to play the explanatory role Prawitz assigns to evidence in his explanation of inference, it must be epistemically transparent. This point is clearly illustrated by the following example, due to Roderick Firth:

> Let us suppose, for example, that Holmes knows at a certain time *t* that the coachman committed the murder. Holmes has studied the mud on the wheels of the carriage and from this and other evidence has reached a correct conclusion by rational inference. We may then employ the term "warranted" to say two quite different things. We may say that the proposition "The coachman did it" is warranted for Holmes at *t*. It is warranted *for* Holmes and not *for* Watson because it is warranted on the basis of evidence possessed only by Holmes. But we can also say that Holmes, because his conclusion is based rationally on the evidence, is warranted *in believing* that the coachman did it.
>
> This distinction between propositional and doxastic warrant is dramatized if we now suppose that Holmes shows Watson the mud and gives him all the other relevant evidence he has, without telling him what conclusion he has drawn from it. In one important respect a change has occurred in Watson's epistemic condition. We may express this fact by saying that the proposition "The coachman did it" is now warranted for Watson. It is warranted for Watson whether or not he believes that the coachman did it. But even if Watson does believe that the coachman did it, we cannot therefore conclude that Watson, like Holmes, is warranted *in believing* that the coachman did it. Believing a proposition *p* is a necessary condition for being warranted in believing *p*. But Watson's belief might not be based rationally on the evidence. (Firth, 1978: 218)

When Holmes shows him the mud and all the other relevant evidence he has, Watson is in possession of evidence for the proposition "The coachman did it", but he doesn't know that he is, because he is not capable to recognize the mud and the other signs *as* evidence for that proposition; and this is the reason why we cannot say that Watson is warranted *in believing* that the coachman did it.[90]

Let me briefly discuss another, distinct reason Prawitz adduces for the non-transparency of the possession of grounds. The reason is the non-transparency of the *notion* of ground itself. Clearly, *if* the notion of ground for α is not transparent, then neither possession of a ground for α is; but *why* is the notion non-transparent? Prawitz argues for this claim in the following way. He starts by remarking that

> because of Gödel's incompleteness result [...] already for first order arithmetical assertions there is no closed language of grounds in which all grounds for them can be defined. (Prawitz, 2015: 98)

In other terms, for every formal theory $\mathcal{T}$ of grounds there is a first order arithmetical sentence α and a ground g such that g is an intuitive ground for α, but the proposition translating "g is a ground for α" is not a theorem of $\mathcal{T}$. At this point Prawitz states what he considers to be the crucial question:

> The crucial question is therefore if it is decidable for an arbitrary definition of an operation, which we may contemplate to add to a given closed language of grounds, whether it always produces a ground of a particular type when applied to grounds in its domain? (Ibid.)

and he concludes:

> it seems to me that we must be sceptical of such an idea, and therefore also of the idea that the condition for something to be a proof or to constitute evidence is luminous. (Ibid.)

I must avow that I do not see any reason for this conclusion. Suppose that the correct answer to the crucial question is negative, namely that there is an operation O[91] on grounds represented in the formal theory of grounds $\mathcal{T}$ by a term K such that if $t_1, \ldots, t_n$ are terms denoting grounds for $\alpha_1, \ldots, \alpha_n$, respectively, then $K(t_1, \ldots, t_n)$ denotes a ground for β, but neither the sentence of $\mathcal{T}$ translating

(55) $K(t_1, \ldots, t_n)$ is a ground for β

nor the sentence translating

(56) $K(t_1, \ldots, t_n)$ is not a ground for β

is a theorem of $\mathcal{T}$. There is no reason to conclude that it is not *intuitively* evident that $K(t_1, \ldots, t_n)$) is a ground for β, and therefore that intuitive evidence is not epistemically transparent: the fact that (55) and (56) are not theorems of $\mathcal{T}$ is logically independent of the fact that (55) is (or is not) intuitively evident. Concededly, formal provability is *intended* to catch intuitive evidence, but sometimes it does not succeed, as just Gödel's theorem shows; when this happens, we don't infer that intuitive evidence is different from what it appears to be (for instance, that Gödel's sentence is not intuitively true/evident), but that formal provability is incomplete. Analogously,

---

[90] Of course there are differences between the case considered by Firth and the ones that are encompassed by ToG (for instance, it is a case of abductive, not deductive, inference; its conclusion is atomic, not logically complex); but they do not seem relevant to the point I am making.

[91] I assume for simplicity that O is defined on grounds for sentences.

the conclusion to draw from Prawitz's example should be that provability in the theory $\mathcal{T}$ is not a good *explicans* of the intuitive notion of evidence. In conclusion, the reasons Prawitz adduces for the claim that evidence is not epistemically transparent, if I understand them correctly, do not seem to be to the point.

Moreover, it seems to me that Prawitz's scepticism about the epistemic transparency of grounds conflicts with one of the roles ToG assigns to grounds. The notion of ground plays in ToG the role of the key notion of the theory of meaning, in the sense that «the meaning of a sentence is explained in terms of what counts as a ground for asserting it» (Prawitz, 2015: 89); as a consequence, knowing the meaning of a sentence α amounts to knowing what counts as a ground for asserting α; and I cannot see any way of understanding the phrase "knowing what counts as a ground for asserting α" other than being in a position to recognize a ground for α when presented with one—the intuitionistic way of understanding it. But if the notion of ground is non-transparent, as we have just seen to be claimed by Prawitz, we have the unwelcome consequence that a sentence whose grounds we are not in a position to recognize is either a sentence we don't understand or a sentence whose ground-conditions are epistemically transcendent, like classical truth-conditions. The necessity to escape this dilemma seems to be the deep reason of what Kreisel calls «the basic intuitionistic idealization that we can recognize a proof when we see one».

It is interesting to observe that, according to Prawitz, also possession of an intuitionistic construction for α risks *not* to be epistemically transparent. He writes about it:

> The situation is parallel to the one [of PTS (*g.u.*)]. To be in possession of a construction of for instance A→B is to know an effective operation that applied to a construction of A yields a construction of B, and hence it is to *know how* to find a construction of B given one of A, but it is not to *know that* there is a such an effective operation. (Prawitz, 2015: 87)

It is not clear why, exactly, Prawitz holds that the situation of the possession of intuitionistic proofs is parallel to the one of the possession of valid arguments as conceived by PTS. In the cases of PTS and ToG the parallelism is clear, as we have seen above: when we are in possession of a non-canonical valid argument (or of a non-canonical term *t* denoting a ground) for α, we know *how* to reduce it to a canonical argument (or term), but in general we do not know *that* the reduction procedure will terminate; knowing-how and knowing-that are two clearly distinct kinds of knowledge, so it is clearly possible that one subsists and the other does not. But the appeal to the difference between knowing-how and knowing-that is based on the distinction between canonical and non-canonical arguments (and, parallelly, terms), and this distinction is *not* present in Heyting's conceptual framework. As we have seen, in some of his papers Prawitz considers this absence as something 'misleading', presumably because he holds there are cogent reasons for the distinction. I have argued above against such reasons; here I only stress that, in absence of a distinction between canonical and non-canonical proofs, the difference between knowing-how and knowing-that cannot be invoked to give substance to the gap between knowing a proof of α and knowing that it is a proof of α.

Prawitz would probably insist that, in spite of Heyting's omission, the parallelism wih PTS and ToG still subsists: in the BHK-interpretations proofs are understood as (abstract) objects; hence we get to know them as we get to know all abstract entities, i.e. via descriptions, and this fact has consequences analogous to the ones drawn by Prawitz about grounds: being in possession of a proof-as-object of α amounts to knowing a term that denotes that proof, and at this point it is necessary to distinguish between knowing a proof-as-object with such and such properties and knowing that it has such and such properties; in other terms, it is possible that one has a term denoting a proof-as-object of α without knowing that one has a term denoting a proof-as-object of α.

In conclusion, the lack of epistemic transparency of the possession of an intuitionistic proof of α is—according to Prawitz—a consequence of equating such possession with knowing a term that denotes that proof. This however, far from suggesting that the intuitive notion of evidence for α is not epistemically transparent, can be taken as showing that the possession of intuitive evidence for α is not adequately analysed as the possession of a name denoting such evidence, since the possession of intuitive evidence for α *is* epistemically transparent. So it becomes interesting to explore the possibility of a direct access to evidence, not mediated by terms or descriptions; and since the mediation of terms seems unavoidable in order to make reference to objects, it becomes interesting to explore the possibility of *not* reifying evidence for α, i.e. of conceiving it and possession of it as *not* consisting in an (abstract) object and in possession of an object, respectively. This is what I shall try to do in Chap. 3. Finally, since the availability of an epistemically transparent notion of evidence is crucial for intuitionism, it is natural to look at intuitionistic proofs as a source of inspiration. This is what I shall try to do in Chap. 5.

2.4.3.2 Transparency and Type Theory

In Sect. 2.3.4 I have explained why Martin-Löf's theory of types (TT) *seems* capable to suggest a way out of the difficulties pointed out by Prawitz concerning the decidability of the relation

(57) c is a proof of α.

Here I shall argue that it *is not* capable to.

When I said (in Sect. 2.3.4) that Martin-Löf's intuitionistic theory of types systematically exploits Curry-Howard's isomorphism, I meant that the correspondence between λ-terms and intuitionistic proof-objects is to be understood in a strong sense, i.e. not simply in the sense that λ-terms *denote* proofs, but in the sense that they *are*, or *become*, proofs when they are formed and manipulated according to the use rules of TT—rules which, according to the slogan «Meaning is use», codify our understanding of their meaning: «A mathematical object is the same as a meaningful mathematical expression.» (Martin-Löf, 1987: 17) This permits to interpret the relation (57) in a way different from ToG's.

In ToG the relation (57) is equated to the property

(58) the term t denotes a ground for α,

which in turn is analyzed as:

(59) there is a term t' such that (t has the same denotation as t' and t' is in canonical form and and t' denotes g and g is a ground for α).

In general, the property (59) holds for terms of formalized theories, whose only non-introductive rules are eliminations; but ToG is an open theory, with a non-specified class of non-introductive rules. For reasons related to the fact that the property "t is riducible to normal form" is not decidable,[92] neither the property (58) is; it may therefore happen that one knows *how* to reduce a term to canonical form, but not *that* it is reducible, or—equivalently—that a term denotes a ground for α, but nobody knows it (Prawitz, 2015: 96).

In TT, as I said, terms *are* proof(-object)s, provided that certain conditions are satisfied; and such conditions are stated by rules involving sentences called judgements—formation and inference rules which determine the meaning of the terms involved. Hence we can say that a term *becomes* a proof-object when it is understood by means of an *act* of understanding.[93] It becomes a proof-object *of what*? Of a proposition, and the identity of such a proposition is determined by assigning a type to the term, according to formation rules (for each logical constant); this is the propositions-as-types idea. In this way, the intuitive relation (57) is expressed, in the language of TT, by the judgement t:α, i.e.

(60) t is a term of type α,

which is decidable (by means of a type-checking algorithm).

The question is whether interpreting (57) as (60) accounts for the intuitive meaning of (57). I think not. As we have just seen, interpreting (57) as (60) presupposes an assignment of types to terms; such an assignment is clearly defined in the case of logically complex sentences (in fact, it is a significant component of the theory of meaning of the logical constants associated to TT), but what about atomic sentences? In Martin-Löf (1985: 35) a (canonical) proof (or verification) of the nonmathematical sentence

[92] Cp. Church (1936). On the denotation relation in ToG cp. Piccolomini d'Aragona (2019), Sect. 5.2.4.

[93] Cp. Martin-Löf (1987), 57:

> The expression which stands for a mathematical object is the matter out of which it is made, and the type of the object is its form. In this sense, a mathematical object is a composite of matter and form. But one component has been forgotten, namely, the act of understanding through which the matter receives its form. It is through this act that the object comes into being.

(61) The sun is shining

is defined as «the direct seeing of the shining sun», and a direct thermometer reading is proposed as a (canonical) proof of «The temperature is 10 °C». However, seeing and reading are *acts*, not objects, and acts are in TT, as we have seen, the proofs of judgements, not of propositions. This may be the reason why in later papers Martin-Löf suggests a more 'objectual' characterization; in Martin-Löf (1994: 35), for instance, he argues that the whole complex of (61) taken together with the shining sun is an analytic judgement; presumably, therefore, the shining sun is a proof-object of (61), and "the shining sun" a term denoting it. By analogy we can conjecture that a proof-object of the mathematical sentence.

(62) $3 + 2 = 5$

is something like evaluation of $3 + 2$ to canonical form, and "evaluation of $3 + 2$ to canonical form" is a term denoting it. Assuming that this is what Martin-Löf suggests, it seems to me that the 'reduction' of (57) to (60) is plausible in the case of mathematical sentences, but not in the case of nonmathematical ones. It is plausible in the former case because of a peculiarity of names of mathematical entities: the fact that their denotations can be read into their canonical form, thanks to our 'number faculty'—a factor of our 'science forming faculty', according to Chomsky (1980: 38); in this case, literally, the term becomes its denotation (in our example, "evaluation of $3 + 2$ to canonical form" becomes 5), and Martin-Löf's idea quoted above—that «A mathematical object is the same as a meaningful mathematical expression»—is plausible. However, the 'reduction' is not plausible in the case of nonmathematical sentences[94]: the denotation of "the sun" cannot be read into the term, hence it would be senseless to say that "the sun" becomes the sun, even after an act of understanding the term.

Moreover, the 'objectual' characterization of verifications generates a tension with the analysis of analytic judgements proposed by Martin-Löf. We have seen that, according to him, the complex of (61) plus the shining sun is an analytic judgement; the reason is that «everything is contained in that judgement that you needed in order to convince yourself of it» (Martin-Löf, 1994: 90); but it seems unlikely that the shining sun is enough to convince me of (61), if I do not look at it. As a matter of fact, even the performance of an act of seeing seems largely insufficient to make (61) evident. Why an act of seeing, and not of hearing? Why an act of seeing the sun and not the moon? Why an act of seeing the sun shining and not setting? And so on. Clearly, what is lacking is an explanation of the relations between the meaning of (61), the sun, and shining, between seeing the sun and the sun, between the concept of shining and shining, and so on; the mere postulation that acts generate evidence of propositions (or judgements) is not enough. Of course a thorough description of the mental operations involved in this process is of pertinence of psychology, not of logic nor of the theory of meaning; but the theory of meaning, and the theory of logic,

[94] The answer may be unconvincing even when it is applied to mathematical sentences containing names of mathematical entities different from natural numbers. Cp. Martino and Usberti (1988), 154–155.

should be compatible with the general format of psychological explanation; and, as we have seen in Chap. 1, the best kind of psychological explanation is computational-representatational explanation, which assumes as fundamental the notion of (mental) state, not the (mental) act/object dichotomy.[95]

2.4.4 The Extendability Thesis

I should like to conclude by putting into evidence an aspect of neo-verificationism I find entirely embraceable. It is the idea—which I shall call "the extendability thesis" —that the intuitionistic explanation of the meaning of the logical constants, originally applied only to mathematical sentences, must and can be extended to a full theory of meaning, concerning therefore sentences of every kind, in particular empirical sentences. The thesis is a natural consequence of Dummett's view of intuitionism as including a theory of meaning of the logical constants, which I find convincing in view of Heyting's explanation illustrated above: there seem to be no plausible reason why the logical constants should have different meanings in different cognitive domains.

At first sight the idea seems to be easy to implement: what is necessary is to find a notion capable to play in a generalized theory of meaning the same key role played in the intuitionistic theory by the notion of proof, and capable at the same time to be applied both to mathematical and to empirical sentences.

Since a proof is, in mathematics, what warrants one to conclusively establish a proposition as true, a natural generalization of this notion is the notion of *verification.* This is the neo-verificationist proposal, whose originality with respect to neo-positivism is partly due to the influence exerted on it by Quine's ideas about verification. In the last section of "Two dogmas of empiricism" Quine sketched a model of language according to which language forms an articulated and interconnected structure, with some sentences lying at the periphery and others at the interior, and such that the impact of experience is transmitted from the periphery to the interior. Dummett judges this model particularly interesting because it yields a first approximation to a view of language capable to escape the difficulties encountered by the verificationism of logical positivists, but still essentially verificationist in spirit. The basic idea of any verificationist approach is that the fundamental notion of the theory of meaning ought to be not truth but verification,[96] understood as what permits us to recognize that a sentence is true. Quine's model of language remains verificationist in the sense that our understanding of language consists, also for Quine, in our capacity to *recognize* which experiences force us to revise our assignments of truth-values to the sentences of the language, and which revisions are efficacious

[95] For an interesting attempt at founding an epistemology for Martin-Löf's type theory on the notion of cognitive act see van der Schaar (2011).

[96] Or falsification, understood as what permits us to recognize that a sentence is false. I shall not take into consideration here the possbility of a falsificationist theory of meaning.

as remedies. The difference from positivism lies in the conception of the verification of a sentence. For the positivists it is a complex of sensory experiences; since sensory experiences are not available for mathematical or logical propositions, they were forced to postulate a dichotomy between logico-mathematical and empirical truths, and to give a completely different characterization of the meaning of the two kinds of sentences. According to Quine's model of sense, on the contrary, we come to recognize a sentence, of any kind, as true through an 'argument' (in the widest sense) having that sentence as conclusion, and among its possible premisses also reports of empirical observations. From this point of view the difference between observation and mathematical sentences is no longer qualitative but quantitative: for the former the argument necessary to recognize the sentence as true will use premisses that are reports of observation, for the latter it will not.

However, the extension of the intuitionistic theory of meaning to empirical sentences faces a considerable difficulty, owing to the fact that, while proofs warrant one to conclusively establish mathematical propositions, verifications of empirical propositions are in general not conclusive. This fact generates to any anti-realist theory of meaning the problem that the assertibility of α can no longer be equated with the possession of a verification of α; but to neo-verificationists it creates a more specific difficulty, connected with the Fundamental Assumption.

Consider for example what might be called "empirical generalizations", i.e. universal sentences concerning physical phenomena, requiring for a verification both sense-experiences and inferences that are not necessarily deductive:

> At a given stage of science, there are of course certain standards that determine the kind of evidence on the basis of which such sentences may be correctly asserted, requiring e.g. that empirical investigations of a certain kind have been carried out. As science progresses, such standards are often raised; for instance, one becomes aware that certain factors that were neglected earlier [...] must be kept constant. But it does not seem reasonable to say that, when this occurs, there is a change in the meaning of the sentence in question. Rather one would say that the earlier standards for the assertion of the sentence have been shown to be mistaken.
>
> [...] [T]he problem with sentences of the kind now considered seems to be that we have no idea of what would constitute a direct verification of them independent of what we count as sufficient grounds for the assertion of them. (Prawitz, 1987: 143)

For sentences of this kind it is also unplausible to say that, if they are assertibile, then there is a canonical verification of them, as the Fundamental Assumption would require.

At least three answers to these problems seem to be possible: (i) to abandon the extendability thesis, because the key notion of an extended theory of meaning could not explain assertibility; (ii) to try to implement the extension programme by still using conclusive verification as the key notion; (iii) to try to implement the extension programme by using a non-conclusive notion as the key notion.

Neo-verificationists generally adopt the second alternative[97]; in the next chapter I will criticize this choice, and I shall take a few steps in the direction of alternative (iii).

[97] See for instance Prawitz (2002).

2.5 Conclusion

From the preceding discussion at least three reasons emerge for choosing, as a basis for an explanation of the internalist meaning of empirical and mathematical sentences, intuitionism instead of neo-verificationism. The first has to do with transparency, the second with impredicativity, the third with realism about the mental.

2.5.1 Transparency

Non-transparency of proofs, which we have observed both in PTS and in ToG and in Martin-Löf theory of meaning, depends on how the process of justification of rules is conceived, hence ultimately on the lack of a distinction between evidential and inferential proofs. The process involves (i) the reduction of terms to canonical form, and (ii) the denotation of proofs by terms. Reduction (i) is exposed to the risk of non-transparency because it may happen that the process terminates without the subject knowing; point (ii) is based on the assumption that it is possible to establish the denotation of a term on the basis of its form. This assumption is not always true; in particular, the structure of a term denoting a ground for an atomic statement P(n) does not reveal anything about the nature of its denotation.

This does not happen within the intuitionistic framework. Concerning (i): there is no need to reduce terms to canonical form, because evidential proofs are distinct from inferential proofs, and the notion of canonical evidential proof is senseless; on the other hand, inferential proofs are defined as sequences of evidential proofs, hence they transmit evidence. Concerning (ii): there is no need to denote proofs by terms, because logic does not come before mathematics, and there is no autonomous logical structure; denotation of proofs by terms may be of great help in order to reconstruct the generation process of a proof, to verify its correctness, etc.; but it plays no role in the justification of inference, from an intuitionistic standpoint. The crucial role is played by the principle of transparency, and by the assumption that rules, functions, are transparent, in the sense that our mind is capable to grasp them.

If what I have argued is correct, the problem arises of defining a theoretical notion of evidence whose possession is epistemically transparent. This is what I shall try to do in Chaps. 4 and 5.

As the first reason of the lack of epistemic transparency of ground possession in ToG is the factiveness of ground possession for atomic empirical sentences, if we look for an *explicans* of evidence whose possession is transparent, we cannot look for a factive notion. Since non-factiveness entails defeasibility,[98] we cannot look for an indefeasible notion either: a defeasible notion seems to be the right key notion of a theory of meaning of empirical sentences.

---

[98] This is an antirealist thesis, as is clear from the following argument: if one has warrant to assert α and α is not true, there must be some reason why α is not true; so, if one comes to know that reason, one has no longer warrant to assert α. For a realist counterargument see Chap. 3, Sect. 3.1.2.

2.5.2 *Impredicativity*

We have seen in Sect. 2.2.2 that the objection of impredicativity does not hit Heyting's clause (H′→), if we interpret the clause as part of a definition of the notion of evidential proof, not of inferential proof.[99] But we have seen also that the distinction between evidential and inferential proofs is blurred by the neo-verificationists, owing to the fact that they conceive proofs exclusively as inferential proofs. As a consequence, the objection of impredicativity does hit the neo-verificationist (i.e. Gentzen's) explanation of implication and, more generally, of proofs.

2.5.3 *Convergence*

I conclude with some general remarks on some more general reasons of convergence between intuitionistic theory of meaning and internalist semantics as advocated by Chomsky.

One of these reasons, of crucial importance, is of a methodological nature: according to Chomsky, only if the entities denoted by names and predicates are of a mental nature is it possible to conceive knowledge of meaning as a system of computational structures and processes, thereby giving a scientific account of semantic competence, i.e. of cognitive preconditions of linguistic use. From this standpoint internalist semantics is just the theory of semantic competence, a part of cognitive psychology; therefore the reasons that justify it are essentially the same that justify Chomsky's methodological internalism. On the other hand, remember the passage from Heyting (1956b: 228) quoted in Sect. 2.2.2: it is now possible to perceive the methodological flavor of this principle: what can be studied on a scientific basis is a *logique du savoir* because its basic notions are «as immediate as possible». Moreover, if we keep present (i) that Heyting explains the meaning of each logical constant **C** by specifying what a proof of an arbitrary sentence having **C** as its principal operator amounts to; (ii) that he postulates that knowing the meaning of such a sentence amounts to being capable to recognize its proofs; and (iii) that such proofs are (inductively) defined in terms of specific mental operations on entities having the nature of mental representations (in the non-relational sense of "representation"), then Heyting's *logique du savoir* looks very much like a theory of our semantic competence of the logical constants, or of the internal structure of our 'deductive faculty'.

In this context it may be interesting to quote an opinion of Gödel's:

> To Sue Toledo he said (at some point in the period 1972–1975) that 'intuitionism involves [an] extra-mathematical element. Namely, the mind of the mathematician + his ego', and he described intuitionism to her as 'essential a priori psychology' [...]. (Van Atten, 2015: 194)

[99] Another sort of impredicativity, related to the fact that standards of evidence change through time, may affect Heyting's explanation as well. Cp. Footnote 26.

Mark van Atten writes:

> Like phenomenology, intuitionism studies essential, structural properties of consciousness, not those of any particular individual's consciousness. Brouwer characterised intuitionism as 'inner architecture' (Brouwer, 1949, 1249), and was interested in the question what mathematical constructions this inner architecture in principle allows, given unlimited memory, time, and so on. [...] To use Noam Chomsky's distinction: Intuitionism does not study the performance of human consciousness in making certain constructions, but its competence. [...]
>
> Like phenomenology, intuitionism recognises that the fundamental notion of subject is not psychological but transcendental. (van Atten, 2015: 260)[100]

It seems to me that the two characterizations of the transcendental subject as 'single' and 'ideal' permit to grasp what Gödel meant by the apparent oxymoron "a priori psychology". Concerning ideality: if the object of psychology are the 'mundane' human subjects, i.e. individual subjects immersed in the world of experience, and intuitionism concerns mathematics as an activity of constitution of objects based on the a priori intuition of time, then intuitionistic logic conceived as a *logique du savoir* cannot be psychology[101]; but it is possible to conceive the study of *the* ideal subject, i.e. of the a priori structure underlying the activity of mathematical constitution: and this would be an 'a priori psychology'. Van Atten's reference to Chomsky in this context is particularly to the point, as it is just in this sense of "psychology" that Chomsky conceives linguistics as a part of psychology. Concerning singularity: Brouwer's claim that «there is no plurality of mind» (Brouwer, 1949: 485) can now be understood not as the manifestation of a solipsistic attitude, but as the expression of the idea that the transcendental subject, i.e. the a priori structure underlying mathematical activity (or, in other terms, our mathematical faculty), is unique; and this idea can be related to Chomsky's repeated claim that, the faculty of language, as well as any other organon of the human being (*the* heart, *the* liver, etc.) is unique, in the sense of invariant through the human species.

A second, and equally significant, reason of convergence is of more general character. Chomsky has repeatedly stressed that empirical sciences in general have undergone a major shift in their standards of intellegibility after Descartes:

---

[100] Van Atten is commenting upon a principle of Brouwer's he himself states in the following way:

> C11. The subject that is correlate to these acts is not a psychological, but a transcendental subject, considered in its essential properties; this motivates the introduction of the notion of a single, ideal subject. (van Atten, 2015: 247)

[101] As a matter of fact, van Atten seems to understand "psychology" in this way:

> The notion of subject that is correlate to Brouwer's acts of mathematical construction is therefore not that of a subject in the world, and cannot be psychological. (van Atten, 2015: 261)

> The idea that there are principles of action and reaction, interaction, growth, development and so on that are just not of the mechanical type – over time, it caused a real shift in the standards of intelligibility for science. It's not the world that's going to be intelligible; we give that up. But the theories have to be intelligible. So we want intelligible theories of the world that we can work with and that meet our epistemological criteria, which are just other aspects of our cognitive system. [...] It's not that people give up the commonsense models [...] But you know that there's a gap, and that your intuitive, commonsense understanding of the world is simply not a guide to what the world is. That's an important change, and it leads in other directions. (Chomsky, 2012: 73)

Applying this to psychology amounts to claiming that psychology has to do with the mental not only because mind is its object of investigation, but also in a sense in which *all* empirical sciences are concerned with the mental: in the sense that it aims at being an intelligible theory, i.e. at elaborating models of (aspects of) world satisfying epistemological criteria which, in turn, result from the structure of our mind. If we take seriously Heyting's view of intuitionistic logic as a *logique du savoir*, i.e. as a theory of the most abstract mechanisms and procedures by which our mind acquires knowledge in specific domains, a consequence of the shift described by Chomsky is that the very logic of empirical sciences should satisfy the epistemological criteria elaborated by intuitionistic logic. In particular semantics, *qua* theory of linguistic competence, should be based on a theory of meaning of intuitionistic inspiration, according to which knowing the meaning of α is not knowing the truth-conditions of α (since such knowledge cannot be described as a mental state), but being able to recognize what constitutes the evidence of α: proofs, in the mathematical domain, *justifications*, in empirical domains. That's why the key notion of the semantics I propose is the notion of justification for α.

Summing up, the reasons of convergence I have tried to point out suggest that intuitionists share with Chomsky, and more generally with a large part of cognitive psychology, a common attitude towards the mental: a realistic one. If this suggestion is accepted, Dummett's view of intuitionism as a variety of antirealism can still be accepted, but only if understood as anti-platonism about mathematical entities and anti-realism about the external word; about mental entities, on the contrary, it seems to exemplify a realistic attitude. For an intuitionist, who equates truth with actual possession of a proof, Dummett's choice of bivalence as a criterion of realism still works: having and not having a proof of α depends on a certain mental entity x being or not being a proof of α, and this does hold according to the intuitionist, since the relation "x proves α" is decidable. From an intuitionistic standpoint, the fact that the *esse* of proofs is their *concipi* amounts to the fact that the laws of evidence are objective in the sense of intersubjectively valid, not arbitrary. In this sense intuitionism is a form of objectivism, not of subjectivism as it is classified on the basis of Dummett's criterion based on the principle of valence (8).[102]

[102] In the conclusion of Chap. 6 I shall explain why I find Dummett's criterion unsatisfactory.

References

Belnap, N. (1962). Tonk, plonk and plink. *Analysis, 22*, 130–134.
Brouwer, L. E. J. (1907). *Over de grondslagen der wiskunde.* Thesis, Amsterdam. (Part. Engl. tr. On the foundations of mathematics. In Brouwer 1975: 13–101.)
Brouwer, L. E. J. (1949). Consciousness, philosophy and mathematics. In *Proceedings of 10th International Congress on Philosophy* (pp. 1235–1249). Amsterdam. (Now in Brouwer 1975: 480–494.)
Brouwer, L. E. J. (1954). Points and spaces. *Canadian Journal of Mathematics*, *6*, 1–17. (Also in Brouwer (1975: 522–38)).
Brouwer, L. E. J. (1975). *Collected works, vol. I: Philosophy and foundations of mathematics.* In A. Heyting (Ed.). North Holland.
Carnap, R. (1934). *Logische syntax der sprache.* Springer-Verlag. (Engl. Tr. *The logical syntax of language*, Kegan Paul, Trench, Trubner & Co Ltd., 1937).
Casalegno, P. (1992). II paradigma di Frege. In M. Santambrogio (Ed.), *Introduzione alla filosofia analitica del linguaggio* (pp. 3–40). Laterza.
Chomsky, N. (1980). *Rules and representations.* Columbia University Press.
Chomsky, N. (2012). *The science of language. Interviews with James McGilvray.* Cambridge University Press.
Church, A. (1936). An unsolvable problem of elementary number theory. *American Journal of Mathematics, 58*(2), 345–363.
Cozzo, C. (1994). *Meaning and argument*. Almqvist & Wiksell.
Dummett, M. (1973). *Frege: Philosophy of language.* Duckworth.
Dummett, M. (1975). The philosophical basis of intuitionistic logic. In H. E. Rose & J. C. Sheperdson (Eds.), *Logic Colloquium '73* (pp. 5–40). North Holland. (Now in Dummett (1978) (pp. 215–247).)
Dummett, M. (1976). What is a theory of meaning? (II). In G. Evans & J. Mcdowell (eds.), *Truth and meaning: Essays in semantics* (pp. 67–137). Clarendon Press. (Now in Dummett (1993) (pp. 34–93).)
Dummett, M. (1978). *Truth and other enigmas.* Duckworth.
Dummett, M. (1982). Realism. *Synthese, 52(1)*, 55–112. (Reprinted in Dummett (1993) (pp. 230–276).)
Dummett, M. (1991a). *Frege. Philosophy of mathematics.* Duckworth.
Dummett, M. (1991b). *The logical basis of metaphysics.* Duckworth.
Dummett, M. (1992). Realism and anti-Realism. In Dummett (1993) (pp. 462–478).
Dummett, M. (1993). *The seas of language.* Oxford University Press.
Dummett, M. (1995). Bivalence and Vagueness. *Theoria LXI, 3*, 201–216.
Dummett, M. (2000). *Elements of intuitionism.* Clarendon Press (First Edition 1977).
Firth, R. (1978). Are epistemic concepts reducible to ethical concepts? In A. Goldman & J. Kim (Eds.), *Values and morals* (pp. 215–229). Reidel.
Frege, G. (1884). *Die Grundlagen der Arithmetik*. Koebner.
Frege, G. (1979). *Posthumous Writings.* In H. Hermes, F. Kambartel & F. Kaulbach (Ed.). Blackwell.
Frege, G. (1980). *Philosophical and mathematical correspondence.* In G. Gabriel, H. Hermes, F. Kambartel, C. Thiel & A. Veraart (Ed.). Blackwell.
Frege, G. (1984). *Collected papers on mathematics, logic, and philosophy.* In B. Mc Guinness (Ed.). Blackwell
Gentzen, G. (1935). Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift, 39*, 176–210, 405–431. (Engl. Tr. Investigations into logical deduction. In Gentzen (1969) (pp. 68–131).)
Gentzen, G. (1936). Die Wiederspruchsfreiheit der Reinen Zahlentheorie. *Mathematische Annalen, 112*, 493–565. (Engl. Tr. The consistency of elementary number theory. In Gentzen (1969) (pp. 132–213).)
Gentzen, G. (1969). *Collected papers.* In M. E. Szabo (Ed.). North Holland.

Gödel, K. (1941). In what sense is intuitionistic logic constructive?. Unpublished Lecture, In Gödel (1995) (pp. 189–200).

Gödel, K. (1958). Uber eine bisher noch nicht benutzte Erweiterung des finiten Standpunktes. *Dialectica, 12*, 280–7. (Engl. tr. by W. Hodges and B. Watson, On a hitherto unexploited extension of the finitary standpoint. *Journal of Philosophical Logic, 9*, 1980, 133–142.)

Gödel, K. (1995). *Collected works, vol. 3*. In Feferman et al. (Ed.). Oxford University Press.

Harman, G. (1986). The meanings of logical constants. In E. LePore (Ed.), *Truth and interpretation* (pp. 125–134). Blackwell.

Heyting, A. (1931). Die intuitionistische Grundlegung der Mathematik. *Erkenntnis, 2*, 106–15. (Engl. tr. The intuitionist foundations of mathematics. In P. Benacerraf & H. Putnam (Eds.), *Philosophy of mathematics* (pp. 52–61). Prentice-Hall, 1983[2].)

Heyting, A. (1934). *Mathematische Grundlagenforschung. Intuitionismus. Beweistheorie*. Springer.

Heyting, A. (1939). Les fondements des mathematiques du point de vue intuitionniste. In F. Gonseth, *Philosophie mathematique. Avec cinq declarations de MM. A. Church, W. Ackermann, A. Heyting, P. Bernays et L. Chwistek* (pp. 73–75). Hermann et C.ie.

Heyting, A. (1956a). *Intuitionism. An introduction*. North Holland.

Heyting, A. (1956b). La conception intuitionniste de la logique. *Les Études Philosophiques, 2*, 226–233.

Heyting, A. (1958). On truth in mathematics. *Verslag van de plechtige viering van het honderdvijftigjarig bestaan der Koninklijke Nederlandse Akademie van Wetenschappen* (pp. 277–279). North Holland.

Heyting, A. (1960). Remarques sur le constructivisme. *Logique et Analyse, 3*, 177–182.

Heyting, A., et al. (1961). Axiomatic method and intuitionism. In J. Bar-Hillel (Ed.), *Essays on the foundations of mathematics* (pp. 237–247). North Holland.

Heyting, A. (1974). Intuitionistic views on the nature of mathematics. *Synthese, 27*, 79–91.

Heyting, A. (1979). Intuizionismo. In *Enciclopedia del Novecento* (vol. 3, pp. 846–55). Istituto dell'Enciclopedia Italiana.

Hilbert, D., & Bernays, P. (1934). *Grundlagen der Mathematik*. Springer.

Kleene, S. C. (1952). *Introduction to metamathematics*. North Holland Publishing Co.

Kolmogorov, A. (1932). Zur Deutung der intuitionistischen Logik. *Mathematische Zeitschrift, 35*, 58–65.

Kreisel, G. (1962). Foundations of intuitionistic logic. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Logic, methodology and philosophy of science* (pp. 198–210). Stanford University Press.

Martin-Löf, P. (1985). On the meaning and justification of logical laws. In C. Bernardi & P. Pagli (Eds.), *Atti degli incontri di logica matematica* (vol. II, pp. 291–340). Università di Siena. (Reprinted in *Nordic Journal of Philosophical Logic, 1*(1), 1996, 11–60.)

Martin-Löf, P. (1987). *Philosophical implications of type theory*. Lectures at the Institute of Philosophy, University of Florence, 15 March–15 May. Manuscript, 1–220.

Martin-Löf, P. (1994). Analytic and synthetic judgments in type theory. In P. Parrini (Ed.), *Kant and contemporary epistemology* (pp. 87–99). Kluwer.

Martino, E. & Usberti G. (1988). The impredicativity of the intuitionistic meaning of logical constants. Manuscript. (Now in Martino (2018) (pp. 147–156).)

Martino, E. & Usberti, G. (1991). Propositions and judgements in Martin-Löf. In G. Usberti (Ed.), *Problemi fondazionali nella teoria del significato* (pp. 125–136). (Now in Martino (2018) (pp. 75–84).)

Piccolomini d'Aragona, A. (2019). *Dag Prawitz's theory of grounds*. Doctoral Dissertation, Universities of Aix-Marseille and Rome La Sapienza.

Prawitz, D. (1965). *Natural deduction. A proof-theoretical study*. Ålmqvist & Wiksell.

Prawitz, D. (1971). Ideas and results in proof theory. In J. E. Fenstad (Ed.), *Proceedings of the second Scandinavian logic symposium* (pp. 235–307). North Holland.

Prawitz, D. (1977). Meaning and proofs: On the conflict between classical and intuitionistic logic. *Theoria, 43*, 2–40.

Prawitz, D. (1978). Proofs and the meaning and completeness of the logical constants. In J. Hintikka, I. Niiniluoto, & E. Saarinen (Eds.), *Essays on mathematical and philosophical logic* (pp. 25–40). Reidel.

Prawitz, D. (1980). Intuitionistic logic: A philosophical challenge. In G. H. von Wright (Ed.), *Logic and philosophy* (pp. 1–10). Nijhoff.

Prawitz, D. (1985). Remarks on some approaches to the concept of logical consequence. *Synthese, 62*, 153–171.

Prawitz, D. (1987). Dummett on a theory of meaning and its impact on logic. In B. Taylor (Ed.), *Michael Dummett: Contributions to philosophy* (pp. 117–165). Nijhoff.

Prawitz, D. (2002). Problems for a generalization of a verificationist theory of meaning. *Topoi, 21*, 87–92.

Prawitz, D. (2005). Logical consequence from a constructivist point of view. In S. Shapiro (Ed.), *The Oxford handbook of philosophy of mathematics and logic* (pp. 671–695). Oxford University Press.

Prawitz, D. (2009). Inference and knowledge. In M. Peliš (Ed.), *The logica yearbook 2008* (pp. 175–192). College Publications, King's College London.

Prawitz, D. (2011). Proofs and perfect syllogisms. In C. Cellucci, E. Grozhols, & E. Ippoliti (Eds.), *Logic and knowledge* (pp. 385–402). Cambridge Scholars Publishing.

Prawitz, D. (2012). The epistemic significance of valid inference. *Synthese, 187*(3), 887–898.

Prawitz, D. (2015). Explaining deductive inference. In H. Wansing (Ed.), *Dag Prawitz on proofs and meaning* (pp. 65–100). Springer.

Prior, A. N. (1960). The runabout inference ticket. *Analysis, 21*, 38–39.

Prior, A. N. (1964). Conjunction and contonktion revisited. *Analysis, 24*, 191–195.

Sundholm, G. (1983). Constructions, proofs and the meaning of logical constants. *Journal of Philosophical Logic, 12*, 151–172.

Sundholm, G. (2020). Kreisel's dictum. In P. Weingartner & H. Leeb (Eds.), *Kreisel's interests: On the foundations of logic and mathematics* (pp. 33–43). College Publications.

Suppes, P. (1957). *Introduction to logic*. Van Nostrand.

Szabó, Z. G. (2020). Compositionality. *Stanford Encyclopedia of Philosophy.*

Tarski, A. (1983). *Logic, semantics, metamathematics: Papers from 1923 to 1938*. In J. Corcoran (Ed.). Hackett Publishing Company.

Troelstra, A. S. (1981). Arend Heyting and his contribution to intuitionism. *Nieuw Archief voor Wiskunde, 3*, XXIX, 1–23.

Troelstra, A. S. (1977). Aspects of constructive mathematics. In J. Barwise (Ed.), *Handbook of mathematical logic* (pp. 973–1052). North-Holland.

Troelstra, A. S., & van Dalen, D. (1988). *Constructivism in mathematics* (Vol. 1). North-Holland.

Usberti, G. (1995). *Significato e conoscenza*. Guerini.

van der Schaar, M. (2011). The cognitive act and the first-person perspective: an epistemology for constructive type theory. *Synthese*, 391–417.

van Atten, M. (2009). On the hypothetical judgement in the history of intuitionistic logic. In C. Glymour, W. Wang, & D. Westerståhl (Eds.), *Logic, methodology, and philosophy of science 13* (pp. 122–136). King's College Publications.

van Atten, M. (2015). *Essays on Gödel's reception of Leibniz, Husserl, and Brouwer*. Springer.

van Atten, M. (2017). The development on intuitionistic logic. *Stanford Encyclopedia of Philosophy.* https://plato.stanford.edu/archives/win2017/entries/intuitionistic-logic-development/

van Atten, M. (2018). *Predicativity and parametric polymorphism of Brouwerian implication.* arXiv.org>math>arXiv:1710.07704v2

Williamson, T. (2000). *Knowledge and its limits.* Oxford University Press.

Wittgenstein, L. (1958). *Philosophical investigations.* Blackwell.

Chapter 3
Epistemic Justifications as Cognitive States

Abstract The aim of this chapter is to introduce the notion of justification, or warrant, as the key notion of the internalist theory of meaning I am going to develop, whose starting idea is that understanding a sentence amounts to having a criterion for establishing what is a justification *for* that sentence. Because of its intended role, justification has to be introduced as a theoretical notion; but it is clear that the formal notion cannot fail to share some characteristics with the intuitive one. In Sect. 3.1 such characteristics are made explicit: the *ex ante* nature of (epistemic) justification, its defeasibility, its non-factiveness. In Sect. 3.2 some other features are introduced as indispensable for a theoretical notion of justification: epistemic transparency, k-factiveness, and its consisting in a cognitive state. In the Conclusion (Sect. 3.3) a more explicit formulation is given of the sort of anti-realism about external world and realism about mental world I am advocating.

Keywords Epistemic Justification · Transparency · Cognitive states · Casalegno · Truth-ground · Internalism

The aim of this chapter is to introduce the notion of epistemic justification, or warrant, as the key notion of the internalist theory of meaning I am going to develop, whose starting idea is that understanding a sentence amounts to having a criterion for establishing what is a justification *for* that sentence. Because of its intended role, justification has to be introduced as a theoretical notion; but it is clear that the formal notion cannot fail to have some systematic relations with the intuitive one; in fact, it must be as faithful as possible to it. There are at least two reasons for this. First, the logical constants occur in natural languages, and a theory of meaning could not pretend to be a theory of meaning if its language did not contain constants translating the logical constants of natural languages; therefore, given a (binary) constant **C** of the theory, the inductive clause that defines justifications for $\alpha\mathbf{C}\beta$ must be as faithful as possible to the meaning of the natural language constant of which **C** is the intended transalation. Second, the inductive definition of "justification for α" will contain a base clause, for the case where α is atomic; for this clause to be a plausible characterization of the class of *justifications* for α (as opposed to a class

G. Usberti, *Meaning and Justification. An Internalist Theory of Meaning*, Logic, Epistemology, and the Unity of Science 59,
https://doi.org/10.1007/978-3-031-24605-0_3

of entities of any other sort), it will be necessary to make reference to the intuitive notion of justification (for an atomic sentence).[1] We have therefore a problem of *material adequacy* of the definition we are going to give; that is to say, we have to ask ourselves when such a definition will be in agreement with our intuitions about the notion of justification. So, first of all, we must make explicit such intuitions; this will be the task of Sect. 3.1.

On the other hand, it is equally clear that the theoretical notion must satisfy the internal needs of the theory of meaning we are going to build: we have seen in Chap. 1 how far the key notions of an internalist semantics conceived as a computational-representational theory must be from the intuitive notions of truth and denotation. We must therefore be prepared for the eventuality of an internal tension between the two exigencies of material and theoretical adequacy; it is therefore necessary to consider both the intuitive notion of justification, and the theoretical notion we need, in order to see whether it is possible to reach a balance between the two exigencies; this will be the task of Sect. 3.2.

In the Conclusion (Sect. 3.3) a more explicit formulation is given of the sort of anti-realism about external world and realism about mental world I am advocating.

3.1 The Intuitive Notion of (Epistemic) Justification

As we have seen in Chap. 2, Sect. 2.4.3.1, there are two senses in which a belief can be epistemically justified: a propositional one, according to which a subject $\mathcal{S}$ has propositional justification for believing a proposition p whenever it is epistemically appropriate for $\mathcal{S}$ to believe p, whether or not $\mathcal{S}$ actually believes p; and a doxastic one, according to which $\mathcal{S}$ is doxastically justified in believing p whenever $\mathcal{S}$ has propositional justification for believing p and actually believes p in virtue of that justification. As far as I know, the first to distinguish these two senses was Roderick Firth, in the passage of Firth (1978) quoted in Chap. 2. Coming back to Firth's distinction, Alvin Goldman has introduced a terminology I shall adopt because I find it more telling (Goldman, 1979: 21).

3.1.1 Ex Ante *and* Ex Post *Justifications*

Instead of doxastic and propositional justification, Goldman speaks of *ex post* and *ex ante* justification, respectively; the reason is that, in order to be doxastically justified, it is necessary to have a belief in advance, so that the problem of the justification

[1] An analogous remark applies to Tarski's definition of truth. In the case of such an atomic sentence as "Snow is white" the base clause stipulates that it has the property T iff the object denoted by "snow" belongs to the set denoted by "white"; but in order that property T can plausibily be equated with the property of being *true* it is necessary to appeal to two fact: (i) that the set denoted by "white" is the set of white things, and (ii) that the object denoted by "snow" is snow.

(or better justifiedness) of belief can only arise *ex post*, that is after the belief has been adopted; whereas a proposition can be justified without being believed, so that the justification one may have to believe a proposition can only be *ex ante*, in the sense that the question whether something is or is not a justification for the sentence expressing that proposition must be answered before the belief has been adopted. The two types of justification play typically different, and in a sense complementary, roles: while *ex ante* justifications *orient* the belief-forming processes, in the sense that they guide a subject in selecting the propositions to believe, the problem of *ex post* justifiedness only concerns processes that have been accomplished, for example when it is necessary to verify their adequacy to their purpose or to test their reliability.

It is evident that when I said at the beginning that the starting idea of the theory of meaning I am going to propose is that understanding a sentence amounts to having a criterion for establishing what is a justification *for* that sentence (i.e. to believe the proposition expressed by it), I was referring to *ex ante* justifications. For a justification *for* a sentence is intuitively something warranting a subject to believe the corresponding proposition *independently of the fact that the subject believes it*. As a consequence, if we want our starting idea not to be deprived of sense, we must make sure that the notion of *ex ante* justification plays an irreplaceable role, in particular that it is not reducible to the notion of *ex post* justifiedness. In point of fact, the thesis that *ex ante* justifiedness is reducible to *ex post* justifiedness is precisely an essential ingredient of the well known theory of justifiedness developed by Goldman (Goldman, 1979: 21 f.); it is therefore necessary to examine Goldman's theory.

He proposes an explanation of justifiedness that applies in a natural way to *ex post* justifiedness; then he tries to extend his explanation to *ex ante* justifiedness by showing how the latter can be defined in terms of the former. Let us see first how he suggests to carry out this reduction:

(1)

> «Person S is *ex ante* justified in believing p at t if and only if there is a reliable belief-forming operation available to S which is such that if S applied that operation to his total cognitive state at t, S would believe p at t-plus-delta (for a suitably small delta), and that belief would be *ex post* justified.» (Goldman, 1979: 21.)[2]

Of course, in order for this reduction to work it is necessary that the definition of *ex post* justifiedness does not make reference to the notion of *ex ante* justification, otherwise it would enter into a vicious circle. Let us therefore see how Goldman defines *ex post* justifiedness.

His basic idea can be summarized in two points: (i) the justifiedness of a belief depends on features of the psychological processes and/or the methods leading to the formation/fixation of that belief; (ii) the crucial feature a process or a method must have in order to produce a justified belief is its being *reliable* or generally held reliable, where «reliable» means that the beliefs it produces are generally true.

[2] In Goldman's (a bit misleading) terminology one can be justified *in believing* p even if one does not actually believe p.

But this is only a first approximation to the definition of justified belief; according to Goldman it needs a substantial amendment. Let us see why. He suggests the following counterexample to his former definition:

> Suppose that Jones is told on fully reliable authority that a certain class of his memory beliefs is almost all mistaken. His parents fabricate a wholly false story that Jones suffered from amnesia when he was seven but later developed pseudo-memories of that period. Though Jones listens to what his parents say and has excellent reason to trust them, he persists in believing the ostensible memories from his seven-year-old past. Are these memory beliefs justified? Intuitively, they are not justified. But since these beliefs result from genuine memory and original perceptions, which are adequately reliable processes, our theory says that these beliefs are justified. (Goldman, 1979: 18)

Goldman's diagnosis is that, although Jones has strong evidence against certain propositions concerning his past, he doesn't *use* this evidence, that is, he fails to do something which, epistemically, he could and should have done. Of course, the reference to what Jones could and *should* have done from an epistemic point of view is not legitimate for Goldman, who wants to explain in non-epistemic terms when a belief is justified (Goldman, 1979: 13). Consequently he proposes to modify his former (sufficient) condition of justifiedness in the following way:

(2)

> «If S's belief in p at t results from a reliable cognitive process, and there is no reliable or conditionally reliable process available to S which, had it been used by S in addition to the process actually used, would have resulted in S's not believing p at t, then S's belief in p at t is justified.» (Goldman, 1979: 20)

Here every reference to what, epistemically, should have been done has been eliminated in favour of (existential) quantification on processes and the demand that these are available—where the notion of availability, or accessibility, is evidently a non-epistemic one (Goldman, 1988: 57–58).

The problem with the condition in (2) is that it is *too* restrictive, in that it excludes situations that we should intuitively classify as cases of justified belief. Consider for instance the following modification of Goldman's example. Suppose (i) that the story of his amnesia has been told to Jones not by his parents but by a distant relative of his who during Jones' childhood never lived near him and about whose reliability he has no motivated opinion; (ii) that in a drawer of Jones' bedroom there is a letter he has never read, which reveals that his distant relative is a quite reliable person and that the story she told him is in fact true. In this case our intuition clearly suggests that Jones' trust in his memories from his seven-year-old past *is* justified: having no reason to trust his distant relative rather than his own memories, he chooses to rely upon the latter, as any reasonable person would do. But, on the basis of Goldman's condition (2), Jones is *not* justified, since *there is* a *reliable* cognitive process, *available* to him (namely, reading the letter), such that, had it been used by Jones in addition to the process he actually used, would have resulted in Jones not believing his remote memories. As a consequence the condition stated in (2), even if it could be proposed as a sufficient condition of justifiedness,[3] could not be suggested as a *necessary* condition.

The moral to draw from this counterexample seems to me clear. The reason why we intuitively feel that Jones is justified in trusting in his memories from his seven-year-old past is that, in spite of the fact that the process of reading the letter is available to him, we do not consider that process as *epistemically relevant* for him to be justified in believing what he believes, whereas his relying upon his memories from his seven-year-old past *is* relevant. Of course, it is not easy at all to explain *why* we have such intuitions; but the intuitions themselves are very clear.[4]

My suggestion is therefore that Goldman's condition in (2) cannot be offered as a necessary condition of justified belief unless the class of cognitive processes on which the existential quantifier operates is restricted with an extra condition to the effect that these processes are *epistemically relevant to the proposition p which* S *believes*. The important thing to notice here is that the epistemic relevance of a process (or a method) can be evaluated by a subject *ex ante*. Consider for instance the epistemic relevance of Jones' distant relative's story; if it could be established by Jones only *ex post*, in particular after he has acknowledged its reliability and the unreliability of his remote memories, it is clear that his trust in his own memories would *not* be justified, whereas intuition definitely suggests that it is.

In conclusion, for our definition of *ex post* justifiedness to be adequate we must make reference to the epistemic relevance of belief-forming processes, and this relevance (besides being a clearly *epistemic* notion) can be tested *ex ante*; therefore an adequate definition of *ex post* justifiedness makes implicit reference to the notion of *ex ante* justification, and the reduction proposed by Goldman does not work.[5]

3.1.2 Non-conclusiveness

It is often remarked that the intuitive justifications we normally have for empirical sentences are non-conclusive; on the other hand, we have seen in Sect. 2.4.4 of Chap. 2 that the neo-verificationists, and in particular Prawitz, opt for the adoption of a conclusive notion of justification as the key notion of an anti-realistic theory of meaning extended to empirical sentences; it is legitimate to ask for the rationale of this choice, and to bring out some of its consequences.

First of all it is useful to distinguish two senses of "conclusive":

> [S]ometimes it means *factive*, sometimes *indefeasible*. A kind of verification is factive just in case necessarily only true statements have verifications of that kind. A kind of verification is

---

[3] In fact Peacocke (1992: 810), provides an example that casts doubt on its being a sufficient condition.

[4] I shall sketch an explanation in Chap. 4.

[5] For a similar argument against the reducibility of propositional justification to doxastic justification see Kvanvig & Menzel (1990).

> indefeasible just in case necessarily any statement with a verification of that kind continues to have such a verification whatever new information is received. (Williamson, 1998: 67)[6]

"Non-conclusive" may therefore be understood either as defeasible or as non-factive. In this section I will argue that the intuitive justifications we normally have for empirical sentences are defeasible through a discussion of the neo-verificationist choice of an indefeasible notion of justification; but before that it may be interesting to discuss the relations between factiveness and indefeasibility.

According to the realist the two senses are different; for example, Williamson writes:

> Factiveness does not entail indefeasibility. Knowing *p* is always a factive way of having warrant to assert *p*; it is almost never an indefeasible way. New evidence can almost always undermine old knowledge [...].
>
> By itself, indefeasibility does not entail factiveness. If warrant to assert *p* consisted merely in good reason to believe *p*, then the inhabitants of a universe created six thousand years ago with every appearance of having existed for millions of years might have an indefeasible non-factive warrant to assert that they are not inhabitants of a universe created six thousand years ago with every appearance of having existed for millions of years. (Williamson, 2000: 265–266)

The position of the antirealist is not univocal, due to the fact that there are different anti-realist attitudes towards the notion of truth, in terms of which factiveness is defined. I shall discuss the neo-verificationist and intuitionistic notion(s) of truth in Chap. 6; here I only remark that Williamson's counterexample to the second entailment is acceptable only to someone who accepts the whole realist notion of truth, hence not to an anti-realist.[7]

Let α be the sentence "the universe we inhabit was created six thousand years ago", and β the sentence "the universe we inhabit has every appearance of having existed for millions of years"; then $(\alpha\wedge\beta)$ is true, since β is true and, by stipulation, also α is. But - Williamson says - we have a good reason, hence a justification j, to believe $\neg(\alpha\wedge\beta)$. What would j look like? I guess like a justification for $\beta \rightarrow \neg\alpha$. Since, by β, there is no appearance that would defeat j, j is indefeasible; but is j non-factive? The only reason to say that it is non-factive is Williamson's stipulation that α is true; but an antirealist might object that, after all, j is exactly the justification actual science has to believe that α is false, and that to assert that α is true we should have a reason, hence an 'appearance' that would defeat j - which by hypothesis is not

[6] It may be useful to define factiveness and indefeasibility in relation to possession of warrant; here are Williamson's definitions:

> Define a way of having warrant to assert *p* to be defeasible just in case one can have warrant to assert *p* in that way and then cease to have warrant to assert *p* merely in virtue of gaining new evidence. A way of having warrant to assert *p* is indefeasible just in case it is not defeasible. [...] A way of having warrant to assert *p* is factive just in case a necessary condition of having warrant to assert *p* in that way is that *p* is true. (Williamson 2000: 265).

I shall use either definition according to convenience.

[7] See Chap. 2, Sect. 2.5.1.

available. Hence Williamson's stipulation is acceptable only if we accept a notion of truth independent of the existence of any justification.

Another counterexample to the same effect might be the following.[8] Suppose that close to Nelson at the point of death there were only two persons, Hardy and Mohammed, that Hardy heard him say: «Kiss me, Hardy», and Mohammed heard him say: «Kismet, Hardy». After Nelson's death Hardy and Mohammed probably believed, respectively, (3) and (4):

(3) «Kiss me, Hardy» were Nelson's last words
(4) «Kismet, Hardy» were Nelson's last words,

and we feel both justified in believing their respective propositions. In this situation they could not acquire new relevant information (and to disregard the irreversibility of time does not seem to be a 'good' idealization); so both Hardy's and Mohammed's justifications are indefeasible. On the other hand, (3) and (4) are incompatible, i.e. they cannot be both true; so the proposition expressed by one of them is false; hence the justification for *that* proposition is indefeasible and non-factive. The intuitionist may retort that the realist has not exhibited a sentence α such that α is false and indefeasible: he has exhibited *two* indefeasible and incompatible sentences; then he has invoked the principle $\neg(\alpha\wedge\beta) \rightarrow (\neg\alpha\vee\neg\beta)$ to conclude that one of the two sentences (we don't know which one) is false; but that principle is intuitionistically invalid: its acceptance crucially depends on the acceptance of the principle of bivalence, hence of the concept of classical truth.

The moral to draw from this discusion is that, until a coherent anti-realist sense is given to "true", the only clear sense an anti-realist can assign to "non-conclusive" is the sense of "defeasible"; hence, the remark made at the beginning of this section is more properly rephrased by saying that the intuitive justifications we normally have for empirical sentences are defeasible. The defeasibility of empirical justifications, i.e. of justifications for empirical sentences, is described by Prawitz in the following passage:

> If a sentence is asserted in mathematics on the basis of what one thinks is a proof of it and it later turns out that the sentence is false, one would ordinarily say that the alleged proof was not a proof and that therefore the sentence was incorrectly asserted. But outside mathematics, one may want to say that a sentence was correctly asserted (on sufficient ground) although it later turned out that the sentence was false, i.e., the grounds on which the sentence was asserted are still regarded as having been sufficient in the situation in question (although they are not so anymore). (Prawitz, 1980: 8)

The possibility envisaged here is the following: (i) a subject $\mathcal{S}$ has at time *t* a justification j for a proposition α; (ii) at *t'* α turns out to be false, hence $\mathcal{S}$ has no justification for it; but (iii) it still holds at *t'* that $\mathcal{S}$ had the justification j for α *at time t*; hence j was a justification for α at t, and is no longer a justification for α at t': by definition, j is defeasible, and this is the crucial difference from what happens within mathematics, where the fact (ii) is a reason to infer that *already at t* j was not a justification for α.

[8] The example is a modification of one given by Dummett (1991a, 1991b: 267) for different purposes.

Prawitz explicitly refuses this possibility for several reasons, which will be discussed in due time[9]; here I would like to stress the costs of Prawitz's adoption of an indefeasible notion. His proposal is sketched in the following passage:

> It is sometimes maintained that it is an illusion to speak at all about conclusive evidence, even for mathematical sentences *and observation sentences*, since we can never rule out the possibility that we have made a mistake. But in the case of a mistaken assertion of a mathematical sentence *or an observation sentence* made in good faith, the reasonable thing to say is that what we thought was a certain proof *or sense-experience* turned out not to be a proof *or a sense-experience of the appropriate kind*. (Prawitz, 1987: 145. *Italics mine.*)

Consider for instance the candidate Prawitz proposes as justification for the observation sentence

(5) It is raining:

the experience of seeing that it rains; as we know that such experience might be the result of an hallucination, or of the electrical stimulation of some points of the cerebral cortex, and so on, there is no reason to propose it as a justification for (5). Presumably Prawitz would say that hallucinating that it rains is not «a sense-experience of the appropriate kind»; but this would not give us the reason we are looking for. As we have seen, the fundamental role of epistemic justification is to guide subjects about what to believe; the point is that, for the experience of seeing to be apt to play this guidance role, there should be some feature of the *experience* of seeing that it rains, *available to the subject who has it*, that permits the subject to discriminate between that experience and the experience of hallucinating that it rains; but there is no feature of this sort: the two kinds of sense-*experience* are identical. Concluding, if we propose indefeasible sense-experiences as justifications for observation sentences, we have no reason to deem justified any observation sentence, since there is no way, for a subject having them, to distinguish indefeasible from defeasible ones. This is the first drawback.

A second one concerns atomic empirical sentences of other sorts. In these cases we usually convince ourselves of their truth through some kind of inference to the best explanation[10]; but an explanation of α can lose its status of best explanation as new information is acquired; if we assume that empirical justifications are indefeasible, we are no longer in a position to explain empirical inferences as inferences to the best explanation. Moreover, most examples of Gettier problems are just cases of justified belief acquired by abduction: to deny that they are cases of justified belief seems too easy a solution to those problems.

A third drawback is a consequence of the first. According to Prawitz,

> In mathematics we operate with the notion of conclusive proofs although of course we may also make mistakes there, in which case we say that what we thought was a proof was not really a proof. Similarly I would say that we may think to have seen something but that it later turns out that we did not see it. (Prawitz, 2002: 90–91)

[9] See Sect. 3.2.2 and Sect. 3.2.3.

[10] I shall analyze this sort of justifications in the next chapter.

In other terms, "seeing" must be taken in a veridical sense; as a consequence, a subject $\mathcal{S}_1$ who believes that it is raining because she has the visual experience of presently falling rain (while she is actually exposed to a cortical stimulation) is making a mistake, exactly like a subject $\mathcal{S}_2$ who believes a mathematical proposition α because he accepts an argument for α containing a flaw. It seems to me that the analogy established by Prawitz is misleading, since it neglects the crucial difference between the two cases: the flawed argument presents some feature, available to $\mathcal{S}_2$, that would have permitted him to realize that it is not correct; on the contrary, no such feature is available to $\mathcal{S}_1$ to distinguish seeing from hallucinating, as we saw above. The availability of the feature is essential from the epistemic point of view: because of it we can say that $\mathcal{S}_2$ has neglected, or has not paid attention to, some information he *should* have taken into consideration; whereas $\mathcal{S}_1$ *is* intuitively justified in asserting (5), and has made no mistake, since there is no information she *should* have taken into consideration.

I do not mean that the analogy postulated by Prawitz between mathematical and perceptual, or more generally empirical, domains is in some sense inconsistent; my opinion is that it is seriously misleading because it obliterates a discrepancy that is significant. In mathematics it is quite natural to say that something *is* or *is not* a proof of α, without further qualifications; in an empirical domain we find it much more natural to say that something is or is not a justification (for α) *for a subject*. For example, if a student believes a mathematical proposition α because someone gave him an argument for α, it is natural for an observer to ask whether the student is right or wrong *tout court* in believing α. On the contrary, if a subject $\mathcal{S}_3$ believes an empirical proposition β - for instance, that John is married—because someone shew him John's marriage certificate, it is natural for an observer to say that $\mathcal{S}_3$ is right (namely justified) in believing that John is married, but equally natural to say that a subject $\mathcal{S}_4$ is right in believing that John is not married because (s)he believes that John's marriage certificate is a fake; in other terms, it is natural for the observer to say that John's marriage certificate is a justification for β *for* $\mathcal{S}_3$, but not *for* $\mathcal{S}_4$.

It seems to me that the reason (or one of the reasons) of this discrepancy lies in a deeper difference between our intuitions about mathematical and empirical evidence: the following principle is felt as true of the former, while it is felt as blatantly false of the latter:

(6) Every subject is in a position to acquire all relevant information about evidence for α.

If α is a mathematical sentence, we see no *a priori* limitation to the possibility of acquiring new relevant information about proofs of it, so all subjects are ideally considered on a par with respect to their being in a position to establish whether something is or is not a proof of α. The situation is completely different if α is an empirical sentence: in the case of the two subjects $\mathcal{S}_3$ and $\mathcal{S}_4$ introduced a moment ago, it may happen that for one of the two, or for both, it is impossible to acquire new relevant information about John's marriage certificate; if this is the case, to say that neither subject is justified to believe her/his respective proposition seems to be an illegitimate regimentation of linguistic usage; nor does it seem better to say that only one is justified, since there is no difference between the quality of $\mathcal{S}_3$'s and $\mathcal{S}_4$'s information about John's marriage certificate that can justify such different qualifications; we are rather inclined to say that "John is married" is justified *for*$\mathcal{S}_3$ and "John is not married" is justified *for*$\mathcal{S}_4$: in empirical domains we admit *conflicting* pieces of evidence, i.e. justifications for incompatible propositions, stemming from different individual locations in space or in time, from different and 'incomparable' amounts of available information, and so on.

A consequence of this disparity between empirical and mathematical domains is that, when we find ourselves in a situation of informational lack, we follow typically different epistemic maxims in the two domains. In mathematics we refrain from adopting a belief; in the case of $\mathcal{S}_2$, for example, we feel that he *ought* to abstain from believing that a proof of α had been presented to him. The rationale for this maxim seems to be the fact that principle (6) is true in mathematics (where evidence is conferred by proofs), so that the possibility is open to the subject of acquiring new relevant information, and it is rational for her/him not to adopt a belief that might be discarded. In the empirical domain, on the contrary, we do not abstain from belief: we normally adopt the belief we are warranted to adopt by available information; in the case of the subjects $\mathcal{S}_3$ and $\mathcal{S}_4$, we feel that *both* are justified in believing their respective propositions. The rationale for this maxim is—I suggest—the fact that principle (6) is not true in the perceptual domain (where evidence is conferred by justifications): there are propositions such that there is no possibility for a subject to acquire new relevant information about evidence for them,[11] and in such cases it is often more useful to *have* a belief than to have none[12]; consequently it is rational to adopt a belief even if later it might be discarded.

I have argued that the intuitive truth of (6) in the case α is a mathematical sentence, and its falsity in the case α is empirical, is the reason of our intuition that the notion of mathematical proof of α is absolute, while the notion of justificaton is relative to subjects; it may be interesting to investigate about the *reasons* of the intuitive truth and falsity of (6) in the two cases. It seems to me that one reason is that, in the case of $\mathcal{S}_2$, we feel that there must be some *intrinsic* feature of the argument such that, had $\mathcal{S}_2$ noticed it, he would have realized that it was not a proof of α. "Intrinsic" means that it was a feature of the argument the subject could have known through a careful analysis of the argument itself, without resorting to other, external pieces of information: that feature was not *immediately* available, but it *was* available, and he did not notice it because of some lack of attention or memory, or because he unduly trusted some of his informants, etc.; he therefore failed to do his epistemic duty, according to which he ought to have refrained from believing that a proof of α had been presented to him until he had carefully analysed all the available data. On the contrary, in the case of $\mathcal{S}_3$ and $\mathcal{S}_4$ there seems to be no *intrinsic* feature of the experience of one of them such that, had (s)he noticed it, (s)he would have realized that her/his experience was not an experience of (veridical) seeing; the only way for

[11] Think of the Hardy-Mohammed example given above.

[12] For example, it may be very useful to have a belief about who will win the next election, even if the results of the exit-polls are far from guaranteeing that the winner will be Mr. X.

her/him to realize it would consist in acquiring new relevant information, which is *not* available in her/his present cognitive state. Therefore, if one subject believes that John is married and the other that John is not married, neither has infringed her/his own epistemic duty, and both their beliefs are justified.

3.2 Requirements for a Theoretical Notion of Justification

I have extracted from the intuitive notion(s) of justification some characteristics that a theoretical notion should share in order to be capable of fulfilling the role of key notion of a theory of meaning. Now I will introduce other characteristics that a theoretical notion should have in order to fulfill that role, independently of their being shared by the intuitive notion(s).

3.2.1 Epistemic Transparency

If a theoretical notion of justification for a sentence is introduced as the key notion of a theory of meaning of empirical and mathematical sentences, in the sense that knowing the meaning of α is equated to being capable to recognize justifications for α, then justifications for α must be epistemically transparent. The reasons for this requirement of transparency are the same as the ones given in Chap. 2 for an analogous requirement imposed onto proofs of mathematical sentences, and I will not repeat them here. I only add that a consequence of this requirement is that there is not a point of view from which something can be judged *to be* a justification for α in spite of the fact that no subject who knows the meaning of α is in a position to recognize that it is. To justifications the same *dictum* applies as Dummett used for intuitionistic proofs: their *esse* is their *concipi*. Of course it may happen that an empirical subject mistakes something for a justification, or that the subject doesn't realize that something is a justification; but this is a consequence of limitations of memory, attention, and so on, from which we make abstraction when we appeal to an *idealized* subject.

Someone might object that, although epistemic transparency is a necessary characteristic of a theoretical notion of justification to which the role of key notion of a theory of meaning is conferred, it makes this notion so different from the intuitive one that it is no longer plausible to call it "justification". My answer is, first, that resemblance to intuitive notions cannot be the only criterion orienting our search for a theoretical notion; it is clear that intuitive notions may turn out to be contradictory, when we try to subsume them into the conceptual framework of a science; a shocking example is the intuitive notion of reference, as we saw in Chap. 1, if we accept Chomsky's arguments. Secondly, when we consult intuition about fundamental notions, it does not give us so definite answers as the objector implies. What is, for instance, *the* intuitive notion of possibility, of cause, of set, of probability, or of truth? If there were a unique answer there would simply be no space for so many philosophical discussions as there still are. Often, all that a philosophical discussion can do is bring to light the fact that what seemed to be one intuitive notion is in fact a cluster of notions bearing a family resemblance. Sometimes it is sufficient to make them explicit in order to extricate notions that are simply different; it is the case of possibility: there is not ‘the’ intuitive notion of possibility, but there are logical, deontic, epistemic possibility, and so on; and no one is ‘more intuitive’ than the others. Sometimes things are more complicated; in the case of truth, for example, after having extricated the metaphysical notion from the epistemic ones, we have to do with notions that are not simply different: they vie for the role of ‘correct’ notion of truth, or for the role of fundamental notion of the theory of meaning. In such cases (and I hold that the case of the notion of justification is similar) it is not by appealing to intuition that we can settle the question; what is decisive are considerations concerning the nature of the theories we can build on the basis of each notion, their coherence, their explicative power, and so on—provided that the choice of a notion as fundamental has clear motivations, and that the notion is intelligible to anyone who has different philosophical views.

In the present case, we have seen a decisive reason for an epistemically transparent notion of justification: its role in the intuitionistic theory of meaning; on the other hand, such a notion should certainly be intelligible to a realist, because the use of epistemically transparent basic notions seems to be ineludible also for the realist in several theoretical domains. I shall only mention some of them. First, even if mathematics is conceived as an activity of discovery of a realm of entities *an sich* subsisting, the mathematician has to be absolutely confident in the reliability of proofs, the very tools by means of which mathematical truths are discovered; and when a proof is based on some essentially *new* method, about whose reliability we simply have no evidence, there seems to be no other way of obtaining such a confidence than by postulating the epistemic transparency of proofs. Second, in the theory of explanation

> [we] do not appear to know how to make the contrast between understanding and merely seeming to understand in a way that would make sense of the possibility that most of the things that meet all our standards for explanation might nonetheless not really explain. (Lipton, 2004: 22)

Third, I have remarked in Chap. 1 that semantic competence includes knowledge of entailment relations, and that, in the context of a linguistic explanation, a language user’s knowledge of an entailment relation between α and β must be conceived as independent of knowledge of the truth-values of α and β. This means that the subject’s knowledge of an entailment relation must be conceivable as the result of a computational process, hence that it cannot happen that an entailment relation subsists unless the subject is in a position to recognize that it does. This suggests - and this is a fourth example - that Tarski’s definition of logical consequence is of no explanatory value in the justification of inference, as Etchemendy observed in the passage quoted in the Conclusion of Chap. 1; again, epistemically transparent notions seem to be necessary in order to justify inference.[13]

3.2.2 Justifications as Cognitive States

In Sect. 3.1.2 I have argued for the necessity of basing a theory of meaning for empirical sentences on a defeasible notion of justification. In this section I consider two objections to this idea.

3.2.2.1 Casalegno's Argument

The first objection been formulated by Paolo Casalegno (Casalegno, 2002).[14] Literally, Casalegno argued that the idea of non-conclusive assertibility conditions, as it is usually understood by the verificationists, is inconsistent; but his argument can be easily rephrased as directed against the notion of non-conclusive justification for a sentence. He first defines defeasible assertibility conditions in the following way:

> To say that C is non-conclusive means that the following is possible:
>
> (i) at a time t X believes that C, and therefore feels entitled to assert "S";
>
> (ii) at a later time t' X is still convinced that at t it was the case that C and that therefore he was then entitled to assert "S", nevertheless, because of new information acquired in the meantime, at t' X no longer believes that at t it was the case that S and is therefore ready to withdraw an assertion of "S" made at t. (Casalegno, 2002: 76)

Then he argues that «saying that C is a non-conclusive assertibility condition is virtually equivalent to saying that C is *not* an assertibility condition.» (Casalegno, 2002: 78) He considers an example:

> Assume that the presence of puddles in the streets is an assertibility condition of the sentence "It has been raining" for John. […] At time t John leaves the house and sees puddles in the streets; since he believes that there are puddles in the streets, he feels entitled to assert "It has been raining". At a later time t' he is told that, as a matter of fact, it has *not* been raining and that the puddles are there because during the night the streets have been washed. He believes what he is told and as a consequence he withdraws the assertion made at t. In this case John withdraws at t' the assertion made at t, but at t' he has not changed his mind as to the fact that at t the relevant assertibility condition was satisfied and that he was therefore entitled to make that assertion. This case shows that the assertibility condition consisting in the presence of puddles in the streets is indeed non-conclusive. (*Ibid.*)

We are therefore confronted with the situation described in the passage from Prawitz (1980) quoted in Sect. 3.1.2. The problem arises at this point:

---

[13] An interesting discussion of the role of epistemic transparency within realist theories of content is contained in Boghossian (1994); see also Boghossian (2011).

[14] Casalegno's objection has been explicitly endorsed by Prawitz in Prawitz (2002), fn. 1.

> So far so good. But now notice that the information which, in the situation just described, John acquires at *t*' could have been available to him already at *t*: in other words, John could have already been informed that the streets had been washed when he left the house and saw the puddles. Also notice that, if this had been the case, his seeing that there were puddles in the streets and his consequent belief that there were puddles in the streets would have not produced in him the belief that it had been raining and it would have not made him feel entitled to assert "It has been raining". Since all this is perfectly possible, it is false that John believes that it has been raining and feels entitled to assert "It has been raining" whenever he believes that there are puddles in the streets. But then, after all, the presence of puddles in the streets cannot be an assertibility condition of "It has been raining" for John. (Casalegno, 2002: 79)

Of course the problem arises even if we do not equate the meaning of a sentence to its assertibility conditions, but to a criterion for distinguishing its justifications, because it concerns the very notion of justification, in virtue of the obvious intuitive connection there is between justification and assertibility.[15]

The central step of Casalegno's argument is based on the assumption that empirical justifications, as he conceives them, bear modal properties, in the sense that it can be meaningfully asked, of a certain justification j, whether in another possible situation it has a certain property (for instance, making a subject feel entitled to assert a certain proposition); and this assumption, in turn, rests on another, and crucial, one: that the very same 'thing'—the presence of puddles in the streets—is a justification for, hence an assertibility condition of, the sentence.

(7) It has been raining

at t, and not at t'. As a consequence, if we choose, as a justification for the assertion of that sentence, something that cannot remain the same as the time passes, the argument is blocked. If we take possible situations as indiciated by times, situations change as time passes; and what certainly changes in passing from one situation to another is the situation itself. "Situation" must be understood, in this context, as all information available to the subject about that situation—in other terms, as the *cognitive state* of the subject. Hence what changes as time passes are the cognitive states of the subject. The way out consists therefore in conceiving of justifications as cognitive states. Anticipating the definition that will given in the next chapter and drastically simplifying, we could say that a cognitive state σ is a justification for (7) iff the hypothesis that it has been raining is the best explanation of information I available in σ (in symbols, BE((7),I)), where I includes the piece of information that there are puddles in the streets. Casalegno's argument would apply if the presence of puddles in the streets were the assertibility condition of the sentence (7), i.e. if, in general, C were defined as the assertibility condition of a sentence α iff BE(α,C). But this is not the view I am suggesting; according to my proposal, it is *the whole cognitive state* σ that is an assertibility condition of α iff BE(α,I), where I is the total amount of information available in σ. As a consequence, it is not true, of *this* assertibility condition, that *it* may not make the subject feel entitled to assert (7): the obtaining or

[15] The connection may be provisorily stated by saying that α is assertible by a subject $\mathcal{S}$ iff $\mathcal{S}$ has a justification for α. This connection will be investigated in the next section.

not obtaining of the relation "BE(α,I)" depends (on the subject's cognitive apparatus and) on information the subject *actually* has in σ, not on information he *may* acquire (not, for instance on the piece of informtion that during the night the streets have been washed): by definition, to acquire new information is to attain a *different* cognitive state; and what is crucial in order to establish whether a cognitive state is a justification for a sentence is information *actually* present in that state, not information *potentially* attainable. It would be senseless to say, of the former cognitive state σ, that *it* has lost its status of justification at t': a cognitive state gives rise to another one as new information is added, and in general there is no question of one and the same cognitive state undergoing a transformation. The argument is therefore blocked.

At the end of his paper Casalegno observes:

> In the foregoing, I have assumed that assertibility conditions [...] are to be conceived as something which can be made explicit and characterized verbally. Admittedly, without this assumption my arguments have no force; but the assumption is – I think – an essential one for a verificationist, and it goes far beyond what is obvious and undisputable. (Casalegno, 2002: 84)

I agree with the first remark. Actually, the assertibility condition of an empirical sentence, as I have suggested to conceive it, is *not* something which, in general, can be made explicit and characterized verbally. It is a cognitive state, and cognitive states cannot in general be identified with lists of known propositions.

As for the second remark—that the assumption in question is essential for a verificationist—I think it is necessary to make a distinction. The verificationism Casalegno has in mind—essentially, Dummett's neo-verificationism—has two main components: an anti-realistic inspiration, coming *grosso modo* from mathematical intuitionism, and a use theory of meaning, coming from the later Wittgenstein and Gentzen. The two components are in principle independent of each other, in the sense that neither a use theory of meaning entails an anti-realistic attitude about the external world, nor does such an anti-realist inspiration entail a use theory of meaning.[16] Casalegno's argument is essentially addressed to the use-theory-of-meaning component, as is clear from the reason he gives of why the assumption is essential for the verificationist:

> The reason is precisely that what verificationists have in mind when they speak of "assertibility conditions" is not the fact that our brain is so programmed that, on some occasions, we happen to feel confident that a certain sentence is assertible, but rather the idea that our capacity to recognize the situations in which a sentence is assertible is the result of our having acquired a body of knowledge whose content could in principle be made fully explicit. (*Ibid.*)

This idea is common to all the supporters of a use theory of meaning, whether realist or anti-realist. As we have seen in Chap. 2, the neo-verificationists *are* supporters of a use theory of meaning, in particular of the idea that the meaning of the logical constants is to be explained in terms of rules for their correct use in inferential practice: by means of such rules arguments are built, and proofs are essentially what is given by

[16] In fact, one of the claims of this book is that there is an irremediable tension between these components.

valid arguments. On the contrary, mathematical proofs, as the intuitionists conceive them, depend essentially on the information available to the subject, and on the structure of her/his mind; and neither information nor her/his own mental structure are required to be consciously accessible to the subject or verbally expressible.[17] The same can be said about justifications for empirical statements, and consequently about their assertibility conditions. Therefore, Casalegno's argument offers a strong reason against the use-theory-of-meaning component of neo-verificationism, but not against an anti-realism about external world coupled with a realism about the mental, as I have argued that intuitionism is.

3.2.2.2 Prawitz on Content

The second objection to the idea of basing a theory of meaning on a defeasible key notion comes from the neo-verificationist side. Here is how Prawitz states it:

> It is not possible to explain the content of such sentences [sentences for which the notion of conclusive verification does not apply (*g.u.*)] in terms of possible justifications. The situation can easily be illustrated by considering ordinary universal sentences. To give an account of when the assertion of such a sentence is justified we may have to deal with all the problems of induction. However, even if we solved all problems of that kind and gave a completely accurate account of all the tests that have to be performed to be justified in asserting a universal sentence, it would be of no help for saying what the content of a universal sentence is. As long as the justifications in question are not conclusive, there is always the possibility that the sentence is refuted by a counterinstance although the justification of the assertion at an earlier time was completely in order. The assertion has then to be withdrawn. But there would be no reason to do so, if the content of the sentence were explained in terms of the possibility for a suitably placed person to justify the assertion of the sentence; *the existence of such a justification would be all that had been asserted, and the speaker would be right about that.* This clearly shows that the proposed understanding of such sentences in terms of possible justifications does not square with our linguistic practice with respect to these sentences. (Prawitz, 2002: 91. *Emphasis added*).

The crucial passage is the italicized remark, which shows also the presupposition behind the whole argument, namely that the content of the assertion of α is that there is a verification of α. I shall argue against this presupposition in Sect. 3 of Chap. 6; the content of the assertion of α—I will claim—is simply the proposition that α. If we give up Prawitz's assumption, his argument is blocked: the content of the assertion that all ravens are black is not that we actually have, nor that there exists in an atemporal sense, a justification for the statement "All ravens are black", but simply (the proposition) that all ravens are black, and the practice of withdrawing that assertion, when a white raven is found, can be accounted for in a natural way: since having a justification for α is a defeasible way of having warrant to assert α, it may happen that at t one has a justification for α and at a later time t' one has no

[17] Think of Brouwer's negative attitude towards mathematical language mentioned in Chap. 2. An argument for the propositional nature of evidence is given in Chap. 9 of Williamson (2000); but not necessarily of all kinds of evidence.

justification for α, and therefore that the assertion made at t has to be withdrawn at t'.

Someone might object that the notion of justification I am proposing is, again, an indefeasible notion: according to the picture I have proposed, after all, a cognitive state that is a justification for α does not lose its status of justification for α, but is simply replaced by another cognitive state which is not a justification for α. This remark is correct, but it misses the point. What is defeasible or not is not, properly speaking, the notion of justification or verification itself, but the notion of *having* a justification/verification; once justifications are conceived as cognitive states, *having* a justification for α amounts to *being in*, or *occupying*, a cognitive state which is a justification for α; and being in a specific cognitive state is defeasible.[18] As for factiveness, the answer depends on how one answers the question: when does one have warrant to assert α? I consider this issue in the next section.

3.2.2.3 On the Nature of Cognitive States

The way out of Casalegno's objection I have proposed has induced a transformation of our usual, intuitive way of conceiving justifications. Intuitive justifications are such 'things' as facts, arguments, perceptual experiences, proofs, pieces of information, memories, and so on; I have proposed to conceive theoretical justifications not as 'things' but as cognitive *states*. I should like to conclude this section by showing that, independently of the problem pointed out by Casalegno, there are also intuitive reasons to conceive justifications as cognitive states.

Let us start from facts. The presence of puddles in the streets was suggested by Casalegno as an example of justification John may have for the sentence (7). However, on second thought John's justification cannot be the objective fact that there are puddles in the streets; suppose that *there are* puddles in the streets and that John does not see them: of course he will not have a justification for (7). We might therefore suggest that justifications are not to be sought for among the facts of the external world, but among such inner states of the subject as perceptual experiences or beliefs whose content can be characterized in purely internalist terms.

---

[18] According to Williamson,

> even grasping a proof of a mathematical proposition is a defeasible way of having warrant to assert it. One can have warrant to assert a mathematical proposition by grasping a proof of it, and then cease to have warrant to assert it merely in virtue of gaining new evidence about expert mathematicians' utterances, without forgetting anything. (Williamson 2000: 265).

This is plausible if "proof" is understood as meaning a written proof or something similar; if a proof is conceived as a whole cognitive state, and having that proof is equated with being in that state, Williamson's view becomes much less plausible. If the subject $\mathcal{S}$ of Williamson's example has been in a mental state σ_1 that is a proof of α and later is in a mental state σ_2 that is not a proof of α, then either $\mathcal{S}$ has forgotten σ_1 when he is in σ_2, or he has chosen to trust the mathematicians instead of his own memory; the first case can be neglected: subjects are necessarily idealized, and idealized subjects have no limitation of memory, attention, etc.; in the second case $\mathcal{S}$ has made a mistake, hence he is *not* justified in believing that he has not warrant to assert α.

The difficulty with perceptual experiences, as with *qualia*, is that they seem to be essentially private, and therefore uncapable to provide a foundation for an intersubjective notion of meaning. With "intersubjective" I do not mean public, but governed by epistemic principles. For instance, when we say that John has a justification for (7) we are using "justification" to mean that it is epistemically *correct* for John to adopt the belief that it has been raining, and this not because John has peculiarities that distinguish him from every other human being; on the contrary, we would say that for every other human being, if (s)he were *in John's position*, it would be correct to adopt that belief. The problem is therefore to identify John's *mental* position in such a way as to obtain the intersubjectivity of justifications as a consequence.

A much better candidate than perceptual experiences in this respect is perceptual *representations*. Throughout this book I will use the term "representation" in the technical sense of a mental entity codifying structured information. Examples of mental representations might be PHON and SEM, the phonetic and the semantic representation of a sentence according to Chomsky's linguistic theory; or the primal sketch, the 2 1/2-D sketch and the 3-D model representation, the stages into which the computational process of vision is articulated according to Marr; or, in our case, the visual representation of puddles in the streets. On the one hand it is important to remember what has been said in Chap. 1: « "representation" is not to be understood relationally, as "representation of"» (Chomsky, 2000: 159): «there is nothing "represented" in the sense of representative theories of ideas, for example.» (Chomsky, 2000: 173).[19] On the other hand, representations are intersubjective in the sense explained above: since it is legitimate to assume that knowing subjects, as members of the same species, implement essentially the same computational apparatus, it is legitimate to assume that for *every* subject, when (s)he has access to the same representations—i.e., to the same information -, it is correct to adopt the same beliefs.

However, even mental representations are not sufficient. Suppose John has seen puddles in the streets, but did not know that normally rain leaves this sort of tracks: again he would not have a justification for the sentence (7). This suggests that only in presence of a certain amount of *background knowledge* is the visual representation of puddles in the streets a justification for that sentence. It is therefore more appropriate to say that the real justification for that proposition is the whole *cognitive state* in which the subject has the visual representation of puddles in the streets together with the necessary background knowledge. I shall introduce the technical notion of cognitive state in the next chapter; here it is sufficient to say that a cognitive state

[19] Therefore, when I spoke of «visual representation *of* puddles in the streets», this must be understood not in the sense that there is a certain relation between puddles in the streets and John's visual representation, but in the sense that John's visual representation has a *property*: the property of producing a certain visual experience. To quote Chomsky (2000), 160 once more:

> The representations are postulated mental entities, to be understood in the manner of a mental image of a rotating cube, whether it is the consequence of tachistoscopic presentations of a real rotating cube, or stimulation of the retina in some other way; or imagined, for that matter.

is intended to contain a certain amount of information, not necessarily verbal nor conscious, articulated into background information and actual information. In the case considered before, a visual representation of a puddle in the street is an example of actual non-verbal information; an example of background non-verbal information might be the memory of rain forming a puddle.

It is important to observe that, according to this way of conceiving cognitive states, and therefore justifications, a justification for a sentence cannot be identified with a list of (known) propositions, and this for at least two reasons. First, both background and actual information may be non-verbal: in that case there simply is not a definite list of propositions corresponding to it. Consider for instance the representation of a face actually perceived: which list of propositions codifies the information contained in it? Does the proposition that John is brown-haired belong to it? Or that John is clever? Or that some given person is John? The natural answer would be that it depends on the sentence for which we are characterizing a justification; but that would not do: background knowledge must be specifiable *before* such a sentence is formulated, otherwise no informative answer to the question "What is a justification for α?" would be at hand. In our example, once we have said that the sentence for which we are characterizing a justification is "John is brown-haired", the only piece of verbal information that might be suggested as a justification is that John is brown-haired; but such a suggestion would be of no explanatory value. Second, even if such a list were well defined, it would be not only potentially infinite, but also indefinitely extensible, since the activity of inferring proposition from a basis is most likely to be conceived as an open-ended process. The relation between cognitive states and lists of known propositions is under this respect analogue to the one between possible worlds and state-descriptions: as there are much more possible worlds than state-descriptions, owing to the lack of names for innumerable objects, so there are much more cognitive states than lists of known propositions, owing to the two reasons just described. As a consequence, when we characterize a cognitive state through a list of linguistic conditions, we do not in fact specify a unique cognitive state, but a class of states.

When justifications are conceived as cognitive states, facts and perceptual experiences are no longer good examples of justifications. What about proofs? Proofs are ordinarily considered as the typical justifications for mathematical sentences. However, in ordinary usage the word "proof" is typically ambiguous: it may refer to an argument presented by a teacher to her class, to a sequence of written lines on a textbook, or to some sort of complex mental process or act through which we come to see that a certain sentence is true. Of course, an argument or a sequence of written lines in a textbook cannot confer evidence on any proposition, and therefore be a justification for it, unless it is understood; and the result of understanding a written proof may be plausibly conceived as a cognitive state. Mental processes or acts may seem capable to confer evidence on a sentence; however, on the one hand, it is not the execution of an act in itself that can confer evidence, but the execution of an act according to certain programs; on the other hand, every act/process ruled by a program can be characterized as the application of an input mental state into an output mental state; hence proof as acts/processes are not something essentially new with respect to proofs as cognitive states, once the internal structure of the computing mind is taken into consideration.[20]

3.2.3 K-Factiveness and Truth-Grounds

We have seen in Chap. 2 that the notion of proof plays a double role within the intuitionistic theory of meaning: on the one hand it explains understanding, hence meaning, of mathematical sentences (as capacity to recognize proofs); on the one hand it explains their assertion conditions (as possession of a proof). If we now try to extend this theory to empirical sentences by replacing the notion of proof with a defeasible notion of *ex ante* justification (as I have argued to be necessary in Sect. 3.1.2), we can still explain understanding, hence meaning, of empirical sentences (as capacity to recognize their justifications); the question is how can we explain their assertion conditions.

One strategy is to require that empirical justifications are factive. However, in many empirical cases factiveness and epistemic transparency of justifications are incompatible. Let us come back to the paradigmatic example given above of a justification for the empirical atomic sentence “it is raining”; it is described by Prawitz, in a paper in which he expounds his ToG, in the following terms:

> a ground for a proposition ‘it is raining’ is taken to consist in seeing that it rains; taking “seeing” in a veridical sense, it constitutes a conclusive ground. (Prawitz, 2009: 186)

If we take “seeing” in a veridical sense, then seeing that α is a factive way of having warrant to assert that α; now, if a ground for “it is raining” consists in seeing that it rains, and seeing is a factive way of having warrant to assert “it is raining”, then it is illegitimate to assume that having such a ground is epistemically transparent since, as we have noticed above, the evidence actually available to a subject is not sufficient for him to discriminate between seeing factively that it rains and seeing non-factively that it rains, namely merely having the impression to see that it rains: the experience of the subject is the same, whether he is seeing or merely having the impression to see. It may therefore happen that a subject is in possession of a ground of “it is raining” without being in a position to know that he is: his visual experience being a case of veridical seeing or of non-veridical seeing *exclusively* depends on how external reality actually is, on the existence of a mere fact, the existence of this fact may transcend the cognitive capacities of the subject. We cannot therefore require that the notion of empirical justification is factive. Moreover, we have noticed above that factiveness is a realist notion, and that if one adopts an anti-realist approach to the theory of meaning, one had better to avoid it until a coherent anti-realist notion of truth has been defined.

A second possibility is based on the remark that, in order to grant assertibility, it is not strictly necessary that a justification is factive: what is really necessary is that

[20] What I mean by “program”, “input state”, “internal structure of the computing mind” etc. will become clear in the next chapter.

it warrants *knowledge*, since it is knowledge that warrants assertibility. I shall call *k-factiveness* this property, and *truth-grounds* justifications of such kind:

(8) A justification is *k-factive* just in case necessarily only known statements have justifications of that kind.
(9) A *truth-ground* of α is a k-factive justification for α.

That k-factiveness grants assertibility is equally true in mathematical and emprical domains; the difference between the two kinds of domains, as we saw above, is that mathematical justifications are indefeasible, empirical ones are defeasible; as a consequence, mathematical knowledge (hence assertibility) is stable through time, empirical knowledge (hence assertibility) is not. This is the strategy I shall explore.

An obvious objection is that if having a truth-ground of α is k-factive, then it is also factive, since knowledge is factive; as a consequence, the second strategy is open to the same objection as the first. However, the introduction of the notion of k-factiveness, hence of knowledge, has changed the situation: the incompatibility between factiveness and epistemic transparency depended on the fact that factiveness is defined in terms of the bivalent notion of truth; k-factiveness is defined in terms of the notion of knowledge; hence, if knowledge can be defined without making reference to the bivalent notion of truth, the second strategy can be implemented, provided that truth-grounds can be defined in such a way as to be epistemically transparent.

I shall propose a definition of knowledge which does not make reference to realistic truth in Chap. 8; and I shall show that the theoretical notion of truth-ground of α is epistemically transparent in Chap. 4. Here I shall argue that the strategy I am proposing is not an *ad hoc* way out of the predicament generated by the incompatibility between factiveness and transparency, because there are others reasons, independent of that predicament, for defining knowledge without making reference to realistic truth.

The Platonic definition of knowledge is open to the following objection, different from, and independent of, Gettier problems, which I shall examine in Chap. 8. According to the definition, $\mathcal{S}$ knows that α iff $\mathcal{S}$ believes that α, $\mathcal{S}$ is justified in believing that α, and α is true. Since the condition that α is true occurs in the *definiens*, the concept of truth is presupposed as primitive or as definable independently of knowledge. This means that, in order to establish (ascertain) whether $\mathcal{S}$ knows that α, an observer $\mathcal{S}$' must have established in advance whether α is true, or at least be in a position to do it independently of establishing whether $\mathcal{S}$ knows that α. This induces a serious difficulty for the explanation of scientific practice, which is characterized by the fact that, in most cases, the observer $\mathcal{S}$‘ is the same as the observed scientist $\mathcal{S}$ since there simply is not an observer having a way to access the truth of α different from the one $\mathcal{S}$ has built or is building. In order to explain this practice it is necessary to admit the possibility that $\mathcal{S}$ knows that α (and knows that (s)he knows that α) *without having established in advance whether* α *is true*.[21]

[21] We might say, equivalently, that, in order to explain scientific practice, we need a definition of knowledge providing a *criterion* of knowledge to the subject who is seeking knowledge.

This objection to the Platonic definition of knowledge is analogous to the objection to Tarski's definition of logical consequence quoted at the end of Sect. 3.2.1.[22] Let me try to make the argument explicit:

(10) For every subject $\mathcal{S}$, a valid inference $\mathcal{I}$ from α to β is intuitively useful for $\mathcal{S}$ if, and only if, for every time t, $\mathcal{S}$ knows at t that β only if there is a time t' such that (i) t' < t; (ii) $\mathcal{S}$ knows at t' that $\mathcal{I}$ is valid; (iii) $\mathcal{S}$ knows at t' that α; (iii) $\mathcal{S}$ does not know at t' that β.

(11) A good explanation of the validity of an inference $\mathcal{I}$ must account for its utility, i.e. explain how $\mathcal{I}$ can be at the same time valid and useful.

(12) Assume that we define an inference as valid if, and only if, it preserves (realist) truth.

(13) From (12) and (10) it follows that, for every subject $\mathcal{S}$, a truth-preserving inference $\mathcal{I}$ is useful for $\mathcal{S}$ if, and only if, for every time t, $\mathcal{S}$ knows at t that β only if there is a time t' such that (i) t' < t; (ii) $\mathcal{S}$ knows at t' that $\mathcal{I}$ is truth-preserving; (iii) $\mathcal{S}$ knows at t' that α; (iv) $\mathcal{S}$ does not know at t' that β.

(14) Condition (13)(ii) means, by Tarski's definition of truth-preserving inference, that, for every model $\mathcal{M}$, either $\mathcal{S}$ knows at t' that α is false in $\mathcal{M}$ or $\mathcal{S}$ knows at t' that β is true in $\mathcal{M}$.

(15) Then there cannot be a time t' satisfying the conditions (i)-(iv) specified in (13); for, if (ii) holds, then, by (14), for every model $\mathcal{M}$, either $\mathcal{S}$ knows at t' that α is false in $\mathcal{M}$ or $\mathcal{S}$ knows at t' that β is true in $\mathcal{M}$; by (iii), $\mathcal{S}$ does not know at t' that α is false in $\mathcal{M}$; hence $\mathcal{S}$ knows at t' that β is true in $\mathcal{M}$, in contradiction with (iv).

(16) Hence, if we equate the validity of an inference with its being truth-preserving, a valid inference cannot be useful; by (11), the definition of validity as truth preservation does not account for its utility.[23]

The analogy between Etchemendy's (and Prawitz's) objection to Tarski's definition of logical consequence, on the one hand, and my objection to the Platonic definition of knowledge, on the other, can be spelled out as follows.

- In both cases we have to do with inference: deductive inference in the former case, abductive inference in the latter.
- In the former case β is inferred because one knows that α and one knows that $\mathcal{I}$ (the deductive inference of β from α) is valid; in the latter β is inferred because it is the best answer to the question "Why α?"
- In the former case the question arising is: "How should the validity of a deductive inference be defined if one wants to explain at the same time its utility?"; in the latter case the question is: "How should the validity of abductive inference be defined if one wants to explain at the same time its utility?"

---

[22] As we have seen in Chap. 1, Sect. 1.2.2., essentially the same objection is stated in Prawitz (2005: 675).

[23] If this reconstruction is correct, a supporter of Tarski's definition of logical consequence would presumably question (14); but (s)he would still owe us an explanation of what knowing that an inference is truth-preserving amounts to.

- Prawitz's answer to the former question (Prawitz, 2005: 681) consists in reversing conceptual priorities: instead of defining the validity of deductive inference in terms of logical consequence (hence of truth), as Tarski made, we should define logical consequence (hence truth) in terms of the validity of inference, and define this notion in terms of the notion of evidence.[24] In the same vein, the answer I suggest to the latter question (hence to the objection I have levelled at the Platonic definition) is based on an analogous reversal of conceptual priorities: instead of defining the validity of abductive inference (hence knowledge of its conclusion) in terms of truth (as in the Platonic definition of knowledge), we should define truth in terms of validity of abductive inference (hence of knowledge), and define this notion without any appeal to the notion of truth.

The *practice* of inferring the truth of α from the fact that α is known is the norm in the case of mathematical knowledge: we infer that Fermat's last theorem is true from the fact that Andrew Wiles has proved it, and proving a mathematical statement is the standard way of knowing it. The (realistic) truth of Fermat's last theorem plays therefore no role in the acquisition of our knowledge of it; if it did, it would be impossible to know Fermat's last theorem without knowing in advance that it is true. However, in the *theory* of deductive inference Prawitz's answer to the question of utility is a minority view, and is perceived as a reversal of established conceptual priorities.

In the case of empirical knowledge the situation is more complex, due to two facts: (i) while mathematical knowledge is essentially based on scientific methods, empirical knowledge may be based either on scientific methods or on 'intuitive' procedures; (ii) in empirical domains there are often several independent ways of acquiring knowledge about a proposition α. An important consequence of (ii) is that in many cases we are in a position to judge a proposition α (to recognize its truth or its falsity) independently of having evaluated a justification procedure j for α; in such cases the Platonic definition of knowledge may be a useful tool to evaluate just the reliability of j; and its utility in these cases may hide its uselessness to acquire essentially new knowledge. When essentially new knowledge is at stake, the *practice* is essentially similar to the one adopted in mathematical domains. An important way of acquiring empirical knowledge is inference to the best explanation: a scientist $\mathcal{S}$ observes certain phenomena, produces a set of potential explanations of them, chooses one of these explanations, say α, as the best one on the basis of certain criteria, and infers α. As soon as $\mathcal{S}$ chooses α as the best explanation, it is legitimate to say that $\mathcal{S}$ knows that α. Of course there are additional complexities; for instance, another scientist $\mathcal{S}$ ' might infer β from the same phenomena, and as far as α and β are

[24] The idea of this reversal was already present in a paper of 1980:

> [T]he classical theory of meaning and truth is certainly unable to explicate the conditions for correct assertion. It is rather the other way around: the notion of truth gets its fundamental characteristics from its relation to the conditions for correct assertion and is explicated in terms of these conditions. (Prawitz, 1980: 8).

incompatible we would not say either that $\mathcal{S}$ knows that α nor that $\mathcal{S}$ ' knows that β; but even in such cases our leading criterion to establish which of them really knows is which of the explanations they propose is really the best. The crucial remark is that the reason why an explanation E is chosen by a subject $\mathcal{S}$ as better than another explanation E' is a cluster of properties $\mathcal{S}$ recognizes E as having, otherwise $\mathcal{S}$ could not choose it; since the truth of α is by hypothesis unknown (and in some cases unknowable), it cannot be one of the properties of the cluster; as a consequence, "s knows that α" characterizes a state of $\mathcal{S}$ that entails the truth of α, but of which the (realistic) truth of α is not a constitutive characteristic, in the sense that it is not part of its definition. Like in mathematics, we infer that α is true from the fact that we know that α, hence the truth of α plays no role in the definition of knowing that α; if it did, it would be impossible to know that α without knowing in advance that it is true, and the explanation of empirical inference as a practice of inference to the best explanation would become an intractable problem.

Differently from what happens in the case of mathematics, in the *theory* of empirical inference the anti-realistic conceptual priorities are not a minority view. A significant example is just the theory of inference to the best explanation, according to which our inductive inferences are best explained by saying that, given certain data and certain background beliefs, we infer what provides the best of the possible explanations we can generate of those data. Since of course no explanation is better than the true one, it would seem natural to define an inference of this kind as valid when it enables us to infer, from true data, the true (or actual) explanation of them; however, such a characterization is not adopted:

> it would not characterize the process of inference in a way we could follow, since we can only tell whether something is an actual explanation after we have settled the inferential question. It does not give us what we want, which is an account of the way explanatory considerations can serve as a guide to the truth. (Lipton, 2004: 58)

The characterization usually adopted reverses, again, conceptual priorities: a distinction is introduced between actual (true) and potential explanations, and an inference is said to be valid when it enables us to pass from the data to their best potential explanation, where the notion of potential explanation and the relation "explanation E is better than explanation E'" are defined without any reference to the truth of the explanations.

The constraint of k-factiveness requires a solution to Gettier problems; for, on the one hand, k-factiveness is a property of a cognitive state that (together with others) is sufficient for knowledge; on the other hand, Gettier problems just show the insufficiency of justified true belief for knowledge. The definition of the notion of $\mathcal{C}$-truth-ground is just required to fill the gap between justified true belief and knowledge, hence yielding a solution to Gettier problems. As I said, an analysis of these problems will be proposed in Chap. 8.

3.3 Conclusion. Internalist Anti-Realism Revisited

It is now possible to give a more explicit formulation of the sort of internalism about justification I find akin to Chomsky's internalism, and of the sort of anti-realism I am advocating.

The rationale for the most common internalist conceptions of justification is reconstructed in the following way by Goldman:

(1) The *guidance-deontological* (GD) *conception of justification* is posited.

(2) A certain constraint on the determiners of justification is derived from the GD conception, that is, the constraint that all justification determiners must be *accessible to*, or *knowable by*, the epistemic agent.

(3) The accessibility or knowability constraint is taken to imply that only internal conditions qualify as legitimate determiners of justification. So justification must be a purely internal affair. (Goldman, 1999: 272)

According to the GD conception, the fundamental role of (epistemic) justification is to guide us about what to believe. This conception is often paired with the deontological conception, according to which «being justified in believing a proposition *p* consists in being (intellectually) required or permitted to believe *p*; and being unjustified in believing *p* consists in not being permitted, or being forbidden, to believe *p*.» (Goldman, 1999: 273) The reason of the strict association of the two conceptions is that, in these varieties of internalism, the knowability constraint is in fact derived from the deontological conception. The strategy of the derivation is to invoke the general principle that, if one has a duty, it must be possible for him to perform it; and the implementation of this strategy in this specific case consists in characterizing the epistemic duty as the duty to «believe what is supported or justified by one's evidence and to avoid believing what is not supported by one's evidence» (Feldman, 1988: 254); applying the general principle to this case one infers that the evidence/justification for a proposition must be accessible: if a subject has the epistemic duty to believe an evident proposition, (s)he must know what her/his duty is, hence (s)he must be in position to recognize the evidence of that proposition. Since evidence has an obvious guiding role, the guidance conception is strictly associated to the deontological conception in such a derivation.

This is *not* the way I propose to motivate the accessibility/ knowability constraint: $\mathcal{C}$-justifications guide our conceptual-intentional systems, but the notion of $\mathcal{C}$-justification is not a deontological (nor an evaluative) notion. The motivation I suggest has been given in the preceding chapter: the accessibility/knowability constraint is a version of the epistemic transparency requirement imposed onto $\mathcal{C}$-justifications, and this is motivated by the fundamental idea that knowing the meaning of a sentence α consists in being in a position to recognize justifications for α: a competent speaker of the language to which α belongs must know the meaning of α, hence must be capable to recognize the justifications for α.

As I explained in Sect. 3.2.2, the accessibility/knowability constraint is not to be understood as a requirement of accessibility to conscious states. Consequently, the internalism about justifications advocated here is of the sort Goldman (1999) would call "weak"; in particular, it involves no appeal to introspection as a privileged form of access to one's own internal states. Indeed, the very notion of *internal* state does not play any role in the present approach—and this constitutes a crucial difference from most varieties of internalism. The central role is played by cognitive states, and cognitive states are conceived as states of a computational device; they are therefore 'internal' only in the very weak and vague sense that they are states of a computational device involving representations, and that some part of mind is identified with such a computational device. Much more significant is the role attributed to internal states in most varieties of internalism. Recall for example Williamson's characterization of the internalism/externalism opposition described in fn. 20 of Chap. 1:

> A case α is *internally like* a case β if and only if the total internal physical state of the agent in α is exactly the same as the total internal physical state of the agent in β. A condition C is *narrow* if and only if for all cases α and β, if α is internally like β then C obtains in α if and only if C obtains in β. [...] C is *broad* if and only if it is not narrow. A state S is narrow if and only if the condition that one is in S is narrow; otherwise S is broad. Internalism is the claim that all purely mental states are narrow; externalism is the denial of internalism. (Williamson, 2000: 52)

The definition of the relation of internal likeness between states depends therefore on the definition of internal physical state. The problem with this characterization is that neither the notion of 'internal' nor the notion of 'physical' seem to be clear enough. As Williamson explains (Williamson, 2000: 51), an *internal* state is a state that occurs within the spatio-temporal boundaries of the agent's body at the time of action—in fact of the agent's brain at the time of action. But suppose an agent has a bullet in his brain, and that it affects his actual action; shall we say that the physical state consisting in the brain-and-the bullett's state is an internal state? According to the proposed criterion yes, but it is unlikely that many internalists would accept this answer. On the other hand, we have seen in Chap. 1, Sect. 1.1, the reasons why Chomsky holds that neither a monist (materialist) nor a dualist thesis can be formulated; for exactly the same reasons the distinction between mental and physical states is not well-grounded from the scientific point of view. Therefore an internalism about justification intended to be congenial to Chomsky's internalism cannot be characterized in this way; a re-definition is needed. It will be implicitly given in the next chapter and in Chap. 5, through the definition of the notions of atomic cognitive state and cognitive state *tout court*.

To conclude, the version of anti-realism I find most congenial to Chomskyan internalism is inspired to some basic ideas of the Proof Explanation, and tries to extend it in two senses: (i) by introducing as the key notion of the theory of meaning a suitable generalization of the notion of proof, capable of playing the role of evidence for empirical sentences, as they are used in everyday language; I will call this notion "$\mathcal{C}$-*justification for* a sentence" ("$\mathcal{C}$" for "cognitive", but also for "computational"); (ii) by integrating the Proof Explanation with a definition of the notion of $\mathcal{C}$-justification for *atomic* sentences, in order to obtain an overall theory of meaning for sentences as used in everyday language.

In particular, the following claims will be inspired by intuitionistic ideas, and at the same time seem to be congenial to Chomskyan internalism:

(17) A language is a matter of an individual, not of a community.
(18) Meaning concerns the relations between language and mind, not the relations between language and external reality.
(19) Knowing the meaning of a sentence α amounts to being in a complex mental state, in which one is in a position to recognize a justification for α.
(20) Justifications for sentences should be defined by induction on the logical complexity of those sentences, and without any reference to an external reality.
(21) The truth of a sentence α consists in the existence of a justification for α. As a consequence truth too concerns the relations between language and mind, not the relations between language and external reality.

These claims constitute the general framework of the theory of meaning I am going to develop in this book; only the framework: in order to arrive at a plausible theory it will be necessary not only to add other claims, but also to refine the present ones.

References

Boghossian, P. (1994). The transparency of mental content. *Philosophical Perspectives, 8*, 33–50.

Boghossian, P. (2011). The transparency of mental content revisited. *Philosophical Studies, 155*(3), 457–465.

Casalegno, P. (2002). The problem of non-conclusiveness. *Topoi, 21*, 75–86.

Chomsky, N. (2000). *New Horizons in the study of language and mind*. Cambridge University Press.

Dummett, M. (1991a). *Frege. Philosophy of mathematics*. Duckworth.

Dummett, M. (1991b). *The logical basis of metaphysics*. Duckworth.

Feldman, R. (1988). Epistemic Obligations. *Philosophical Perspectives* (vol. 2, 235–256), Epistemology.

Firth, R. (1978). Are epistemic concepts reducible to ethical concepts? In A. Goldman & J. Kim (Eds.), *Values and morals* (pp. 215–229). Reidel.

Goldman, A. I. (1979). What is justified belief? In G. S. Pappas (Ed.), *Justification and knowledge* (pp. 1–23). Reidel.

Goldman, A. I. (1988). Strong and weak justification. *Philosophical Perspectives* (vol. 2, 51–69), Epistemology.

Goldman, A. I. (1999). Internalism exposed. *The Journal of Philosophy, 96*(6), 271–293.

Kvanvig, J. & Menzel, Ch. (1990). The basic notion of justification. *Philosophical Studies, 59*(3), 235–261.

Lipton, P. (2004). *Inference to the best explanation.* Second Edition. Routledge. (First Edition 1991).

Peacocke, C. (1992). Sense and justification. *Mind, 101*, 793–816.

Prawitz, D. (1980). Intuitionistic logic: A philosophical challenge. In G. H. von Wright (Ed.), *Logic and philosophy* (pp. 1–10). Nijhoff.

Prawitz, D. (1987). Dummett on a theory of meaning and its impact on logic. In B. Taylor (Ed.), *Michael Dummett: Contributions to philosophy* (pp. 117–165). Nijhoff.

Prawitz, D. (2002). Problems for a generalization of a verificationist theory of meaning. *Topoi, 21*, 87–92.

Prawitz, D. (2005). Logical consequence from a constructivist point of view. In S. Shapiro (Ed.), *The Oxford handbook of philosophy of mathematics and logic* (pp. 671–695). Oxford University Press.

Prawitz, D. (2009). Inference and Knowledge. In M. Peliš (Ed.), *The Logica Yearbook 2008* (pp. 175–192). College Publications, King's College London.

Williamson, T. (1988). Knowability and constructivism. *Philosophical Quarterly, 38*, 422–432.

Williamson, T. (1998). Review of G. Usberti, *Significato e conoscenza. Dialectica, 52*(1), 63–69.

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

Chapter 4
$\mathcal{C}$-Justifications for Atomic Sentences. Names and Predicates, $\mathcal{C}$-Objects and $\mathcal{C}$-Concepts

Abstract The aim of this chapter is to define the theoretical notion of $\mathcal{C}$-justification ("$\mathcal{C}$" for "cognitive", but also for "computational") for atomic sentences. As a preliminary the notion of prelinguistic cognitive state is characterized (Sect. 4.1) in terms of which the basic notions of an internalist ontology—$\mathcal{C}$-object and $\mathcal{C}$-concept—are defined. Linguistic atomic cognitive states are introduced in Sect. 4.2, and with them the notions of $\mathcal{C}$-authorization to use a singular term to refer to a given object and $\mathcal{C}$-authorization to use a predicate to apply a manageable concept to $\mathcal{C}$-objects. Predication is conceived as a case of inference to the best explanation, and this, in turn, as a particular kind of computation (Sect. 4.3). The notions of $\mathcal{C}$-justification for, and of $\mathcal{C}$-truth-ground of, an atomic sentence are defined in Sect. 4.4, where it is argued for their epistemic transparency. In Sect. 4.5 it is shown how the 'uninteresting' problems generated, according to Chomsky, by the realist assumption of externalist semantics vanish within the framework of the present internalist semantics, and how, on the other hand, a direction can be suggested for the solution of some 'interesting' problems. The next two sections are devoted to a reformulation of Frege's distinction between sense and denotation: in Sect. 4.6 a theoretical *explicans* of the intuitive notion of synonymy is individuated through an analysis of our intuitions about synonymy of expressions of various categories; in Sect. 4.7 the present approach is compared with the neo-verificationist model of sense, based on a distinction between direct and indirect methods of object identification. The chapter terminates (Sect. 4.8) with a discussion of the so-called frame problem as it has been formulated by Fodor.

Keywords Theory of meaning · Justificationist semantics · Anti-realism · Internalist semantics · Justification · Names · Predicates · Concepts · Cognitive states · Explanation · Truth-ground · Chomsky · Neo-verificationism · Frame problem

G. Usberti, *Meaning and Justification. An Internalist Theory of Meaning*, Logic, Epistemology, and the Unity of Science 59,
https://doi.org/10.1007/978-3-031-24605-0_4

The aim of this and of the next chapters is to define the theoretical notion of justification—which I call $\mathcal{C}$-justification ("$\mathcal{C}$" for "cognitive", but also for "computational")—for the sentences of the language $\mathcal{L}$; the present chapter is devoted to atomic sentences, the next to logically complex ones.

The strategy I adopt is compositional: the semantic value of $\pi(\nu_1, \ldots, \nu_k)$ in the cognitive state σ will be defined in such a way that it depends on the semantic values of π, ν_1, ..., ν_k in σ and on σ.[1] As a preliminary to their definition, the notion of prelinguistic cognitive state is characterized (Sect. 4.1) in terms of which the basic notions of an internalist ontology—$\mathcal{C}$-object and $\mathcal{C}$-concept—are defined. Linguistic atomic cognitive states are then introduced (Sect. 4.2), and with them the notions of $\mathcal{C}$-authorization to use a singular term to refer to a given object and $\mathcal{C}$-authorization to use a predicate to apply a manageable concept to $\mathcal{C}$-objects. Predication is conceived as a case of inference to the best explanation, and this, in turn, as a particular kind of computation (Sect. 4.3). The notions of $\mathcal{C}$-justification for, and of $\mathcal{C}$-truth-ground of, an atomic sentence are defined in Sect. 4.4, where it is argued for their epistemic transparency. In Sect. 4.5 it is shown how the 'uninteresting' problems generated, according to Chomsky, by the realist assumption of externalist semantics vanish within the framework of the present internalist semantics, and how, on the other hand, a direction can be suggested for the solution of some 'interesting' problems. The next two sections are devoted to a reformulation of Frege's distinction between sense and denotation: in Sect. 4.6 a theoretical *explicans* of the intuitive notion of synonymy is individuated through an analysis of our intuitions about synonymy of expressions of various categories; in Sect. 4.7 the present approach is compared with the neo-verificationist model of sense, based on a distinction between direct and indirect methods of object identification. The chapter terminates (Sect. 4.8) with a discussion of the so-called frame problem as it has been formulated by Fodor.

I will start with an informal discussion of the problems and an equally informal presentation of the notions and ideas introduced to solve them; formal definitions will follow in Sect. 4.4.1.

4.1 Prelinguistic Cognitive States

Cognitive states are states of the mind of a subject, once mind is conceived as a computing apparatus. Since $\mathcal{C}$-justifications are justifications for sentences, I will be especially concerned with 'linguistic' cognitive states, i.e. with states of a mind endowed with a language faculty. However, also a 'prelinguistic' mind has cognitive states; moreover, they have in my view a logical priority over linguistic ones, as we will see. It will be therefore necessary to introduce some general assumptions about the structure of 'prelinguistic' mind. I shall try to extract them from the consideration of two elementary competences that, plausibly, human mind shares with the minds

[1] The meaning of "on σ" will be explained in Sect. 4.3.3.

of many animals: perceiving, or more generally being given, objects; and grasping, or managing, concepts.

4.1.1 A Caveat

I want to stress preliminarily that my intention in this section is not to explain either the process of constitution of objects or the nature of concepts in a prelinguistic mind: this could not be the goal of a semantic theory; I shall only try to individuate a notion of object and a notion of concept that are 'compatible' with the kind of computational-representatational explanations cognitive psychology gives of object perception. In particular, if perception is to be explained as a computational process, it is necessary that perceived objects are computable and classifiable by means of computations; within the conceptual framework I will propose, this means that it is legitimate to assume that prelinguistic $\mathcal{C}$-concepts are manageable by a subject, and that the subject can establish by means of computations whether a term of the internal representational system (**IRS**) belongs to a $\mathcal{C}$-object, whether two $\mathcal{C}$-objects are identical, whether an object file matches a $\mathcal{C}$-object stored in memory. I hope I will make plausible these assumptions, which I condense in the claim that the cognitive states of a 'prelinguistic' mind are inhabited by $\mathcal{C}$-objects, and that several $\mathcal{C}$-concepts are manageable. This does not exclude that many new $\mathcal{C}$-objects can be created within linguistic cognitive states, nor that many new $\mathcal{C}$-concepts can become manageable, as we will see in Sect. 4.2.

4.1.2 A Given Object

What do we mean when we say that a specific object is *given* to a subject $\mathcal{S}$? And, even before this: what *is* an object? The questions may seem senseless, owing to the fact that there is an unpredictable variety of ways an object can be given—by vision, touch, hearing, by memory, attention, imagination, and so on. However, on the one hand we do qualify all these ways as ways of giving an object—in other terms, we do consider intuitively legitimate such a general notion of object; on the other hand, a semantic theory requires just such an extremely general notion of object, simply because it requires entities to interpret names onto. This need subsists independently of the externalist or internalist nature of the semantics we adopt; it is a conceptual necessity, due, among other reasons,[2] to the fact that we want—and must—explain

[2] Another reason will be given in Sect. 4.2.2.

the meaning of the identity sign: if two *different* signs as "a" and "b" could not denote *the same object*, "a = b" could not mean that *a* and *b* are one object.[3]

The difference (one of the differences) between externalist and internalist semantics lies in the answer to the starting questions. According to the former, objects are individuals of the external world,[4] and they are given to us when they cause some of our mental states.[5] We have seen in Chap. 1 that this choice entails the adoption of the ontology of common sense,[6] and that this prevents semantics from being a scientific enterprise. So, an internalist answer must meet an important constraint: it must be compatible with the ontology of some science; since the objects internalist semantics is concerned with are mental entities, the science at stake will be psychology, more specifically computational-representatational theories of mental domains.

Keeping these preliminary remarks in mind, let us come back to our starting questions; for definiteness, let us say that $\mathcal{S}$ sees a glass of water: what do we mean? We can mean two very different things: either that what is *in fact* a glass of water is seen by (given to) $\mathcal{S}$, independently of the subject's actually seeing that object as a glass of water; or that $\mathcal{S}$ actually sees something *as* a glass of water, independently of its being in fact a glass of water. I hold that the two meanings are clearly distinct, and that we should try to answer our questions in *both* their readings; however, I hold that the computational significance of the questions is immediately clear only when they are understood in the second sense, so that it will be convenient to start from this interpretation.[7] Let me explain why. There is some instant of time *t* such that, from the point of view of a subject, there is in principle no difference between being presented at *t* with a glass of water itself and being presented at *t* with a hologram of a glass of water, or even with being presented at *t* with nothing at all, and being only stimulated at *t* with an electrode at some point of the cerebral cortex in such a way as to produce the mental image of a glass of water; when I say that there is no difference from the standpoint of the subject, I mean that if our aim is to characterize the mental state of the subject and his mental computations, the subject's mental state at *t* is exactly the same in the two situations: the two situations are very different, but the amount and the quality of information available to the subject at *t* is the same, and therefore his mental computations will be the same; in other terms, what makes the difference between the two situations is entirely irrelevant. What is relevant is only information available to the subject at *t* and the structure of the subject's computing apparatus.[8]

---

[3] We have seen in Chap. 1 that, according to Chomsky, internalist semantics is syntax in the technical sense that «it deals with the properties and arrangements of the symbolic objects» (Chomsky 2000: 174); however, this does not mean that linguistic symbols denote themselves.

[4] «What is an individual? A very good question.» (Scott 1970: 144).

[5] Not all agree on this.

[6] The quotation from Scott (1970: 144) mentioned in fn. 4 continues as follows: «So good, in fact, that we should not even try to answer it. We could assume that being an individual is a primitive concept—that is harmless: any sufficiently clear concept can be made primitive.» But is it sufficiently clear?

[7] I will try to answer the questions in their first reading only in Chap. 7.

[8] Presumably these factors are also the only ones responsible for the subject's perceptual experience.

If we call $\mathcal{C}$*-object* ("$\mathcal{C}$" for computational) what is given to an information-processing system, our starting questions can be rephrased as follows: What does it mean that a $\mathcal{C}$-object is given to a subject? And what is a $\mathcal{C}$-object?

As I have argued right now, it is perfectly possible that there is absolutely *nothing*, in the surrounding environment, that is presented to the subject, even though (s)he perceives something, for instance a glass of water. Well: it is possible that there is nothing *in the environment*, but there is surely something *in the subject's mind*. What? Let us see how objects are conceived within the framework of one influential theory of visual perception of objects[9]:

> Imagine watching a strange man approaching down the street. As he reaches you and stops to greet you he suddenly becomes recognizable as a familiar friend whom you had not expected to meet in this context. Throughout the episode, there was no doubt that a single individual was present; he preserved his unity (in the sense that he remained the *same* individual), although neither his retinal size, his shape, nor his mental label remained constant. Perception appears to define objects more by spatiotemporal constraints than by their sensory properties or by their labeled identity. The perceptual system is also capable of restoring continuity that has been briefly broken in the stream of sensory inputs. The man who reappears after walking behind a car will normally be treated as the same individual who was seen to disappear, provided that the disappearance was short and that the parameters of motion remain more or less constant. (Kahneman, Treisman & Gibbs 1992: 176–177)

The authors distinguish therefore two senses of the term "identity"; according to the first, the approaching man assumed to be a stranger is not identical with the man recognized as a friend; according to the second, the two men are one and the same. In the first sense the identity of an object is the label conferred on it when it is identified; in the second the identity depends on spatiotemporal constraints.

In order to account for this distinction the notion of *object file* is introduced: an object file is addressed by its location at a particular time (not by any feature or identifying label), and contains sensory information that has been received about the object at that location. When the sensory input changes, the information in the files is updated; whenever possible the cognitive system assigns current information to preexisting object files, yielding the perceptual experience of changing or moving objects; when this is impossible a new object file must be set up. The system of object files is distinct from the network of nodes and connections that permits recognition. To mediate recognition, the sensory description in the object file is compared to stored representations of known objects. If and when a match is found, the identification of the object is entered in the file.

Within this framework, the ingredients necessary to explain what is in one's mind when one perceives a visual object seem to be the following. First, a mental representation, or description, which is *activated* at a particular time and encapsulates information—a description belonging, in the terminology of the authors, to an object file. "Representation" and "description" are not used here in a relational sense: it is not a representation *of* an object in the external world, in the sense that to define it

[9] I choose the following quotation for its strong intuitive appeal. For my purposes other computational approaches to vision might have been chosen, including Marr (1982)and Ullman (1996).

is not necessary to make reference to an object of the external world.[10] A representation, in the present context, is simply a structured symbol, a *term* of some internal representational system. I shall call "*at*" the activated term.

The second ingredient are *constitutive criteria* to establish whether two terms of an internal representational system are to be assigned to the same object file. These criteria can be seen as functions taking as arguments pairs of terms and giving as value 1 when the two terms are in some relevant sense equivalent, so that object files are the classes of the partition induced by the equivalence relation. The relevant sense is to be defined, for different contexts or cognitive states, by psychological theories of object perception. For example, when a stationary scene is perceived, the relevant equivalence may consist in having the same location at a particular time; when what is perceived is a moving object, the relevant equivalence may be determined by some of the principles stated in this passage from Spelke (1990):

> the processes by which humans apprehend objects occur relatively late in visual analysis, after the recovery of information for three-dimensional surface arrangements and motions. The processes appear to accord with four principles - cohesion, boundedness, rigidity, and no action at a distance - that reflect basic constraints on the motions of physical bodies. These principles may be central both to human perception of objects and to human reasoning about object motion [...].(Spelke, 1990: 30-31)

For example, according to the cohesion and the boundedness principles, two surface points lie on the same object if and only if there is a path of connected surface points linking them (Spelke, 1990: 49); the relevant equivalence may therefore consist in being linked by a path of connected surface points.

The third ingredient is a *catalogue* of stored representations of known objects, a «set of nodes in long-term memory that represent its parts, properties, and categories of membership» through a system of features and labels.

The fourth ingredient is a *matching operation* comparing object files with stored objects, thereby permitting recognition. I submit that this operation works in the following way. Object files and stored objects are classes of representations, i.e. of terms of **IRS**; each term either contains features (i.e. labels occurring in the term),[11]

---

[10] Of course it may happen, and often it does happen, that it *is* a representation of an object in the external world; but this is an entirely different question. This technical sense of "representation" is common in cognitive psychology, as we have seen in Chap. 1 speaking of Chomsky. Marr uses "description", reserving "representation" to designate the formal system descriptions belong to; although he informally speaks of descriptions *of* entities, he explicitly emphasizes the non-relational nature of descriptions:

> A *representation* is a formal system for making explicit certain entities or types of information […]. I shall call the result of using a representation to describe a given entity a *description* of the entity in that representation […]. For example, a representation for shape would be a formal scheme for representing some aspects of shape, together with rules that specify how the scheme is applied to any particular shape. […] The phrase "formal scheme" is critical to the definition […]. To say that something is a formal scheme means only that it is a set of symbols with rules for putting them together - no more and no less. (Marr 1982: 20–21)

[11] On the notion of feature see McGilvray (1998: Sect. 3.2.3).

or is associated with a collection of features; there is matching between an object file o and a stored object o' when there are some terms belonging to them which share a sufficient number of features. The problem is how to characterize the 'sufficient' number. My suggestion is that a number n is sufficient for $< t, t' >$ ($t \in o$, $t' \in o'$) iff the fact that t and t' share n features is a sufficient condition for o and o' to be identical. At this point the operation of matching can be defined as shown in (4)(ii) below.

Is it legitimate to extrapolate from this approach to visual perception a view of how objects in general are given to subjects in prelinguistic cognitive states? That is my working hypothesis. Cognitive states are conceived as the states of components of the mind that might be called, following Chomsky, *conceptual-intentional systems*. I will introduce the drastically simplifying assumption that, for every subject, there is a unique conceptual-intentional system, **CIS**, with a unique internal representational system, **IRS**. Further assumptions are the following:

(1) Sensory organs, memory and imagination have access to **CIS**.
(2) Cognitive states subsist in time; a cognitive state is individuated, among other things, by the activation, at a certain time, of one or more mental representations or descriptions, more abstractly of terms of **IRS**.
(3) In long-term memory a catalogue of known objects is stored, which represents their parts, properties, and categories of membership through a system of features and labels.
(4) Some component of a subject's mind (maybe **CIS**[12]) implements algorithms computing:

  (i) a function *const* such that, if *t* is a term of **IRS** and σ a cognitive state, $const_t(\sigma) = \{t' | \text{there is in } \sigma \text{ one relevant equivalence relation } R_\sigma$[13] such that *t'* is a term of **IRS** and $tR_\sigma t'\}$.[14]
  (ii) a function *id* such that, if o and o' are two stored objects in σ, *id*(o,o') = 1 iff they contain the same elements (i.e., the same terms of **IRS**).
  (iii) a function *inf* such that, if t is a term of **IRS** and σ is a cognitive state, $inf_t(\sigma)$ is a set of features.

---

[12] This conjecture is based, on the one hand, on the hypothesis (Hauser, Chomsky & Fitch (2002)) that **CIS** is part of the Faculty of Language in the Broad Sense, on the other, on Spelke's following remarks about the central, amodal character of the mechanisms ruling object perception:

> [Y]oung infants can sometimes perceive the unity of partly hidden objects, the boundaries of adjacent objects, and the identity or distinctness of objects that move fully out of view. The mechanisms that accomplish these tasks appear to be central in three respects: (1) They take as input representations of three-dimensional surface arrangements and motions, (2) they are amodal, and (3) they yield representations of parts or states of an object that cannot be seen directly. (Spelke 1990: 48)

[13] It is up to psychology to specify, for any cognitive state, the relevant equivalence relation. Examples are "having the same spatio-temporal location", "having the same face", "having the same form", etc.

[14] I will also denote with "$/t/_{R_\sigma}$" the equivalence class of t induced by R_σ.

(iv) a function *match* such that, if o is an object file in σ and o' a stored object in σ, $match(\langle o,o' \rangle, \sigma) = 1$ iff there are terms t, t' belonging to o and o', resp., and features $f_1, \ldots, f_n$ such that their occurrence in both $inf_t(\sigma)$ and $inf_{t'}(\sigma)$ is sufficient for o and o' to be identical.

Under these assumptions, we can say that a $\mathcal{C}$-object is *given* when, through the activation of a term, an object file (i.e. a set of terms) is created by means of *const* and, by means of *match*, it is either recognized, i.e. identified with a stored object, or not recognized and 'created', i.e. added to the catalogue as a new entry.

Let me conclude by observing that, in passing from one cognitive state to another, the cognitive domain (i.e. the catalogue of stored $\mathcal{C}$-objects) may be updated, in the sense that a $\mathcal{C}$-object may acquire or lose some terms; or an old $\mathcal{C}$-object may be suppressed; or a new $\mathcal{C}$-object may be created; or one $\mathcal{C}$-object may split into two; or two $\mathcal{C}$-objects may merge into one. I am not concerned with the mechanisms underlying these operations; I shall simply assume that such operations can be characterized as computations.

Gareth Evans has called into question the claim that recognition can be explained in the way suggested above (Evans, 1982: Sect. 8.5). As far as I can see, Evans' argument can be reconstructed as follows. Concentrating upon the concept of similarity, or 'looking like', involved in every explanation of recognition along the lines suggested above, he remarks that

> for one thing to strike me as like another [...] is not a judgement, to which the question of truth or falsity can significantly be applied [...], which can be only when the possibility of error has been provided for. (Evans, 1982: 293–294)

The only alternative account, according to Evans, is that

> for one thing to strike me as like another is simply a *reaction* which those things occasion in me [.]. (Evans, 1982: 293)

At this point he poses the crucial question: «Is there any necessity that what would remind X of Y will remind people in general of Y?» (Evans 1982:294) According to him there is no good answer to this question:

> Might not X be rather insensitive to certain properties of faces to which other humans are very sensitive, but sensitive to other features which other humans ignore, so that the picture that the police artist produces is a good likeness for X but not for others? Why should X not have an idiosyncratic similarity space in this area? (*Ibid.*)

moreover—he argues—the retreat to the claim that the relevant appearance-property is that of looking like this *to X*, hence a *private* concept, would be disastrous:

> Anyone influenced by Wittgenstein will argue that there *is* no such concept, no such property. A concept is something abstracted from the practice of judging - a capacity exercised by someone in the course of making a judgement; which is a performance assessable as being correct or incorrect. (Evans, 1982: 295)

However, there is an account of looking like, different from conceiving it as either a judgement or a reaction, which Evans does not take into consideration; it consists in conceiving it as a computation. Remember that, on the one hand, the ‘matching’ operation does not take as inputs objects, but mental representations, i.e. structured objective information; on the other hand, perceptual and memory modules of an idealized subject can be conceived as computing devices acting according to some algorithms, where an algorithm is individuated by the totality of its instructions; matching is therefore a computable operation whose input is a pair of mental representations and the output 1 or 0. Concededly, it is not a judgement in Evans’ sense: the question of truth or falsity cannot significantly be applied to it, as matching does not take as inputs objects and mental representations, but pairs of mental representations; but this is exactly the reason why it is a computation, and not a reaction: there *is* a necessity that what is recognized by a subject $\mathcal{S}$ as the $\mathcal{C}$-object *o* is recognized as *o* by any other subject $\mathcal{S}$’ occupying the same cognitive state. It is a methodological assumption, fundamental in psychology, that a subject’s response is the outcome of the interaction of two factors: information available to the subject $\mathcal{S}$ on a certain occasion, and $\mathcal{S}$’s cognitive apparatus. Input information is perfectly objective, in the sense that it is accessible to all the subjects who share a cognitive ‘position’ identifiable in objective terms; for instance, the input to the visual apparatus is, in Marr’s terminology, an ‘image’, an array of intensity values. On the other hand, it is a normal idealizing assumption that the cognitive apparatus of all human beings is essentially the same; of course it is possible that $\mathcal{S}$ has an idiosyncratic similarity space in some area, but this poses (possibly very difficult) practical problems of discrimination between ‘normal’ and idiosyncratic features, not theoretical problems of illegitimacy of idealization, exactly like the fact that some people have the heart to the right does not render illegitimate the idealization that the heart is located slightly to the left. Of course, putting the matter in this way is equivalent to *not* being «influenced by Wittgenstein», i.e. by Wittgenstein’s anti-realism about the mental.

4.1.3 A Manageable Concept

Intuitively one manages the concept HORSE if one can discriminate horses from other objects; under this view the concept HORSE can be equated with a function f taking as arguments objects and such that f(x) = 1 iff x is a horse. Analogously one manages the binary concept (or relation) ‘x is longer than y’ if one can discriminate the pairs of objects standing in that relation from other pairs; under this view the binary concept (or relation) ‘x is longer than y’ can be equated with a function f taking as arguments pairs of objects and such that $f(<x, y>) = 1$ iff x is longer than y. This view of concepts as discrimination functions is essentially Fregean. But Fregean concepts take as arguments (n-tuples of) objects of the external world, whereas $\mathcal{C}$-objects, as they are conceived here, are sets of representations[15]; moreover, Frege was not concerned with the manageability of concepts.

[15] As we have seen, they are not simply sets of representations, but equivalence classes; but this is not relevant here.

If we want to adopt this functional view of concepts we must therefore adapt it to the framework of an internalist semantics, by equating $\mathcal{C}$-concepts with functions mapping n-tuples of $\mathcal{C}$-objects onto $\{1,0\}$, which will no longer be understood as truth-values, but as "Yes" or "No" answers the computational apparatus of **CIS** associates to those inputs; and by equating managing a $\mathcal{C}$-concept with being capable to compute that function. However, while it is intuitively clear what it is for an object of the external world to be a horse, or to be longer than another, it seems less clear what it is for a $\mathcal{C}$-object to have that property, or for two $\mathcal{C}$-objects to stand in that relation. But this is not the relevant question; the notion I am trying to define in this chapter, justification for an atomic sentence, is intended to be an *explicans* of the intuitive conditions at which that sentence is *evident*. So the intuitive relevant question is: under which conditions is it *evident* that a given object has a property, or that two objects stand in a certain relation?

As in the case of object perception, the answer is up to psychology, not to semantics; up to semantics is to lay a framework 'compatible' with good psychological theories and explanations. In order to give an idea of what I mean by "good" I shall first describe an analysis of visual relations I judge not good, and I'll try to explain why I find it not good; this will help to introduce an example of a good theory. Let us start from an idea sketched by Marr in the final chapter of his book on *Vision*:

> The perception of an event or of an object must include the simultaneous computation of several different descriptions of it that capture diverse aspects of the use, purpose, or circumstances of the event or object.
>
> [...] The various descriptions [...] include coarse versions as well as fine ones. These coarse descriptions are a vital link in choosing the appropriate overall scenarios demanded [to tackle the problem of multiple descriptions of objects (*g.u.*)] and in correctly establishing the roles played by the objects and actions that caused those scenarios to be chosen. (Marr, 1982: 358)

Suppose we take this idea concerning the perception of events as a suggestion for an analysis the perception of (some kinds of) relations; from this point of view, the perception of a dog pursuing a cat would be the perception of a relation of pursuing between a dog and a cat in the roles, respectively, of agent and patient; and such a relation would be perceived when the actually derived description matched a stored model of the pursuing relation between a dog and a cat.

One crucial drawback of this idea is that it does not account for the capacity of our visual system to perceive, i.e. to compute, *abstract* relations. The point is clearly illustrated by S. Ullman; he considers the spatial relation "x is inside y", where x is a small X figure and y a single closed curve:

> the concept of being "inside" is abstract, because it does not refer to any particular shape, but can appear in many different forms. More formally, [...] a relation such as "inside" defines a set of configurations that satisfy this relation. Clearly, in many cases the set of [configurations] S that satisfy a [relation] P can be large and unwieldy. It therefore becomes impossible to test a [configuration] for [relation] P by simply comparing it against all the members of S stored in memory. To be more accurate, the problem lies in fact not simply in the size of the set S, but in what may be called the size of the *support* of S. (Ullman, 1996: 274)

The supports of S are, roughly, the properties with respect to which a newly derived configuration is compared to a stored one:

> When the set of supports is small, the recognition of even a large set of objects can be accomplished by simple template matching. This means that a small number of patterns is stored, and matched against the figure in question. When the set of supports is prohibitively large, a template matching decision scheme will become impossible: we cannot store, for example, all instances of closed curves in the image. (Ullman, 1996: 275)

The moral to draw is that, if our perception of spatial relations were explained in terms of recognition and matching, that perception could not be described as a computational process; however, our visual system can efficiently compute abstract relations; this clearly indicates that the preceding analysis is not good, and that a different kind of explanation is required.

Ullman's own explanation is based on the idea that our perception of shape properties of objects and their spatial relations consists in the execution of *visual routines*:

> Consider, for instance, the task of comparing the lengths of two line segments. Faced with this simple task, a draftsman may measure the length of the first line, record the result, measure the second line, and compare the resulting measurements. When the two lines are present simultaneously in the field of view, it is often possible to compare their lengths by 'merely looking'. This capacity raises the problem of how the 'draftsman in our head' operates, without the benefit of a ruler and a scratchpad. More generally, a theory of the perception of spatial relations should aim at unraveling the processes that take place within our visual system when we establish shape properties of objects and their spatial relations by 'merely looking' at them. (Ullman, 1984: 99)

A routine for the relation R is therefore a 'perceptual program' executed by the visual processor to compute a relation R. A noteworthy aspect of this explanation is that conceiving spatial relations as visual routines permits, as Ullman shows, to meet three basic requirements that should be imposed onto the 'visual processor':

> The three requirements are (i) the capacity to establish abstract properties and relations (abstractness), (ii) the capacity to establish a large variety of relations and properties, including newly defined ones (open-endedness), and (iii) the requirement to cope efficiently with the complexity involved in the computation of spatial relations (complexity). (Ullman, 1996: 273-274)

Since similar requirements should be imposed onto many other information-processing systems, Ullman's proposal appears promising for a routine view of a number of emprical relations.

Another nice aspect of the routine view of concepts is that it is absolutely generic, in the sense that no requirement is imposed onto the general notion of routine except that it is an algorithm for the computation of what I have called above a discrimination function.[16] Hence my tentative answer to the intuitive question raised above would be that is it *evident* that a given object has a property, or that two objects stand in

[16] Visual routines are not *only* this: «visual routines should not be thought of merely as predicates, or decision processes that supply 'yes' or 'no' answers.» (Ullman 1984: 114).

a certain relation, when a routine is available to discriminate among objects having and not having that property, or among pairs of objects standing and not standing in that relation.

I shall therefore conceive an n-ary $\mathcal{C}$-concept C as a function f_C from $\mathcal{C}$-objects to {0,1} such that $f_C(x_1,\ldots,x_n) = 1$ iff $x_1,\ldots,x_n$ stand in relation C. The function is the *goal* of the routine, which is individuated, and denoted with the name "C", by the psychologist, not by the subject; what the subject does is to follow that routine.[17] Under these assumptions, the intuitive expression "the concept C″ is systematically ambiguous between the function f_C and the routine computing f_C. Given a cognitive state σ, the n-ary $\mathcal{C}$-concept C is *manageable* in σ iff a program (or routine)[18] computing f_C is available in σ.

4.2 Linguistic Atomic Cognitive States

In human mind **CIS** can be accessed also by the faculty of language, in the sense that the syntactic structures generated by syntax are inputs for **CIS**, which is dedicated to their (semantic) interpretation. In particular **CIS** receives expressions of two different sorts, singular terms and predicates,[19] whose possible uses must meet some general cognitive preconditions; knowledge of these preconditions is plausibly part of our semantic competence.

---

[17] From the psychologist's point of view the individuation of the function comes before the individuation of the algorithm actually employed by the perceptual system. (For a clear illustration of the strategy of a computational theory of information-processing systems see Marr 1982).) Of course the empirical problem arises of «what triggers the execution of different routines during the performance of visual tasks» (Ullman 1996: 279–280); Ullman's suggests a strategy:

> It seems to me that this problem can be best approached by dividing the process of routine selection into two stages. The first stage is the application of what may be called *universal routines*. These are routines that can be usefully applied to any scene to provide some initial analysis. They may be able, for instance, to isolate some prominent parts in the scene and describe, perhaps crudely, some general aspects of their shape, motion, color, the spatial relations among them etc. These universal routines will provide sufficient information to allow initial indexing to a recognition memory, which then serves to guide the application of more specialized routines. (Ullman 1984: 116)

[18] «These stored visual routines constitute "perceptual programs"» (Ullman 1996: 313).

[19] Under the assumption that the reference language is a first order one like $\mathcal{L}$, of course. The more substantial assumption I am making here is that categories similar to predicates and singular terms are present in natural languages as well.

4.2.1 $\mathcal{C}$-Authorizations to Use a Singular Term to Refer to a Given Object. Denotation

A plausible assumption concerning singular terms is that in every cognitive state a certain set of features is attached to each singular term. I will therefore postulate, as a component of a subject's cognitive apparatus, a function *ec* associating to every singular term ν and every cognitive state σ a certain set of features stored in memory, which I call the *epistemic content* $ec_{\nu,\sigma}$ associated to ν in σ. I assume as well that $\mathrm{ec}_{\nu,\sigma}$ is articulated into a *lexical content* $lec_{\nu,\sigma}$, constituted by information coming from the lexicon, and a *situational content* $sec_{\nu,\sigma}$, constituted by information coming from perception, memory, imagery, the belief system, etc.[20] In general, information contained in $\mathrm{lec}_{\nu,\sigma}$ is far from permitting to individuate one $\mathcal{C}$-object as its denotation, but it is sufficient to assign the label of a kind or type to its denotation (for example the label HUMAN BEING to the denotation of "John"), and consequently to apply to its denotation the appropriate criterion of identity (personal identity, in the case of John). Which features are to be put into $lec_{\nu,\sigma}$ and which in $sec_{\nu,\sigma}$ is partly an empirical question pertaining to lexical semantics, partly a question depending on our intuitions about synonymy and related notions; I shall try to make some of those intuitions explicit in Sect. 4.6. From the conceptual point of view the important thing is that *some* divide between lexical and situational content is acknowledged, in order to avoid Quinean inextricability and its holistic consequences.[21] From the intuitive point of view we may remark that, normally, $\mathrm{lec}_{\nu,\sigma}$ remains invariant in passing from σ to σ', while $\mathrm{sec}_{\nu,\sigma}$ is expected to vary considerably; variations in $\mathrm{lec}_{\nu,\sigma}$ are possible as well, but in principle they will be felt as reasons for splitting ν into two different terms ν_1 and ν_2.

I shall not deal with the problem of how information contained in the epistemic content is organized, represented, or accessed; I will simply assume that it is computationally tractable, as it happens in the case of information encoded into lexical entries through features and labels.

Concerning the notion of $\mathcal{C}$-authorization to use a singular term to refer to a given object, the intuitive idea I am pursuing is to define it as a cognitive state in which the given $\mathcal{C}$-object is, in a sense to be defined, recognized as given also by the singular term. Presumably 'linguistic' recognition is not essentially different from perceptual recognition, in the sense that, from a very abstract point of view, the necessary ingredients of both kinds of recognition are the same: stored information, new information, and an operation of comparison or 'matching' between the two.

We have seen that in a prelinguistic cognitive state stored information is contained in the catalogue; in a linguistic one also the epistemic content associated to singular terms is to be considered as stored information.

As for the operation of matching, we can observe that, in linguistic cognitive states, the domain of the function *match* becomes much larger, since there may be

[20] An analogous distinction is introduced and motivated in Bierwisch (1992: 30–32).

[21] See Chap. 3 and, below, Sect. 4.7.1.

matching between the epistemic content associated to a singular term and information associated to a model stored in memory; but the function is defined in the same way as in the prelinguistic case.

Let us consider new information. We have seen that in a prelinguistic cognitive state σ new information is contained in the activated term, and that which term is activated is determined by what is perceived, or focused on, at the temporal index i of σ. This holds also in linguistic cognitive states, but in them also something essentially new happens. The intuitive phenomenon is that, when I hear a friend of mine speaking of Paul, or when I read in a book the sentence "Consider the square root of 64", I may focus on Paul and understand my friend's speech about *him*; moreover, I not only implicitly know that my friend is speaking of *only one* person, but also I myself keep that person distinct from any other object; similarly, I may follow a long argument about the square root of 64, and I have an analogous intuition of uniqueness about the square root of 64. In other terms, in a linguistic cognitive state a *specific* object may be given also through the use of a singular term.[22] How can this be conceived as a computational process? We cannot say that an object file is set up in the same sense as in the visual case; for example, it makes no sense to say that in the case of "Paul" a temporary object is set up which is «addressed by its location at a particular time»; however, also in these cases there is a difference between being given an object and identifying it: I may focus on Paul even if I have no idea of who Paul is (except what I can grasp from lexical epistemic content associated to the name, i.e. that he is a male human being), so that he is an entirely new object of my mental domain; analogously, I may reason about the square root of 64 even if I do not recognize it as the number 8. The difference between object files and $\mathcal{C}$-objects seems therefore to be significant also when objects are given through singular terms: a $\mathcal{C}$-object is an object file that is either recognized, i.e. identified with a stored object, or added to the catalogue as a new entry. So, if we want to give a computational account of this new way of giving $\mathcal{C}$-objects by means of names or descriptions, the starting question is: which is the activated term, in these cases?

It may be helpful to remark that, when my friend speaks of Paul, (s)he is performing the linguistic act we intuitively qualify as *referring* to a specific person; hence our question must be placed on the background of a more general one: how can we give a computational account of referring? Since we can say that, when a subject in a cognitive state σ refers with ν to o, the denotation of ν in σ is o, our general question can be restated as follows: how can we give a computational account of denotation in σ?[23] As we will see, once this question has been answered, also the question of which is the activated term will have got an answer.

[22] Another interesting example of such a linguistic way of giving objects is offered, in natural languages, by pronouns in certain syntactic constructions. We have observed in Sect. 1.1.2.1 of Chap. 1 that in the sentence "He thinks that John is abroad" the pronoun cannot be coreferential with the name; presumably, our computing apparatus introduces into our cognitive domain a $\mathcal{C}$-object *different* from the denotation of "John".

[23] For interesting implementations of a computational approach to the notion of denotation see Tichy (1969), Dummett (1975), Moschovakis (1993), Martin-Löf (2001).

Recognition plays an essential role also at the linguistic level; it is therefore natural to postulate the existence of a linguistic matching operation *MATCH*, which will be slightly different from the prelinguistic one in that it compares stored objects not with object files, but with epistemic contents associated to names, and is consequently conceptually simpler than the prelinguistic one. I will therefore add to (1)–(4) the following assumptions:

(5) Singular terms of $\mathcal{L}$ are treated by **CIS** as terms of **IRS**.
(6) Some component of a subject's mind (maybe **CIS**) implements algorithms computing:

  (i) the function *MATCH* such that, if $ec_{\nu,\sigma}$ is the epistemic content associated to the name ν in the cognitive state σ and o is an object of $\mathcal{D}_\sigma$ (the cognitive domain of σ), $MATCH(ec_{\nu,\sigma},\mathrm{o}) = 1$ iff there is a term t belonging to o such that all features occurring in $ec_{\nu,\sigma}$ occur in t.
  (ii) the function *den* taking as arguments names and domains of cognitive states and defined in the following way:

Definition 1 $den(\nu,\mathcal{D}_\sigma) = \mathrm{o}$ iff
$\mathrm{o} \in \mathcal{D}_\sigma \wedge MATCH(ec_{\nu,\sigma}, \mathrm{o}){=}1 \wedge \forall \mathrm{o}'((\mathrm{o}' \in \mathcal{D}_\sigma \wedge MATCH(ec_{\nu,\sigma}, \mathrm{o}') = 1) \rightarrow \mathrm{o} = \mathrm{o}')$.

To illustrate the meaning of this definition let us distinguish two situations. In the first John is speaking with a person in front of him; we can say that he occupies a cognitive state σ in which the activated term is a description t of a face and t gives (i.e. belongs to) a $\mathcal{C}$-object o of $\mathcal{D}_\sigma$ classified as a male human being; suppose further that the epistemic content associated in σ to the name "Paul" contains, besides the lexical features MALE and HUMAN, the description t' of a face, and that t' MATCHES the object o; since it can plausibly be assumed, as a part of 'perceptual' competence, that no two human beings have exactly the same face, the term t satisfies the uniqueness conditions of Definition 1, hence $den(\mathrm{Paul},\sigma) = \mathrm{o}$.

In the second situation a friend of John's is speaking to him about Ricciolino, whose name John hears for the first time; in this case no term is activated by perception or attention or memory; however, consider the name "Ricciolino" itself: by assumption (5) it is a term of **IRS**; moreover, since syntactical identity ≈ is an equivalence relation,[24] and it is the only relevant one in σ,[25] $/\text{"Ricciolino"}/_{\approx\sigma} \in \mathcal{D}_\sigma$, hence $const_{Ricciolino}(\sigma) = /\text{"Ricciolino"}/_{\approx\sigma}$; obviously $ec_{Ricciolino}$ MATCHES $/\text{"Ricciolino"}/_{\approx\sigma}$, because the features occurring in "Ricciolino" are the same as those occurring in $ec_{Ricciolino}$; finally, $/\text{"Ricciolino"}/_{\approx\sigma}$ is the singleton {"Ricciolino"}, because if t ≈ "Ricciolino" then $/\mathrm{t}/_{\approx\sigma} = /\text{"Ricciolino"}/_{\approx\sigma}$; hence all the conditions

[24] Two expressions e and e' are syntactically identical iff they are tokens of the same syntactical type; I assume that phonological equivalence is constitutive of syntactical identity.

[25] It is the only relevant one because it permits to define a $\mathcal{C}$-object satisfying the uniqueness condition of Definition 1, and other possible candidates (in particular the relation of belonging to a $\mathcal{C}$-object which matches the epistemic content of "Ricciolino") do not.

of Definition 1 are satisfied, and *den*(ν,σ) = {"Ricciolino"}. We have got an answer to our starting question as well: the activated term is the name "Ricciolino".[26]

The application of Definition 1 to the second example shows also one important case of linguistic updating of the cognitive domain: the creation of a new object file, which will become a $\mathcal{C}$-object when it will be added to the catalogue, i.e. to the cognitive domain.[27]

Other cases of linguistic updating of the cognitive domain are the splitting of one $\mathcal{C}$-object into two, and the fusion of two $\mathcal{C}$-objects into one. In order to illustrate the former, suppose John, in his initial cognitive state, believes that Plato was the the author of the *Phaedo* and the tutor of Alexander the Great; later on, in a bookshop, he finds a book about Aristotle, and he reads that he was the tutor of Alexander; now John has several options, i.e. several possible explanations of the data at his disposal: (a) Plato was called also "Aristotle"; (b) There are two persons, Plato and Aristotle, the tutor of Alexander; (c) the book he has in his hands is an April fool's joke; and so on. To make a choice John needs some selection criteria, and perhaps to acquire more information; suppose that, after this work, he selects (b), entering a new cognitive state in which he associates to "Aristotle" the epistemic content TUTOR OF ALEXANDER and to "Plato" the epistemic content AUTHOR OF THE *PHAEDO*: through the procedure described above he constitutes two distinct $\mathcal{C}$-objects from the one he named "Plato".

We can at last answer the central question addressed in this section:

Definition 2 The *denotation of* the name ν *in* the atomic cognitive state[28] σ (in symbols $[[\upsilon]]_\sigma$) is *den*(ν, σ).

Definition 3 A *cognitive authorization to use a singular term* ν *to refer to the* $\mathcal{C}$*-object o* is an atomic cognitive state σ such that $[[\upsilon]]_\sigma = \text{o}$.

4.2.2 $\mathcal{C}$-Authorizations to Use a Predicate to Apply a Concept to $\mathcal{C}$-Objects

What does it mean to know the meaning of a predicate? Like in the case of names, my starting point is that it amounts to being able to recognize a cognitive authorization*:* a *cognitive authorization to use the predicate in order to apply a manageable concept to objects.*

Let me first explain why I use this involved and somewhat abstruse expression instead of the much more simple "authorization to concatenate a predicate with a

[26] In this case the name is a sort of placeholder, and the given object, in a sense, a virtual object.

[27] Plausibly, it will be added to the catalogue when the epistemic content of the name will be enriched with a set of features that, according to principles of the subject's cognitive apparatus, can be jointly attributed to no more than one object.

[28] Atomic cognitive states are defined by Definition 9.

name". I speak of application of a *concept* (I skip "manageable" for a while) to an *object*, instead of concatenation of a *predicate* with a *name*, because I want to stress that application is an operation involving concepts and objects, i.e. the entities *denoted* by predicates and names, and not directly predicates and names. This is correct, I hold, even for a computational approach like the present one (which involves an internalist notion of denotation, as we have seen); the difference from the realistic view concerns the nature of the entities denoted by names and predicates, not the fact that such entities are *distinct* from linguistic entities. The main reason for this is that an important aspect of the cognitive preconditions for the use of predicates is that if a subject is authorized to concatenate a predicate with a name, then (s)he is authorized to concatenate it with any other name of the same object, provided (s)he is authorized to believe that it is a name of the same object. If I am justified to assert, for instance, that the boy in front of me is running, then I am thereby justified to assert that Matthew is running, and that the elder son of my brother is running, provided I am justified to believe that the boy in front of me is Matthew, the elder son of my brother. In more solemn terms we might say that predication, the operation of concatenating a predicate with a name, has an implicit *modal* aspect, in the sense that we do not simply ask ourselves whether we are authorized to concatenate a predicate with a given name, but with any other name we *could* use to refer to *the same object*. This seems to be another important reason why names, and more generally singular terms, cannot simply pick out terms of the internal representation system, but must be used to refer to *objects*.[29]

An immediate consequence of this is that what applies to objects is *concepts*, in the Fregean sense of entities having the nature of functions. But it should be stressed that Frege never speaks of *concepts* as applying to objects, but directly of predicates. As Dummett (1981: 246) observes, for Frege «the crucial notion for the explanation of the sense of a predicate is that of its being true of an object [...]». As a consequence, «the relation between [a predicate] and its referent [i.e., a concept] does not have to be invoked» (*ibid.*); nor *could* it be invoked – I add – because «we can make no suggestion for what it would be to be given a concept.»[30] An almost immediate consequence of this idea is that «The only way we can gain an idea of [a concept] is as the referent of a predicate, [...] we approach it – apprehend it – via language» (Dummett, 1981: 202); and a consequence of this thesis is that neither a human being has concepts before the acquisition of a language, nor a non-human animal has access to concepts. I find this conclusion untenable for many reasons; for one, it is incompatible with the claim, strongly supported by evidence, that

> humans indeed have early-developing core knowledge systems, and these systems permit a range of highly intelligent behaviors and cognitive capacities [...]. In each case [...] nonhuman animals have been found to have capacities that equal or exceed those of human infants. The core knowledge systems that have been studied in human infants so far therefore do not account for uniquely human cognitive achievements. (Spelke, 2003: 289)

---

[29] A first, related reason has been given in Sect. 4.1.1.

[30] Dummett (1981): 241. See also p. 408: «the notion of identifying a concept [...] seems quite inappropriate.»

For this reason I think the mention of 'a manageable concept' is essential in the statement of the starting question concerning predicates: concepts (some concepts) are manageable independently of language. I don't say that they are *given*, like objects, but that they are *manageable*, and that subjects manage them before language comes in.

Like in the case of singular terms, I postulate that in every cognitive state σ a certain set of features is attached to each predicate π as its *epistemic content* $ec_{\pi,\sigma}$, about which I make the assumption, analogous to the one about singular terms, that it is articulated into a lexical content $lec_{\pi,\sigma}$, constituted by information coming from the lexicon, and a situational content $sec_{\pi,\sigma}$, constituted by information coming from perception, memory, etc. The organization of lexical information concerning verbs and other predicates is of the competence of linguistics, and I shall not venture into this field.

In Sect. 4.1.3 I have characterized prelinguistic $\mathcal{C}$-concepts as programs for the computation of discrimination functions, and I have stressed that the individuation of each function is of competence of the psychologist, not of the subject; in a linguistic cognitive state the possibility arises that a prelinguistic $\mathcal{C}$-concept is denoted by a predicate, hence that the subject himself is in a position to consciously manage it. Take for example the unary concept DOG, which I assume to be manageable by an infant; this means that a routine p_{DOG} is available to the infant that computes the function f_{DOG}. Suppose now that her mum shows the infant a dog and tells her "That's a dog": the infant occupies now a cognitive state σ in which the predicate "dog" is associated to the routine p_{DOG}; in other terms, in σ the predicate "dog" denotes the function f_{DOG}, or the $\mathcal{C}$-concept DOG.

While prelinguistic manageable concepts can be assumed to be natural denotations of primitive predicates, the availability of information coming from the lexicon enormously increases the number of manageable $\mathcal{C}$-concepts. New $\mathcal{C}$-concepts may be introduced by associating epistemic contents to predicates, for example through definitions, as when we define the concept PRIME NUMBER; or through the splitting of an old concept into two or more ones, as when the old concept of STAR was split into the concepts of STAR, PLANET, GALAXY, etc.; or in other ways. Take for example the introduction of the predicate "Unicorn" through the stipulation that a unicorn is a horse with a single horn; by means of the definition the epistemic content associated to the predicate "Unicorn" will include the features HORSE, WITH ONE HORN; under the assumption that programs computing f_{HORSE} and $f_{WITHONEHORN}$ are manageable in σ, it will be possible to assemble a program for the computation of the complex function $f_{HORSEWITHONEHORN}$. For example, the routines for f_{HORSE} and $f_{WITHONEHORN}$ might be 'label checking' routines, checking the presence of the labels HORSE and WITH ONE HORN, respectively, in stored models matching with newly derived descriptions[31]; in this case the routine for $f_{HORSEWITHONEHORN}$ would

[31] Notice that, in the long passage from Kahneman, Treisman & Gibbs (1992) quoted above, the possibility is admitted that sensory information about an object is «matched to stored descriptions to identify *or classify* the object» (*emphasis added*).

consist in checking the presence of the two labels HORSE and WITH ONE HORN in one stored model.

The preceding discussion should justify the following definitions:

Definition 4 The *denotation of* the predicate π *in* the atomic cognitive state σ (in symbols $[\![\pi]\!]_\sigma$) is an algorithm computing the function f_C iff f_C is defined by $ec_{\pi,\sigma}$.

Definition 5 A *cognitive authorization to use* π *in order to apply the concept C to objects* is a cognitive state σ such that $[\pi]_\sigma = f_C$.

A final note concerning identity. In (4)(ii) I have postulated the existence, already at the prelinguistic level, of a function *id* corresponding in fact to the $\mathcal{C}$-concept of (extensional) identity f_{id} between $\mathcal{C}$-objects. At the linguistic level I postulate the existence of an (objectual) identity predicate, i.e. of a binary predicate "=" such that

(7) In any cognitive state σ of any subject, $ec_{=,\sigma}$ contains the piece of information that two $\mathcal{C}$-objects are identical if, and only if, they have the same elements (i.e. the same terms of **IRS**).

As a consequence "=" denotes, for every subject, f_{id}. This postulate is justified by the assumption that the notion of objectual identity is innate, as well as the notion of object. An analogous assumption will be made about the logical constants, as we will see in Chap. 5.

I postulate also the existence of a conceptual identity predicate, i.e. of a binary predicate "≡" such that

(8) In any cognitive state σ of any subject, $ec_{\equiv,\sigma}$ contains the piece of information that two $\mathcal{C}$-concepts are identical if, and only if, they apply to the same $\mathcal{C}$-objects.

The rationale for this choice will be discussed in Sect. 4.6.1.2.

4.3 The Justification Question and the Problem of Relevance

If an atomic cognitive state σ has been specified, C^n is a $\mathcal{C}$-concept manageable in σ and $o_1, \ldots, o_n$ are $\mathcal{C}$-objects of $\mathcal{D}_\sigma$, then the answer to the questions

(9)

(i) The *Application Question* for $f_C(o_1 \ldots o_n)$ in σ:
Which is the value of $f_C(o_1 ... o_n)$ in σ?[32]

(ii) The *Application Question* for $f_=(o_1 o_2)$ in σ:
Which is the value of $f_=(o_1 o_2)$ in σ?

(iii) The *Application Question* for $f_\equiv(C_1 C_2)$ in σ:
Which is the value of $f_\equiv(C_1 C_2)$ in σ?

[32] To say that the value of $f_C(o_1 \ldots o_n)$ is 1 in σ I will use "$f_C(o_1 \ldots o_n) =_\sigma 1$".

is determined by a computation, as we have seen. It is important to distinguish these questions from the following:

(10)

(i) The *Justification Question* for $\pi(\nu_1, \ldots, \nu_k)$ in σ:
When is $\pi(\nu_1, \ldots, \nu_n)$ justified in σ?

(ii) The *Justification Question* for $\nu_1 = \nu_2$ in σ:
When is $\nu_1 = \nu_2$ justified in σ?

(iii) The *Justification Question* for $\pi_1 \equiv \pi_2$ in σ:
When is $\pi_1 \equiv \pi_2$ justified in σ?

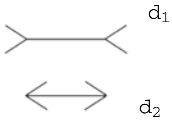
Someone might be induced to think that, if $[\![\pi]\!]_\sigma = f_C$, $[\![v_1]\!]_\sigma = o_1, \ldots, [\![v_n]\!]_\sigma = o_n$, the answers to (9) and (10) are 1 under the same conditions; but it would be a serious error. The answer to (9) is determined by the mere computation of f_C, $o_1, \ldots, o_n$; also the answer to (10) is computationally determined, but the principles governing the latter computation are different from those governing the former, as I will explain in a moment. Before that, let us consider some examples intended to show the gap between (9) and (10).

Case 1. Consider the cognitive state σ_1 of Mary who is sitting in position p_1, looking at a disk d placed on a table, and sees d as round; according to the definitions given above, $f_{ROUND}(d) =_{\sigma 1} 1$; imagine now that at a subsequent time Mary moves to a position p_2, from which she sees the disk as elliptical; in Mary's new cognitive state σ_2, according to our definitions, $f_{ELLIPTICAL}(d) =_{\sigma 2} 1$. It is not intuitively correct to say that in σ_2 Mary is justified to believe that that disk is elliptical, since she remembers that in p_1 the disk looked round, and she knows that being round and being elliptical are incompatible properties; Mary will probably be uncertain about the shape of the disk—in my terminology, $j_{\sigma 2}$(ELLIPTICAL(d))) will be undefined. The example shows that, for the answer to (10) to be 1, it is not sufficient that the answer to (9) is 1. However, after a while Mary will decide that

(11) That disk is round;

why?

Case 2. In the well-known Müller-Lyer illusion the horizontal line d_1 in (12) looks shorter than d_2:

(12)


even if, having measured them with a ruler, we know that

(13) d_1 and d_2 are the same length:

a case in which $f_{SAMELENGTH}(o_1,o_2) =_\sigma 0$ but j_σ(are-the-same-length(d_1,d_2)) should intuitively be 1. The example shows that, for the answer to (10) to be 1, it is not even necessary that the answer to (9) is 1.

The gap between (9) and (10) is in fact much deeper than these examples show, since there are many cognitive states in which the value of $f_C(o_1 \ldots o_n)$ cannot be computed because the $\mathcal{C}$-objects $o_1, \ldots, o_n$ are not perceptually given, but we are intuitively justified to believe the proposition expressed by an atomic sentence $\pi(\nu_1,\ldots,\nu_k)$ such that π denotes f_C and $\nu_1,\ldots,\nu_k$ denote, respectively, $o_1, \ldots, o_n$—I will call *indirect* such cases. Here is a couple of examples.

Case 3. Suppose Ann hears some noises in the room nearby where, as a matter of fact, Jack is running. If she had no other information, Ann would not be justified to believe that.

(14) Jack is running in the room nearby;

but suppose she has at his disposal the following supplementary pieces of information (codified by features): (a) that in the room nearby there is only Jack, and (b) that a person running in the room nearby produces noises similar to the ones he is hearing. In this cognitive state Ann is intuitively justified to believe that Jack is running, although no 'perceptual computation' gives 1 as value.

Case 4 puts into evidence another aspect of the problem. Suppose that when he wakes up John hears at the radio that a demonstration was held last evening, and that the police used fireplugs; since he wants to know whether the demonstration passed through his street, he looks through the window and sees puddles in the street, while he sees no puddles in the street nearby. In this case we would intuitively say that he has a justification for something like

(15) The demonstration passed through my street;

but if John had had a different question in mind it would have been correct to say that his seeing puddles in the street gave him a justification for something different, maybe for the belief that it rained during the night. How to account for this 'interest-relativity' of the notion of justification?[33]

I have anticipated that the computations determining the answers to the application and the justification questions are not independent; what I mean, more exactly, is that the answer to the former question is *presupposed* by the answer to the latter. The value of $f_C(o_1 \ldots o_n)$ in σ exclusively depends on which $\mathcal{C}$-concept f_C is and on which $\mathcal{C}$-objects $o_1,\ldots,o_n$ are, and in every atomic cognitive state the identity of such entities *is* known, as we have seen in Sects. 4.1 and 4.2. The value of $j_\sigma(\pi(\nu_1, \ldots, \nu_n))$ depends on the value of $f_C(o_1 \ldots o_n)$ in σ (under the assumption that $[\![\pi]\!]_\sigma = f_C$, $[\![\upsilon_1]\!]_\sigma = o_1, \ldots, [\![\upsilon_n]\!]_\sigma = o_n$)—this is what I mean when I say that the former presupposes the latter—but also on other things, as the cases 1–4 suggest and as I will argue in a moment.

[33] There are other, more complex, indirect cases. A particularly interesting class involves testimony. I will deal with cases of this kind in Sect. 4.8.1.

4.3.1 Justification and Explanation

The preceding examples show that, if we want that the answer to question (10) is the result of a computation, it is not sufficient that the answer to (9) is the result of a computation, and the definition of the denotation of predicates and of singular terms is the result of a computation: some essential component is lacking in the definition of a cognitive state.

It seems to me that a very natural answer to this problem emerges if we analyze the preceding examples as cases in which the subject is looking for an *explanation* of the data available to her/him.[34] Consider **Case 1**: in σ_2 Mary will probably try to acquire new relevant information, for example by touching the disk, or by changing again her position, or some other way; presumably, after having obtained a haptic representation of a round disk she will decide that the disk is round, thereby choosing the visual representation activated in σ_1 as 'better' than the one activated in σ_2 in the sense—I suggest—that, together with some general laws, it permits to account for the other visual representation as resulting from the form of the disk and the position of the observer. From an abstract point of view, we might say that Mary selects the representation which offers the best explanation of the data. An analogous analysis can be given of **Case 2**: the result of measurement with a ruler is new relevant information, and it is chosen as 'better' than the result of purely visual comparison because it yields a better explanation of the data. In the 'indirect' cases it is even more clear that the subject is looking for an explanation, and that in each case (s)he is intuitively justified to infer, and hence to believe, the proposition (s)he chooses as the best explanation of the data available in the cognitive state (s)he occupies. Notice that also in these cases a crucial role is played by relevant information: in **Case 3** it includes the pieces of information (a) and (b); in **Case 4** relevant information depends on John's interest: if he wants to know whether the demonstration passed through a certain street, the radio news is relevant; if he wants to know which shoes to put on, it is not. But a difference between direct and indirect cases should be stressed. In the former it might be plausibly suggested that relevant information is in some way limited by the meaning of the singular terms and predicates occurring in the sentence under consideration; for example, in **Case 1** it might be suggested that the representation of a round disk is relevant because it is similar to the denotation of "that disk". By contrast, this suggestion is meaningless in indirect cases: in **Case 3** it is not sufficient to make reference to the meaning of "run" in order to know whether the pieces of information (a) and (b) are relevant to a justification of (14).

In these sketched analyses I have made use of an intuitive notion of explanation, which leaves many things unclear; in particular, the appeal to 'relevant' information seems to play an important role, but how is relevant information to be specified? I think the best way to tackle the problem is to consider, on the one hand, a representative class of theories of explanation and, on the other, the constraint that such a theory permits to define the answer to question (9) as the result of a computation; and to

[34] For an implementation of this idea within the conceptual framework of Proof-Theoretic Semantics see Millson and Straßer (2019) and Stovall (2019).

choose the theory that best fits the constraint. To make an example, consider one of the best known theories of explanation, the Deductive-Nomological Theory. According to it a fact is explained when it is *deduced* from a set of premises including at least one law that is necessary for the deduction (Lipton, 2004: 26); since a valid deductive argument cannot be converted into an invalid one by the addition of premises, an explanation is not ruined by adding true premises. For example (Lipton, 2004: 51), a good explanation of the fact that Jones rather than Smith contracted paresis is that only Jones had syphilis; according to the Deductive-Nomological Theory this explanation is not ruined if we add the premise that only Smith was a regular churchgoer. On the one hand this shows the intuitive inadequacy of the theory, for if I say that Jones rather than Smith contracted paresis because only Jones had syphilis *and* only Smith was a regular churchgoer I have given an intuitively incorrect explanation (*Ibid.*); on the other hand, the consequence for the definition of justification I am looking is that there would be no way to circumscribe the class of relevant pieces of information.

Without entering into an analysis of various theories of explanation, I will describe in some detail the one that seems to me to best fit the constraint: van Fraassen's theory, as it is exposed in Chap. 5 of van Fraassen (1980); then I will suggest some modifications of it that seem to me necessary in order to define the notion of $\mathcal{C}$-justification. According to van Fraassen's theory, explanations are answers to why-questions, and why-questions have a *contrastive* nature, in the sense that their logical form is not simply "Why β?" but "Why β *in contrast to* X?", where X is a set of alternatives. From this point of view, in **Case 4**, "The demonstration passed through my street" and "It has been raining" can be seen as answers to two quite different questions—say "Why are there puddles in my street *in contrast to* there not being any in others?" and "Why are there puddles in my street *in contrast to* there not being any?", respectively. In this way, the subject's interest, which was intuitively seen as a disturbing subjective factor, is now transformed into an aspect of the objective situation; as a consequence, there *is* now some objective factor in terms of which a justification for one of the two statements can be differentiated from a justification for the other. However, the interest-dependence of justifications, i.e. of answers to why-questions, cannot be explained away exclusively by means of the contrastive interpretation of why-questions. For example (van Fraassen, 1980: 142), the question "Why does the blood circulate through the body?" can be answered in different ways—for instance "Because the heart pumps the blood through the arteries" or "To bring oxygen to every part of the body tissue"—independently of the contrasting class of alternatives, and depending on the kind of reason requested—a cause or a function, respectively. It seems natural to say that here a relation of *relevance* comes into play: in one case a causal reason is relevant, in the other a functional reason.

To sum up, a why-question expressed, in a given context C, by an interrogative sentence may be identified with a triple $Q = \langle\beta_k, X, Rlv\rangle$, where β_k is the *topic*, $X = \{\beta_1, \ldots, \beta_k, \ldots\}$ is the *contrast-class*, and Rlv is a *relevance relation* between sentences and pairs $\langle \beta_k, X\rangle$. A *presupposition* of a why-question Q is that (i) its topic is true, (ii) the other members of its contrast-class are not true, and that (iii) at least one of the propositions that are relevant to it is true; the conjunction of (i) and (ii) is called *the central presupposition* of the question. The context C includes a body of background knowledge K. A question Q *arises* in a given context C if K implies the central presupposition of Q and does not imply the falsity of (iii).[35] A *direct answer* to a why-question Q is expressed by a sentence of the form

(16) β_k in contrast to (the rest of) X (in context C) because α;

(16) is assumed to claim that β_k and α are true, that the other members of X are not true, and that α is a *reason*, i.e. that α bears relation Rlv to $<\beta_k, X>$.

4.3.2 Explanation and Computation

Van Fraassen's theory is sufficiently articulated and flexible to make explicit all the variables that are implicit in the intuitive relation of explanation, thereby making a computational treatment of explanation possible. However, some modifications seem to be necessary if we want to use it to characterize empirical justifications.

The first concerns the notion of context, which is fundamental in van Fraasen's approach but is left by him unanalyzed. It seems to me that it can be analyzed, at least partially, in terms of the notion of cognitive state introduced above. Van Fraassen conceives of a context of use in the usual way, i.e. as «an actual occasion, which happened at a definite time and place, and in which are identified the speaker […], addressee […], and so on» (van Fraassen, 1980: 135). An important aspect of the intuitive notion, as it results from the passage quoted at the end of the preceding section, is that both «a certain body K of accepted background theory and factual information» is available in a given context; but van Fraassen does not analyze K, apart from saying that «it depends on who the questioner and audience are». Well: if we focus on the contexts in which the speaker and the addressee are one and the same subject, it is not difficult to see how contexts can be defined in terms of cognitive states. Given a cognitive state σ, a context C can be defined in the following way: the subject is defined as the one whom the state σ belongs to (subjects include temporal sequences of cognitive states); the background theory is implicitly specified through the epistemic contents associated in σ to names and predicates; factual information is information encoded into, or associated to, the activated terms.

[35] Hence Q arises even if we do not know whether there is a direct answer (see below), provided (i) and (ii) are true. Van Fraassen stresses that a question's arising or not depends on background knowledge:

> In the context in which the question is posed, there is a certain body K of accepted background theory and factual information. This is a factor in the context, since it depends on who the questioner and audience are. It is this background which determines whether or not the question arises; hence a question may arise (or conversely, be rightly rejected) in one context and not in another. (Van Fraassen 1980: 145)

The second modification concerns the relation of relevance. Van Fraassen poses no explicit constraint on Rlv, simply suggesting that it is determined by the context[36]; on the contrary, for his theory to be useful in a computational account of justification it is essential that Rlv is strictly constrained. Consider for example the **Case 4** in Sect. 4.3.2. Among the data (factual information) available to John in the atomic cognitive state σ he occupies there is the piece of information that there are puddles in his street, and that the police used fireplugs; and John looks for a reason of the data. The crucial point is that, for John's search of a reason to be conceived as a computation, it must be conceivable as the execution of an algorithm; how could such an algorithm be characterized? The general theory of inference to the best explanation suggests that it is articulated into two phases: (a) the generation of a class of potential explanations, and (b) the choice of one as the best; van Fraassen's theory suggests how to generate the potential explanations: what John is doing is looking not simply for a reason of the data, but for the best answer to the why-question "Why are there puddles in my street in contrast to there not being puddles?"; he must therefore first generate a class of potential answers. Intuitively, the potential answers must be relevant to the why-question; on the other hand, for the class of potential answers to be generable, it must be circumscribable in some way. My suggestion is that it is circumscribed just by the existence of a relation Rlv of relevance between words. A plausible general assumption is that words are connected in semantic memory to form semantic fields or networks[37]; in the **Case 4** we may assume that the semantic network for "puddle" makes salient connections like the following[38]:

(17)

```
                          puddle
                            |
       ----------------------------------------
       |          |          |          |          |
   washing       rain       snow      flood     fireplug
                                                   |
                                            -------------
                                            |           |
                                         fireman      police
                                            |           |
                                          fire    demonstration
```

[36] Kitcher & Salmon (1987: 319) argue that «the lack of any constraints on "relevance" relations allows just about anything to count as the answer to just about any question.» However this objection is evaluated, it should be stressed that it is not addressed to the fact that, in van Fraassen's account of explanation, relevance is taken as a primitive relation. This, in my opinion, is a merit of van Fraassen's approach, since, instead of requiring a definition of relevance in terms of other notions (as for example in Sperber and Wilson's Relevance Theory (Sperber & Wilson, 1995)), it permits a definition of fundamental notions of the theory of explanation, in particular of the notion of reason.

[37] Evidence for this assumption includes for instance the phenomenon of semantic priming. Cp. McNamara (2005).

[38] I am in no way suggesting that the following is the actual structure of the semantic network "puddle" belongs to.

At this point it is natural to equate the relation Rlv with the relation between the nodes of the semantic field, and to require that the potential answers to the why-question arising in σ are sentences built up with predicates belonging to the semantic field of the predicate of the why-question. As a consequence, the following facts severely circumscribe the range of data to be scanned, thereby permitting the generation of the class of potential answers:

(a) the presence of "fireplug" within the semantic field of "puddle";
(b) the presence of "puddles" in the why-question arising in σ;
(c) the fact that, in σ, $sc_{\text{the demonstration}}$ contains the pieces of information

- that last evening a demonstration was held, and
- that the police used fireplugs.

The class of potential answers will therefore include: "It rained last night", "The demonstration passed through the street", "Last night my street has been washed", ecc. Presumably John will not be capable, in σ, to choose one answer as the best; but the fact that his choice is limited to those alternatives will orient his search for more data. For example, he will try to acquire information about the presence of puddles in other streets, about the days of street washing, and so on.

I will therefore assume a relation Rlv of relevance among words, and consequently among $\mathcal{C}$-concepts, as a further component of the computational apparatus of a subject. For a semantic theory it is not necessary to investigate either the nature or the actual extension of this relation; what is important is a consequence of its existence: that the system of $\mathcal{C}$-concepts manageable in a given cognitive state is not 'isotropic' in the sense that, given a cognitive state σ, it is not true, in general, that starting from one $\mathcal{C}$-concept every other $\mathcal{C}$-concept can be 'reached'. I will come back to this point in the concluding remarks at the end of this chapter.

4.3.3 Answering the Justification Questions

Since all the notions of van Fraassen's theory of explanation can be defined in terms of the notion of cognitive state, as we have seen, a computational answer to the Justification Question for atomic sentences becomes possible.

The intuitive idea is that it is necessary to distinguish two kinds of cases: 'direct' ones, in which the answer to the Justification Question is the same as the answer to the corresponding Application Question, and 'indirect' ones, in which inference to the best explanation comes into play. More precisely, a 'direct' case is a cognitive state σ in which the subject, having computationally determined that $[\![\pi]\!]_\sigma = f_C$, $[\![v_1]\!]_\sigma = o_1, \ldots, [\![v_n]\!]_\sigma = o_n$, has sufficient information to give one answer to the Application Questions for $f_C(o_1 \ldots o_n)$ in σ. In such a case, we can say that the *evidential factors* of the answer to the Justification Question, i.e. the data and algorithms the answer depends on, are the same as the evidential factors of the answer to the corresponding Application Question; since the latter factors are the function f_{app} of application of a concept to its objects, the concept f_C and the objects $o_1 \dots o_n$, their complex can be conceived as a single evidential factor and denoted by "$f_C(o_1 \dots o_n)$" (f_{app} is denoted by simple syntactic concatenation). When $f_C(o_1 \dots o_n) = 1$ in σ, we can say that $f_C(o_1 \dots o_n)$ *makes evident* $\pi(\nu_1, \dots, \nu_n)$ in σ.[39]

The 'indirect' cases are cognitive states in which the subject, even having computationally determined that $[\![\pi]\!]_\sigma = f_C, [\![\upsilon_1]\!]_\sigma = o_1, \dots, [\![\upsilon_n]\!]_\sigma = o_n$, has not sufficient information to give one answer to the Application Question for $f_C(o_1 \dots o_n)$ in σ; significant examples are Cases 1–4 described above, in which the value of $f_C(o_1 \dots o_n)$ is not defined because information attached to the situational epistemic content of the predicate and of the singular terms is not sufficient either to give an answer or to choose between two conflicting answers. In such cases the evidential factors of the answer are not determined by the syntactic components of $\pi(\nu_1, \dots, \nu_n)$, nevertheless they are computable, when they exist.[40] Consider, for example, **Case 4:** intuitively, the evidential factor is the piece of information that there are puddles in my street; from the point of view of the theory of explanation illustrated above, the evidential factors are all the pairs < Q, A > such that Q is the why-question arising in σ ("Why are there puddles in my street *in contrast to* there not being any in others?") and A a potential answer to Q; for one potential answer (for instance (15)) to be made evident it is necessary that it is chosen by the subject as the best one. When the best answer is expressed by the very sentence $\pi(\nu_1, \dots, \nu_n)$ we are defining a justification for, we seem to fall into a circle, because we are defining the meaning of $\pi(\nu_1, \dots, \nu_n)$ and at the same time we use $\pi(\nu_1, \dots, \nu_n)$ to state the answer; but it is an appearance: when we use $\pi(\nu_1, \dots, \nu_n)$ to state the answer, the meaning we give it is determined by the fact that $\pi, \nu_1, \dots, \nu_n$ have their respective denotations, and these may be assumed to have been computed without circularity; more concisely, the meaning of $\pi(\nu_1, \dots, \nu_n)$ is that $f_C(o_1 \dots o_n) = 1$.

There is a considerable difference between 'direct' and 'indirect' cases with respect to the computation of the answer to the Justification Question. In the former it's about applying a $\mathcal{C}$-concept to a $\mathcal{C}$-object, and we have seen that a $\mathcal{C}$-concept can be seen as a routine, whose application can be assumed to be an unconscious process; in the latter processes of other kinds take place, e.g. selection of one answer among several possible ones, or *creation* of new routines or concepts (as opposed to the managing of old ones), and these processes may require the intervention of conscious control. By the way, it is possible that just the intervention of these processes makes the difference between human and other animals' reasoning, between human and other animals' 'logical competence'. In any case, the fact that these processes are conscious does not preclude them from being computations.

These considerations are intended to justify the following definitions:

---

[39] A cognitive state in which $f_C(o_1 \dots o_n) = 0$ is not to be equated to a case in which $f_C(o_1 \dots o_n)$ simply does not make evident $\pi(\nu_1, \dots, \nu_n)$. I shall come back to this point in Sect. 4.4.1 of this chapter and in Chap. 5, Sect. 5.2.1.

[40] In 'indirect' cognitive states it is perfectly possible that an atomic sentence has no evidential factor.

Definition 6 Let σ be an arbitrary atomic cognitive state such that $[\![\pi]\!]_\sigma = f_C$, $[\![\nu_1]\!]_\sigma = o_1, \ldots, [\![\nu_n]\!]_\sigma = o_n$; then

(i) the set $EF_{\pi(\nu 1,\ldots,\nu n),\sigma}$ of the *evidential factors* of $\pi(\nu_1, \ldots, \nu_n)$ in σ is $\{f_c(o_1 \ldots o_n)\} \cup \mathcal{R} = \{<Q,\pi(\nu_1, \ldots, \nu_n)> | Q$ is a why-question arising in σ and $f_c(o_1 \ldots o_n) =_\sigma 1$ is a potential answer to Q};
(ii) the set $EF_{\nu 1=\nu 2,\sigma}$ of the *evidential factors* of $\nu_1 = \nu_2$ in σ is $\{f_=(o_1 o_2)\} \cup \mathcal{R} = \{<Q, \nu_1 = \nu_2> | Q$ is a why-question arising in σ and $f_=(o_1 o_2) =_\sigma 1$ is a potential answer to Q};
(iii) the set $EF_{\pi 1\equiv\pi 2,\sigma}$ of the *evidential factors* of $\pi_1 \equiv \pi_2$ in σ is $\{f_\equiv(C_1 C_2)\} \cup \mathcal{R} = \{<Q,\pi_1 = \pi_2> | Q$ is a why-question arising in σ and $f_\equiv(C_1 C_2) =_\sigma 1$ is a potential answer to Q}.

Definition 7

(i) Let $a \in EF_{\pi(\nu 1,\ldots,\nu k),\sigma}$; then a makes evident $\pi(\nu_1, \ldots, \nu_n)$ in σ (in symbols $a \vDash_\sigma \pi(\nu_1, \ldots, \nu_n)$) iff

- either $a = \{f_c(o_1 \ldots o_n)\}$ and $f_c(o_1 \ldots o_n) =_\sigma 1$;
- or $a \in \mathcal{R}$ and there is a why-question Q arising in σ such that $f_c(o_1 \ldots o_n) =_\sigma 1$ is the best answer to Q in σ.

(ii) Let $a \in EF_{\nu 1=\nu 2,\sigma}$; then $a \vDash_\sigma \nu_1 = \nu_2$ iff

- either $a = \{f_=(o_1 o_2)\}$ and $f_=(o_1 o_2) =_\sigma 1$;
- or $a \in \mathcal{R}$ and there is a why-question Q arising in σ such that $f_=(o_1 o_2) =_\sigma 1$ is the best answer to Q in σ.

(iii) Let $a \in EF_{\pi 1\equiv\pi 2,\sigma}$; then $a \vDash_\sigma \pi_1 \equiv \pi_2$ iff

- either $a = \{f_\equiv(C_1 C_2)\}$ and $f_\equiv(C_1 C_2) =_\sigma 1$;
- or $a \in \mathcal{R}$ and there is a why-question Q arising in σ such that $f_\equiv(C_1 C_2) =_\sigma 1$ is the best answer to Q in σ.

Definition 8 Let α be an atomic sentence; then $j_\sigma(\alpha) = 1$ iff there is an $a \in EF_{\alpha,\sigma}$ such that $a \vDash_\sigma \alpha$.

I have said above[41] that the value of $j_\sigma(\pi(\nu_1, \ldots, \nu_n))$ depends not only on the value of $f_C(o_1 \ldots o_n)$ in σ but also on other things; now we know on what: on the data expressed by the topic of the why-question Q arising in σ; we can therefore say that the value of $j_\sigma(\pi(\nu_1, \ldots, \nu_n))$ depends on the semantic values of $\pi, \nu_1, \ldots, \nu_k$ *and on σ*. The same remark applies to identities. This may sound as a departure from the Compositionality Principle as it is usually understood:

(Comp) For every complex expression e in $\mathcal{L}$, the semantic value of e in $\mathcal{L}$ exclusively depends on the structure of e and on the semantic values of the constituents of e.

[41] Immediately before Sect. 4.3.1.

But in fact the semantic value of e standardly depends, in an externalist semantics, also on how the real world actually is. For example, the semantic value of "Ann smokes" depends on the semantic values of "Ann" and "smokes" (i.e. a and S = {x|x smokes}, respectively), but also on the fact that a ∈ S in the real world. More generally, in an externalist possible worlds semantics the semantic value of "Ann smokes" in w depends on the semantic values of "Ann" and "smokes" in w and on whether a ∈ S in w, hence *on w*. It might be retorted, on behalf of (Comp), that the fact that a ∈ S, or that a ∉ S, in w is contained in the definition of the semantic value of S in w. I need not take position on this issue[42]; if the last convention were chosen, my remark on the compositional nature of my strategy would be rephrased by dropping "on σ".

Let us see how definition 8 works by applying it to the cases 1–4.

Case 1. The question arising in σ_2 is something like "Why do I see that disk as round in p_1 and elliptical in p_2 *in contrast to* looking it as round or elliptical in both positions?", and the best answer is something like "Because I look in p_1 frontally, and in p_2 sideways, at it," and this, together with some other premise (something like "A frontal view is more reliable than a lateral one"[43]), entails that the disk is round; in the cognitive state σ_3 in which the best answer has been computed $f_{\mathrm{ROUND}}(\mathrm{d}) =_{\sigma 3} 1$; hence, by Definition 8, $j_{\sigma 3}$(is round(d)) = 1.

Case 2. The question arising in σ_2 (after the measurement) is "Why do I see d_1 as shorter than d_2 *in contrast to* seeing them the same lenghth?", and the best answer is something like "Because the presence of angle brackets at the extremities of d_1 and of inverted angle brackets at the extremities of d_2 produces a distorting effect"; in the cognitive state σ_3 this answer has been chosen, the value $f_{\mathrm{SAMELENGTH}}(d_1, d_2)$ is (no longer 0, but) undefined. The best answer, together with some other premise (something like "a measure with a ruler is more reliable than a visual comparison, in presence of a distorting context") entails that d_1 and d_2 are the same length, i.e. that $f_{\mathrm{SAMELENGTH}}(d_1, d_2) =_{\sigma 3} 1$, hence, by Definition 8, $j_{\sigma 3}$(are the same length(d_1, d_2)) = 1.

Case 3. The cognitive state σ occupied by Ann is characterized by the following facts: (a) the activated term is a representation of such and such noises; (b) sc_{Jack}

[42] However, let me remark that the most convincing account I have found of the difference between knowing the truth-*conditions*, hence the meaning, of π(ν) and knowing its truth-*value* in Frege's semantics is the one, already quoted in Chap. 1, given by Heim and Kratzer (1998) (and inspired by Dummett 1981: 227), according to which we derive the truth-conditions of π(ν), as distinct from its truth-value, only if the denotation of π is defined by specifying a *condition* (as when we say that [smokes] = {x|x smokes}) instead of displaying a *list* (as when we say for instance that [smokes] = {Ann, Jan, Mary}). Under this account, the fact that a∈S, or that a ∉ S, in w is *not* contained in the definition of the semantic value of S in w; in fact, displaying a list

> would have required more world knowledge than we happen to have. We do not know of every existing individual whether or not (s)he smokes. And that's certainly not what we have to know in order to know the meaning of "smoke". (Heim & Kratzer 1998: 21)

[43] I am not saying that this premise is explicitly formulated by Mary.

contains the piece of information that Jack, and no other, is in the the room nearby. The question Q arising in σ is "Why are there such noises in the room nearby *in contrast to* there not being noises?". Since "noises" bears the relation Rlv to "run", "snore", etc., the following sentences belong to the set PA of potential answers: "Jack is snoring in the the room nearby", "Jack is running in the the room nearby", "John is running in the the room nearby", etc.; from this and (b) it follows that (14) is the best answer to Q in σ, i.e. that $f_{\mathrm{ISRUNNING}}(\mathrm{jack}) =_{\sigma} 1$; hence, by Definition 8, j_{σ}(is running(jack)) = 1.

Case 4 has already been analyzed, but it may be useful to reframe the analysis. The cognitive state σ occupied by John is characterized by the following facts: (a) the activated terms are a representation of puddles in my street and a representation of the absence of puddles in the street nearby; (b) $\mathrm{sc}_{\mathrm{demonstration}}$ contains the piece of information that the evening before a demonstration was held; (c) $\mathrm{sc}_{\mathrm{fireplug}}$ contains the piece of information that the police used fireplugs. Since John wants to know whether the demonstration passed through his street, the question Q arising in σ is expressed by "Why are there puddles in this street *in contrast to* there not being any in the other?" Since "puddle" bears the relation Rlv to other words, as shown in (17), the following sentences belong to the set PA of potential answers: "The demonstration passed through my street", "It rained last night", "My street has been washed", etc.; the second may be discarded because of the second activated term; information in (b) and (c) make the first better than the third, and this entails that $f_{\mathrm{PASSEDTHROUGHMYSTREET}}(\text{the demonstration}) =_{\sigma} 1$; hence, by Definition 8, j_{σ}(passed through my street (the demonstration)) = 1.[44]

I conclude that the answer to the Justification Question for an atomic sentence α in σ can be seen as determined by a computation if we conceive that computation as an inference to the best explanation of the data available in σ, provided we do not assume that the Justification Question for α has an answer in every cognitive state.

4.4 Subjects and $\mathcal{C}$-Justifications for Atomic Sentences

The analyses and discussions of Sects. 4.1–4.3 have given rise to hypotheses and assumptions about operations permitting constitution and recognition of $\mathcal{C}$-objects, constitution and applicaion of $\mathcal{C}$-concepts. It is useful to distinguish the operations that are constitutive of the notion of cognitive state from those that can be ascribed to the structure of the computational apparatus of the knowing subject, and that may be assumed as invariant across the multiplicity of subjects. The following brief summary operates this distinction and introduces the definitions I will refer to later.

[44] Of course this is only the rough sketch of an analysis.

4.4.1 Internalist Ontology (General Assumptions)

1. Individual mind is articulated into several components, including the senses (vision, audition, etc.), memory, imagination, the faculty of language, and the conceptual-intentional systems (**CIS**).
2. It is assumed, for the sake of simplicity, that there is a unique **CIS**, with its **IRS**, permitting the derivation of representations, i.e. (structured) terms,[45] according to algorithms/programs; and that the representations coming from the components having access to **CIS** are 'translated' into **IRS**.
3. The senses, memory, imagination and the faculty of language have access to **CIS**.
4. The terms coming from the components having access to **CIS** are organized by **CIS** in equivalence classes ($\mathcal{C}$-objects) and functions from $\mathcal{C}$-objects to {0,1} ($\mathcal{C}$-concepts), by the computational apparatus of (knowing) *subjects* (see below).
5. $\mathcal{C}$-objects and $\mathcal{C}$-concepts belong to (prelinguistic and linguistic) *atomic cognitive states*.

Definition 9 An *atomic cognitive state* is a quadruple $\sigma = < i, \mathcal{A}_\sigma, \mathcal{D}_\sigma, \mathcal{C}_\sigma >$ such that

(i) i is a time;
(ii) $\mathcal{A}_\sigma$ is a non-empty set of terms of **IRS** *activated* at i;
(iii) $\mathcal{D}_\sigma$ is a non-empty set of $\mathcal{C}$-objects (the cognitive domain of σ);
(iv) $\mathcal{C}_\sigma$ is $\cup_{n>0}\mathcal{C}^n$, where each $\mathcal{C}^n$ is a non-empty set of n-ary $\mathcal{C}$-concepts (the manageable concepts of σ).

6. In passing from an atomic cognitive state σ to another σ', a $\mathcal{C}$-object may acquire or lose some terms; or an old $\mathcal{C}$-object may disappear/be suppressed; or a new $\mathcal{C}$-object may appear/be created; or one $\mathcal{C}$-object may split into two; or two $\mathcal{C}$-objects may merge into one. The global result of these operations can be seen as an *updating* of $\mathcal{D}_\sigma$ to $\mathcal{D}_{\sigma'}$, consequent to the acquisition of new information.
7. When the faculty of language accesses **CIS**, semantical interpretation of SEMS takes place, including the process of assigning a meaning to singular terms and predicates; the process is effected, again, by *subjects*.

Definition 10 A (knowing) *subject* is an agent whose cognitive structure includes a triple $\mathcal{S} = < \Sigma, \mathcal{CA}, [[]] >$ such that

1. Σ is a set of atomic cognitive states $\sigma_1, \sigma_2, \ldots$ indexed by natural numbers (thought of as times).
2. $\mathcal{CA}$ (the Computational Apparatus) accesses the catalogue CAT of object descriptions and the relation Rlv of relevance[46] stored in memory, and computes the following functions:

[45] Examples of representations derived in the visual representational system are Marr's 3-D model descriptions; examples of representations derived in the linguistic representational system are logical forms. "Representation" is not used here in a relational sense: a representation is not a representation *of* an object in the real world, but simply a structured symbol or term.

[46] Rlv is a relation between words, hence indirectly between their denotations.

2.1. *ec*, a function such that

2.1.1. if $\nu \in \mathcal{N}_o$, ec_ν is a function such that for each $\sigma \in \Sigma$, $ec_\nu(\sigma)$ is a set of features (the *epistemic content* associated to ν)[47]; $ec_\nu(\sigma)$ is assumed to be articulated into two parts: the *lexical epistemic content* $lec_\nu(\sigma)$, constituted by information coming from the lexicon, and the *situational epistemic content* $sec_\nu(\sigma)$, constituted by information coming from perception, memory, imagery, etc.

2.1.2. if $\pi \in \mathcal{P}^n$, ec_π is a function such that for each $\sigma \in \Sigma$, $ec_\pi(\sigma)$ is a set of features (the *epistemic content* associated to π).[48] $ec_\pi(\sigma)$ is assumed to be articulated in a way analogous to $ec_\nu(\sigma)$.

2.2. $<p_{C1}, p_{C2}, \ldots>$, a collection of algorithms computing the $\mathcal{C}$-concepts f_{C1}, f_{C2}, …[49]

2.3. The function *const* defined in (4)(i).

2.4. The function *id* defined in (4)(ii).

2.5. The function *inf* defined in (4)(iii).

2.6. The function *match* defined in (4)(iv).

2.7. The function *MATCH* defined in (6)(i).

2.8. The function *den* defined in (6)(ii).

2.9. A function *ba* such that, if σ is an atomic cognitive state, Q a question arising in σ and PA the set of potential answers to Q in σ, $ba(<Q, PA>, \sigma)=\alpha$ iff α is the best answer in σ to Q among the ones contained in PA.[50]

---

[47] It is assumed that such information is computationally tractable, as it happens in the case of information encoded into lexical entries through (systems of arrays of) *features*; but features are not the only computationally tractable way of codifying information: Marr'sc 3-D model descriptions are another example.

Information contained in $ec_\nu(\sigma)$ may be either verbal or non-verbal; it is subject to the following restrictions: (i) verbal information must be atomic, in the sense that it must be expressible by means of atomic sentences; (ii) the atomic sentences expressing verbal information associated to ν must contain occurrences of ν; (iii) non-verbal information associated to ν must be explicitly associated to ν. The rationale for these restrictions is the following. For (i): in order to explain the notion of justification for (and therefore the meaning of) an atomic sentence I appeal to the notion of authorization to use a name; if the explanation of this notion made reference to logically complex sentences, the whole approach could not satisfy a molecularity requirement, since the explanation of the meaning of an atomic sentence would presuppose the understanding of the meaning of sentences of unlimited logical complexity. For (ii) and (iii): The reason for these requirements is the necessity of assuring epistemic transparency to the formal notion of authorization: any subject who associates to ν the epistemic content $ec_\nu(\sigma)$ must know that $ec_\nu(\sigma)$ is associated to ν.

[48] With restrictions analogous to the ones for $ec_\nu(\sigma)$.

[49] The collection of primitive concepts is assumed to include the identity concept f_{id}; an algorithm for its computation is suggested by the function *match* (see Sect. 2.6).

[50] I am assuming without discussion the existence of some criteria according to which an answer to a why-question is selected as the best one among several possible ones. For an introduction and discussion see Lipton (2004) and van Fraassen (1980). I am not assuming that *ba* is total; Case 4 is an example of an atomic cognitive state in which its value is not determined. In such cases the subject normally searches for new information; also this search can be seen as ruled by computational procedures.

3. $[[\ ]]_{\mathcal{S}}$,[51] the *meaning-assignment*, is a function such that
    3.1. for every $\nu \in \mathcal{N}_{\text{o}}$, $[[\nu]]_{\mathcal{S}}$ is a function such that, for every $\sigma \in \Sigma$, $[[\nu]]_{\mathcal{S},\sigma} = den(\nu,\mathcal{D}_{\sigma})$.[52]
    3.2. for every $\pi \in \mathcal{P}^{\text{n}}$, $[[\pi]]_{\mathcal{S}}$ is a function such that, for every $\sigma \in \Sigma$, $[[\pi]]_{\mathcal{S},\sigma}$ = the $\mathcal{C}$-concept f_{C} iff $ec_{\pi,\sigma}$ defines f_{C}.[53]
    3.3. $[[=]]_{\mathcal{S}}$ is a function such that, for every $\sigma \in \Sigma$, $[[\pi]]_{\mathcal{S},\sigma}$ = the $\mathcal{C}$-concept of objectual identity defined in (7).
    3.4. $[[\equiv]]_{\mathcal{S}}$ is a function such that, for every $\sigma \in \Sigma$, $[[\pi]]_{\mathcal{S},\sigma}$ = the $\mathcal{C}$-concept of conceptual identity defined in (8).

Definition 11 An $\mathcal{S}$*-justification for an atomic sentence* α is an atomic cognitive state σ of the subject $\mathcal{S}$ such that $\text{j}(\alpha) = 1$.[54]

An important consequence of this definition is that the notion "σ is an $\mathcal{S}$-justification for the atomic sentence α" is semidecidable in the following sense: (i) if a subject $\mathcal{S}$ has a justification for α, i.e. $\mathcal{S}$ occupies an atomic cognitive state σ and σ is an $\mathcal{S}$-justification for α, then $\mathcal{S}$ knows that $\mathcal{S}$ has an $\mathcal{S}$-justification for α; (ii) if $\mathcal{S}$ does not have a justification for α, i.e. $\mathcal{S}$ occupies a cognitive state σ and σ is not an $\mathcal{S}$-justification for α, then in general $\mathcal{S}$ does not know that $\mathcal{S}$ does not have a justification for α. In other terms, possession by $\mathcal{S}$ of an $\mathcal{S}$-justification for α is epistemically transparent for $\mathcal{S}$. Here "knows" is to be understood approximately in the sense of tacit or implicit knowledge—the sense Chomsky assigns to the technical term "cognize": if a subject $\mathcal{S}$ who cognizes that (s)he has an $\mathcal{S}$-justification for α could become conscious of what (s)he cognizes, we would not hesitate to say that (s)he knows that (s)he has an $\mathcal{S}$-justification for α.[55] What the subject experiences, in case (i), is that α is *evident* for her/him. Suppose for example that α is "That is a cat": if John understands α and someone asks him: "Is that a cat?", he will answer "Yes"; if for some reason he does not understand α (maybe because he doesn't know English, or because he is an infant, or because he is a dog, etc.), the best explanation of his behaviour will still be that it is evident for her/him that that is a cat.

In case (ii) $\mathcal{S}$ will answer neither "Yes" nor "No",[56] and the best explanation of her/his behaviour will be neither that it is evident for her/him that that is a cat, nor that it is evident for her/him that that is not a cat. This case should be clearly distinguished from the one in which $\mathcal{S}$ is in a position to answer "No". As we will see in the next chapter, such a cognitive state characterizes the intuitive situation

[51] The subscript "$\mathcal{S}$" will occasionally be omitted.

[52] $[[\nu]]_{\mathcal{S},\sigma}$ will be called the $\mathcal{S}$*-denotation* (or occasionally the $\mathcal{S}$*-extension*) of ν in σ, $[[\nu]]_{\mathcal{S}}$ the $\mathcal{S}$*-intension* of ν.

[53] $[[\pi]]_{\mathcal{S},\sigma}$ will be called the $\mathcal{S}$*-extension* (or occasionally the $\mathcal{S}$*-denotation*) of π in σ, $[[\pi]]_{\mathcal{S}}$ the $\mathcal{S}$*-intension* of ν.

[54] An $\mathcal{S}-$justification for α is a $\mathcal{C}$-justification for α belonging to the cognitive states of the subject $\mathcal{S}$. I will occasionally follow this use.

[55] See Chomsky (1980): 69–70.

[56] This happens in general because in the given atomic cognitive state available data are not enough to select one potential answer as the best one (see for instance Case 4 above in the text).

in which $\mathcal{S}$ has a justification for ~α, where "~" is a sort of empirical negation; of course, in such a situation $\mathcal{S}$ does not have a justification for α, as in case (ii), but the difference from (ii) is that $\mathcal{S}$ *knows* that (s)he has a justification for ~ α, hence (s)he also *knows* that (s)he does not have a justification for α. Hence, many cases intuitively described as cases in which $\mathcal{S}$ simply has not a justification for α, are in fact epistemically transparent, since $\mathcal{S}$ knows that (s)he does not have a justification for α.

Summing up, it is legitimate to assert that the function *j* of *justificational evaluation* computing the answer to the Justification Question in σ for each atomic sentence (i.e. such that, for any atomic cognitive state σ and any atomic sentence α, $j_\sigma(\alpha) \in \{0, 1\}$) is a (partial) calculable function.

4.4.2 $\mathcal{C}$-Truth-Grounds for Atomic Sentences

In Chap. 3 I imposed onto truth-grounds of α the conditions of being k-factive justifications for α, and I observed that possession of a k-factive justification for α is sufficient to warrant the assertibility of α, since it is just knowledge that warrants assertibility. Now it is time to define the notion of truth-ground in such a way as to guarantee its k-factiveness.

The starting question is therefore: under which conditions can we assert that a subject $\mathcal{S}$ knows that α? It is an old question; Plato gave a classical definition of knowledge,[57] according to which

(18) $\mathcal{S}$ knows that $\alpha =_{\text{def}}$

(a) $\mathcal{S}$ believes that α;
(b) $\mathcal{S}$ is justified in believing that α;
(c) $< \alpha >$ is true.

Edmund Gettier (Gettier, 1963) objected, by means of two famous counterexamples, that conditions (a)–(c) are not sufficient to guarantee knowledge. A definition of the notion of truth-ground must therefore contain an (implicit) answer to Gettier problems. I shall analyze Gettier problems in Chap. 8, but here I must anticipate a sketch of my answer; the central ideas are:

(SG)

(i) that the notion of truth-ground is *constitutive* of the notion of truth: a sentence we have a truth-ground of is true because we have a truth-ground of it, not viceversa; I have argued for this idea in Chap. 3.

(ii) that a subject $\mathcal{S}$ who occupies a cognitive state that is a truth-ground of α occupies an *ideal* cognitive state, i.e. a state in which $\mathcal{S}$ has, about α, all relevant information.

[57] See Plato, *Meno* 97a–98b; *Theatetus* 200d–210c.

(iii) that the intuitive reason that justifies a belief in a cognitive state is the same as the reason that justifies that belief in an ideal cognitive state. This idea emerges from the analysis of a number of Gettier cases; I shall explain and justify it in Chap. 8.

Let us concentrate on condition (ii). There is a tension between the property of a cognitive state of being ideal and its epistemic transparency, for two reasons: first, the totality of what, in a given situation, is relevant to know in order to establish whether things are so and so seems not to be a priori circumscribable; second, a cognitive state is ideal if, and only if, it is *optimal*, in the sense that *there is not* a better cognitive state (in some sense of "better"), and optimality may be transcendent, because in order to establish it it may be necessary to inspect an infinite set of cognitive states.

Consider, for example, the following case. Yesterday Bill, the notary public, has received John's marriage certificate; this, together with many other pieces of information, gave him a justification to believe that John is married. But today Bill comes to know that John has an interest at making him believe that he is married (say because of an inheritance), and that in the past he was condemned for cheat; at this point he is no longer sure that the justification he had yesterday is a truth-ground of the statement "John is married". He goes to the Town Hall to control whether the certificate is authentic; the answer is "Yes": so the original justification seems again to him a truth-ground. But later he is informed that John's wife died some months ago: the original justification is no longer a truth-ground. And so on and so forth. Since there is no natural limit to the number of possible oscillations, the only way of eliminating in principle the possibility of an infinite number of oscillations is to allow that the subject has access from the outset to an infinite number of relevant pieces of information. Notice that what is required is not that the subject knows the answer to each test to which the justification might be submitted; this would not yet be sufficient to guarantee the possibility for the subject to decide whether the justification is a truth-ground, because it might happen that infinitely many tests were necessary, and that positive and negative answers to the tests alternated regularly. What is required is something stronger: that the test is one, and that, in order to answer the question, the subject knows all relevant information at once. Of course, this poses an unsurmontable difficulty to whoever wants to determine whether a given cognitive state is optimal: if, in order to determine optimality, an infinite number of pieces of information is necessary, then it cannot be ruled out that a cognitive state is optimal and that no subject is in a position to know that it is.

The solution I propose consists in defining truth-grounds of $\pi(\nu_1, \ldots, \nu_n)$ not in terms of a property of a cognitive state, like its being ideal, but in terms of a binary *relation* between two cognitive states: the relation existing between σ and σ' if both are justifications for $\pi(\nu_1, \ldots, \nu_n)$ and σ' is 'better' than σ, in a sense to be defined, with respect to $\pi(\nu_1, \ldots, \nu_n)$.[58] In this way any reference to the global notion of optimal cognitive state can be avoided, and replaced by the local relation of betterness between states: if the relation of betterness can be defined in such a way that it turns out to be transparent, transparency of truth-grounds is ensured. Here is how I propose to define the betterness relation and the relation "σ' assigns to σ

[58] *Mutatis mutandis*, the proposal may be extended to identities.

the status of $\mathcal{S}$-truth-ground of $\pi(\nu_1, \ldots, \nu_n)$", (or equivalently the property "σ is a $\mathcal{S}$-truth-ground of $\pi(\nu_1, \ldots, \nu_n)$ relative to σ′")[59] (in symbols $\vDash^{\mathcal{S}}_{\sigma,\sigma'} \pi(\nu_1, \ldots, \nu_n)$):

Definition 12 Let $\mathcal{S} = \, < \Sigma, \mathcal{CA}, [\![\ ]\!] >$ be a subject, and $\sigma =< \mathcal{A}_\sigma{}'\mathcal{D}_\sigma{}'\mathcal{C}_\sigma >$, $\sigma' =< \mathcal{A}_\sigma, \mathcal{D}_\sigma{}', \mathcal{C}_\sigma{}' >$ two atomic cognitive states of $\mathcal{S}$; then, for every atomic sentence $\pi(\nu_1, \ldots, \nu_n)$, σ′ is *better than* σ *with respect to* $\pi(\nu_1, \ldots, \nu_n)$ (in symbols $\sigma' \geq_{\pi(\nu 1,\ldots,\nu n)} \sigma$) if, and only if, conditions (a)–(d) are satisfied:

(a) for every at ∈ $\mathcal{A}_\sigma$, $inf_{\mathrm{at}}(\sigma)$ is part of $inf_{\mathrm{at}}(\sigma')$;
(b) for every ν_i ($1 \leq i \leq n$), either
$ec_{\nu i}(\sigma)$ is part of $ec_{\nu i}(\sigma')$, or
$ec_{\nu i}(\sigma)$ is not part of $ec_{\nu i}(\sigma')$, and there is a why-question Q arising in σ about the data contained in $inf_{\mathrm{at}}(\sigma) \cup inf_{\mathrm{at}}(\sigma')$ such that the association to ν_i of $ec_{\nu i}(\sigma')$ yields a better answer to Q than the association of $ec_{\nu i}(\sigma)$;
(c) either
$ec_\pi(\sigma)$ is part of $ec_\pi(\sigma')$, or
$ec_\pi(\sigma)$ is not part of $ec_\pi(\sigma')$, and there is a why-question Q arising in σ about the data available in σ′ such that the association to π of $ec_\pi(\sigma')$ yields a better answer to Q than the association of $ec_\pi(\sigma)$.

It seems plausible to postulate that the relation $\sigma' \geq_{\pi(\nu 1,\ldots,\nu n)} \sigma)$ is reflexive and transitive, since the relations 'x is part of y' and 'x is a better answer that y' are.

Definition 13 Given a subject $\mathcal{S} = < \Sigma, \mathcal{CA}, [\![\ ]\!] >$, a cognitive state $\sigma \in \Sigma$, and a cognitive state σ′, $\vDash^{\mathcal{S}}_{\sigma,\sigma'} \pi(\nu_1, \ldots, \nu_n)$ if, and only if,

(i) $\sigma' \geq_{\pi(\nu 1,\ldots,\nu n)} \sigma$;
(ii) $j_\sigma(\pi(\nu_1, \ldots, \nu_n)) = 1$ and $j_{\sigma'}(\pi(\nu_1, \ldots, \nu_n)) = 1$;
(iii) if $a \vDash_\sigma \pi(\nu_1, \ldots, \nu_n)$ and $a' \vDash_{\sigma'} \pi(\nu_1, \ldots, \nu_n)$, then $a = a'$.

Some comments about this definition are in order.

Clause (iii) of Definition 13 requires that the evidential factor that makes $\pi(\nu_1, \ldots, \nu_n)$ evident in σ and σ′ is the same; it corresponds to condition (SG)(iii), once intuitive 'reasons' are identified with what I have called above evidential factors. In everyday language, when we want to explain a phenomenon, we use "reason" in different senses. For example, John may ask himself: "Why are there puddles in

[59] The predicate "σ is a $\mathcal{S}$-truth-ground of $\pi(\nu_1, \ldots, \nu_n)$ relative to σ' " expresses a property of a cognitive state σ, which is relative to a cognitive state σ'; the predicate "σ' assigns to σ the status of $\mathcal{S}$-truth-ground of $\pi(\nu_1, \ldots, \nu_n)$" expresses a non-relative relation. An analogous distinction between relative properties and (non-relative) relations is introduced by D. Lewis (Lewis 1979: 539) with the following words:

> There is a sense in which "mountain I live on top of" expresses a property relative to any given subject; but also there is a sense in which what it expresses is not relative to subject, and it expresses a relation.

the street?", and find appropriate the answer "Because last night it rained"; in this case we say that last night's rain is, for John, the reason of the puddles in the street. But it may also happen that, in the same situation, Mary asks herself: "Why does John believe that last night it rained?", and finds approriate the answer "Because he has seen puddles in the street"; in this case we say that the puddles in the street are, for Mary, the reason of John's belief that last night it rained. In the former case "reason" means the proposition a subject $\mathcal{S}$ is authorized to believe by its being the best explanation of the data available to $\mathcal{S}$; in the latter case it means the data which the belief a subject has is based on, hence the evidential factors of the believed proposition. I have privileged the second sense.

Is the notion of $\mathcal{S}$-truth-ground epistemically transparent? As the notion expresses a relation between cognitive states, the question is meaningful only if it concerns a cognitive state containing the representation of another cognitive state; if we conventionally call "observer" the subject occupying the latter cognitive state, and "believer" the one occupying the former, we can say that the question of transparency is meaningful only for an observer. Now, the definition given above entails that, if the notion of $\mathcal{S}$-justification for an atomic sentence α, the relation $\sigma' \geq_{\alpha} \sigma$ and the relation of identity between evidential factors in σ and in σ' are transparent, so is the notion of $\mathcal{S}$-truth-ground of α. For the transparency of the notion of $\mathcal{S}$-justification for α I have already argued above. The relation of identity between $\mathcal{S}$-reasons is evidently transparent. As for the relation $\sigma \geq_{\alpha} \sigma'$, let me observe that we are here confronted with a question that is essentially similar to one arising over and over again in any scientific domain: given a set of data about which a why-question arises, find the best answer. That, in specifiable circumstances, the content of this question does not depend on the context, but is uniquely determined, and that the question is decidable, are normal assumptions in every scientific investigation, and the fact that the scientific enterprise has some success suports the idea that they are plausible assumptions.

4.5 Chomskyan Problems Reconsidered

Let us see how the 'uninteresting' problems generated, according to Chomsky, by the externalist assumption illustrated in Chap. 1 vanish within the framework of our internalist semantics; and how, on the other hand, a direction can be suggested for the solution of some 'interesting' problems.

4.5.1 'Uninteresting' Problems Vanish

4.5.1.1 Singular Terms

Let us start from Chomsky's problem about "London". At the basis of the difficulty there was the model theoretic assumption that the denotations of names are *individuals* of some domain. In the semantics I have sketched, on the contrary, names denote $\mathcal{C}$-objects, i.e. *equivalence classes* of terms of **IRS**, and the problem disappears: while it is contradictory that an individual belongs to two disjoint sets, it is not contradictory that a set has subsets that are subsets of disjoint sets. It is therefore consistent to conceive the denotation of "London" as a set of terms of **IRS** internally articulated into subsets or 'parts', each of which is labeled by a label such as LOCATION, PEOPLE, AIR, BUILDINGS, INSTITUTIONS, and so on. The plausibility of this assumption is supported by the fact that it follows from the tenet that the epistemic content associated to a name in an atomic cognitive state specifies (either explicitly or implicitly) such labels; and this tenet is standard in lexical semantics. Analogous remarks can be made about "War and Peace", "this book" and similar examples.

4.5.1.2 Predicates and Compositionality

Let us come back to the example of the predicate "is green" in Chap. 1; the problem, for the externalist semantics, consisted in the fact that for each of the sentences (11)–(14) of that chapter we must assign to the predicate "is green" a different set as extension; since there is not such a thing as *the* extension of P, there is no way to define the truth-conditions of these atomic sentences, and of many others, in a compositional way. Let us now consider the meaning of the Application Question in our internalist semantics: the value of $f_{\text{GREEN}}(\text{o}) = 1$ in no way is constitutively related to a fact of the external world, in the sense that, in order to define that value, it is in no way necessary to make reference to facts of the external world; it is necessary to make reference to the function f_{GREEN}, i.e. to a certain routine, and to the $\mathcal{C}$-object o, a mental entity. What determines the value of f_{GREEN}, when it is applied to o, is information coming from different components of the mind, including perception, memory and, importantly, the lexicon. This means that even a 'prelinguistic' concept, as the concept GREEN presumably is, can be modified by the intervention of information coming from the lexicon. I am not concerned with how this information is coded in the lexicon, but with the fact that, in general, lexical information may be to the effect that the value of a $\mathcal{C}$-concept may depend on the nature of the $\mathcal{C}$-objects it is applied to; for instance, to the effect that $f_{\text{GREEN}}(\text{o}) = 1$ when o is a banana and its exterior surface reflects green light, when o is a stoplight and its lit lamp emits green light, and so on; or to the effect that $f_{\text{WHITE}}(\text{o}) = 1$ when o is wine and it is yellow, when o is hair and it is grey, and so on. More involved is the case of the 'linguistic' concept like HAVE WHEELS; even in this case it is perfectly possible that lexical information is delivered to **CIS** to the effect that $f_{\mathrm{HAVEWHEELS}}(\mathrm{o}) = 1$ when o is a unicycle or a onewheel and $f_{\mathrm{HAVEWHEELS}}(\mathrm{o}) = 1$ when o neither a unicycle nor a onewheel and has more than one wheel; the additional complication is that it is equally possible that lexical information delivered to **CIS** is different: that $f_{\mathrm{HAVEWHEELS}}(\mathrm{o}) = 0$ when o is a unicycle or a onewheel and $f_{\mathrm{HAVEWHEELS}}(\mathrm{o}) = 1$ when o is neither a unicycle nor a onewheel and has more than one wheel. Presumably it is an empirical question whether lexical information is of the former or the latter kind; from the semantical point of view what is relevant is that both alternatives are open, and this is granted by the fact that lexical information is in general constitutive of a $\mathcal{C}$-concept.

A consequence of this possibility is that the identity of some $\mathcal{C}$-concepts is not defined before the value of their application to certain $\mathcal{C}$-objects. This raises a question: if the $\mathcal{C}$-concept $\mathrm{f}_{\mathrm{WHITE}}$, for instance, has no priority over the value of $\mathrm{f}_{\mathrm{WHITE}}(\mathrm{o})$, then to grasp that concept it is necessary to know the value of its application to some $\mathcal{C}$-objects, and it is natural to wonder whether, in general, in order to grasp a $\mathcal{C}$-concept it is necessary to know the values of its application to all $\mathcal{C}$-objects and maybe to grasp other $\mathcal{C}$-concepts, with possible holistic consequences.

We have met an analogous problem concerning linguistic expressions in Chap. 2, when the relations between the compositionality and the context principles were discussed, and Dummett's proposal to conciliate them were illustrated. In the case of a predicate the same strategy can be adopted, by assuming that to grasp its meaning requires us to be able to grasp the thoughts expressed by *atomic* sentences in which it figures, but not to understand, for example, quantified sentences containing it; and an analogous assumption may be made about $\mathcal{C}$-concepts.[60]

In this way compositionality is preserved in the weak sense that the meaning of an expression does not depend on the meaning of logically more complex expressions, and therefore the danger of holism is avoided.

Summing up, the loss of strict compositionality does not jeopardise the possibility of a Frege-style semantics, once the externalist ingredient of Frege's semantics has been dropped and replaced by internalist assumptions about the denotations of singular terms and predicates. Moreover, the fundamental fact remarked above, that the value of $f_{\mathrm{C}}(\mathrm{o}_1 \ldots \mathrm{o}_\mathrm{n})$ is determined by a computation, still subsists even in presence of 'local holisms' as the one I have just stressed: they are relevant to the determination of f_{C}, not to the computational nature of application.

4.5.2 Possible Answers to 'Interesting' Problems

The denotation of a singular term in an atomic cognitive state is determined by computational principles of **CIS**, but also by computational principles of the faculty of language. For example, if we pretend for a moment that our reference language is English, we will be confronted with such sentences as (5) of Chap. 1:

[60] See Footnote 47. On this point see also Dummett (1991b), Sect. 10.1.

(5) He thinks that John is abroad,

about which our linguistic competence informs us that “he” cannot be coreferential with “John”. As the internalist semantics I am illustrating implements (aspects of) our **CIS**, which receives as input SEM(5), it will be impossible that “he” and “John” denote the same $\mathcal{C}$-object in any atomic cognitive state. In this sense syntax constrains the relation of denotation; in other terms, the relation of denotation is ‘lexically driven’. But the interesting question is whether the fact that, in this semantics, singular terms receive an internal denotation can contribute to explaining some phenomena which are interesting for the linguist. I think so.

In Chap. 1 I adduced the linguistic relevance of the distinction between homonymy and polysemy as an argument for the need of an internalist notion of denotation. Commenting on Chomsky’s suggestion that «It seems that a general principle of syntax-semantics interaction is involved», I remarked that such a principle should make reference to an internalist notion of denotation. Now it is possible to state the principle:

(19) Coindexing is possible if and only if the noun phrase and the pronoun are intended to refer to the same $\mathcal{C}$-object.

This formulation would be meaningless if “$\mathcal{C}$-object” were replaced by “object (of the external world)”; for consider the sentence

(20) The bank burned down and then it moved across the street:

the bank that burned down is a building, a concrete object, the bank that moved an institution, an abstract object, hence they cannot be the same object as syntax requires. On the contrary, it is perfectly possible that one $\mathcal{C}$-object includes subsets labelled, respectively, with BUILDING and INSTITUTION, hence (19) can be proposed as explaining the fact that «Referential dependence is preserved across the abstract/concrete divide.» (Chomsky, 2000: 181).

However, if (19) is a plausible «general principle of syntax-semantics interaction», a new question arises: how is it possible that, if all objects, *qua* classes of terms of **CIS**, belong to the same ontological category, our intuition clearly certifies that *there are* disjoint classes of objects?

Let me remind the reader that in Sect. 1.2.4 of Chap. 1 I argued that internalist theoretical notions of denotation and truth should meet four conditions; the question we are considering is only one of a host of problems arising in connection with condition (iv): to account for the essential aspects of the corresponding intuitive notions. I argued that one important reason why they *should* account for the corresponding intuitive notions is that psychology studies the mind, and the mind has representations, in particular representations of objects and of thoughts in a non-relational sense, i.e. representations and beliefs having an immanent content. Immanent content is—I suggest—what the mind experiences. Since mental experience includes objects, it is necessary to define a notion of object that permits us to explain our experience of objects; and since the objects we experience are precisely the objects of common sense, we must explain the common-sense notion (or notions) of object, in particular of object of reference: this is precisely what condition (iv) requires. This does not contrast with Chomsky's idea that

> what we take as objects [...] depend on their place in a matrix of human actions, interests, and intent in respects that lie far outside the potential range of naturalistic inquiry. (Chomsky, 2000: 21)

In other passages (Chomsky, 2000: 127) he mentions acts of human will as factors on wich the status of nameable thing crucially depends. What Chomsky refuses is that

> in studying the natural world (for that matter, in studying these concepts [the concepts of common-sense understanding], as part of the natural world), we view it from the standpoint provided by such concepts. (Chomsky, 2000: 20)

but he explicitly admits the possibility that

> the concepts of common-sense understanding can themselves be studied in some branch of naturalistic inquiry. (*ibid*).

Let us consider now our intuition that there are disjoint classes of objects. In the light of the distinction I have introduced, the problem is how to account for the fact that we experience certain objects as belonging to disjoint categories. The problem can be articulated into two questions:

(i) how to account for the fact that we experience objects as belonging to different categories?
(ii) how to account for the fact that the same object can be experienced as belonging to different categories?

As for the first question, I have already introduced in Sect. 4.5.1.1 the assumption that equivalence classes of terms are internally articulated into labeled subclasses; for example, a term that matches the piece of lexical information 'River' falls naturally into a subclass labeled RIVER; a visual description that matches a stored model of a human face falls into a subclass labeled HUMAN; and so on. The plausibility of this assumption is supported by the fact that it follows from the description of the epistemic content associated to a name in an atomic cognitive state as specifying such labels (either explicitly or implicitly); and this last description is standard in linguistics. I add now a further assumption: that, given an atomic cognitive state σ, we *experience* a $\mathcal{C}$-object o as belonging to a category C of $\mathcal{C}$-objects whenever the equivalence class it consists in is labeled 'C'. I do not see any plausible alternative to this assumption, in the light of the fact that it seems a priori clear that the grounds for assigning a $\mathcal{C}$-object to a category are to be searched for among the intrinsic properties of the class of representations it consists in, *not* in some relation of that set to some external object of that category; for otherwise we could not explain how a $\mathcal{C}$-object can be experienced as concrete although there is no concrete external object which it might stand, for example, in a spatial relation to; nor could we explain how we experience a $\mathcal{C}$-object as abstract or as fictitious. In virtue of these two assumptions, we can account for the fact that two objects, although belonging to the same ontological category—classes of terms of **CIS**—may be experienced as belonging to different, and even disjoint, ‘experiential’ categories.

Let us consider now the second question: how is it possible that the *same* $\mathcal{C}$-object is experienced as belonging to different categories? Observe that it may happen that the epistemic content associated to a name induces in the denotation of the name an articulation into different, and even disjoint, labeled subclasses. This is the case with the name “London”, whose epistemic content may contain such features as “London is a town in England”, “London is the inhabitants of a certain region”, and so on. The actual principles ruling the articulation of a name’s denotation may be extremely intricate; for example, it may be that we have an innate system of conceptual labels, and that the structure of this system imposes restrictions upon the internal organization of the name’s denotation. The only relevant point at present is that the possibility is open that the denotation of a name is articulated into different labeled subclasses. As a consequence of our second assumption, we experience the denotation of that name as belonging to two different, or even disjoint, categories.

The question remains of how it is possible that we experience it as *one* $\mathcal{C}$-object and not two. To this purpose it is useful to remember that in Sect. 4.2.1 it has been remarked that, in a linguistic cognitive state, a $\mathcal{C}$-object may be given through the use of a singular term, and that in such cases the process of constitution of the denotation of the singular term may be driven by the singular term itself; this boils down to the empirical hypothesis that, in the common cases in which a $\mathcal{C}$-object is introduced through the speaker’s use of a singular term, the **CIS** of the hearer takes the syntactical identity of the singular term as the identy criterion of the $\mathcal{C}$-object denoted by it. I shall try to argue for this hypothesis in Sect. 4.6.

To sum up: there is no contradiction in the fact that the same $\mathcal{C}$-object belongs to disjoint categories; the identity of the object is grounded on the syntactical identity of the name denoting it, while the categories it is experienced as belonging to are the labels of its subclasses. This seems to me a way to substantiate Chomsky’s idea that

> a lexical item provides us with a certain range of perspectives for viewing what we take to be things in the world, or what we conceive in other ways. (Chomsky, 2000: 36)

The distinction between what a $\mathcal{C}$-object is and how it is experienced yields also an answer to an objection that looms over any internalist semantics. It may be adapted as follows to the present one. For every atomic cognitive state σ, the principle

(21) “ν” denotes ν in σ

is intuitively valid; from it we obtain

(22) “Trump” denotes Trump in σ;

since “Trump” denotes in σ a $\mathcal{C}$-object, i.e. an equivalence class of terms of **CIS**, it follows that Trump is an equivalence class of terms, against the evidence: Trump is a man, not an equivalence class of terms. The answer at this point is obvious: if we keep present the foregoing distinction, the objection boils down to the true but innocuous remark that Trump is experienced by us as an object of the external world, not as a class of mental representations.

4.6 On Sense and Denotation

The central question of a theory of meaning is: What is the meaning of a sentence? Two basic aspects of this question have been highlighted by Quine, on the one hand, and by Dummett and Chomsky, on the other. Quine has condensed the former in the motto "No entity without identity": in order to admit an entity of some kind into our ontology, we must explain the identity criterion for entities of that kind. Chomsky and Dummett, with different but converging arguments, have stressed the vacuity of any explanation of meaning that is not accompanied by an explanation of what it is to know it.[61] If we combine these two aspects, the central question requires a preliminary answer to the question: When do we know that two sentences have the same meaning, i.e. that they are synonymous? If we keep present Frege's compositionality principle, the last question articulates into two others: "When do we know that two singular terms are synonymous?" and "When do we know that two predicates are synonymous?" Only after having made our intuitions about these questions explicit will I propose *explicantia* for synonymy and, consequently, a definition of the notion of sense for singular terms, predicates and sentences.

4.6.1 On Different Notions of Synonymy

As a matter of fact, it seems to me that we have quite different intuitions about the synonymy of different kinds of expressions, and even about different kinds of designators.[62] One of the merits of the internalist semantics I am proposing is its flexibility concerning the *explicans* it allows to propose for the relation of synonymy of different kinds of designators. In this section I propose *explicantia* on the basis of my own intuitions about the meaning of different kinds of expressions, but different proposals are possible within the same conceptual framework.

Let us begin by distinguishing two possible *explicantia* of the intuitive relation of (relative) synonymy:

Definition 14 Let δ and δ' be designators of $\mathcal{L}$ and $\mathcal{S}$ a subject; then δ and δ' are *equi-intensional for* $\mathcal{S}$ (or $\mathcal{S}$*-equi-intensional*) iff $[\![\delta]\!]_{\mathcal{S}} = [\![\delta']\!]_{\mathcal{S}}$.

Definition 15 Let δ and δ' be designators of $\mathcal{L}$, $\mathcal{S}$ a subject and σ a cognitive state of $\mathcal{S}$; then δ and δ' are *equi-extensional for* $\mathcal{S}$ *in* σ (or $\mathcal{S}$*-equi-extensional in* σ) iff $[\![\delta]\!]_{\mathcal{S},\sigma} = [\![\delta]\!]_{\mathcal{S},\sigma}$.

Let us now gather some intuitive data concerning competence, meaning and synonymy of different kinds of designators, in order to find the most adequate *explicantia*. It should be stressed that the intuitive notions I am invoking here are the

[61] Chomsky conceives a grammar as a theory of linguistic *competence*; Dummett conceives the theory of meaning as a theory of understanding, i.e. of knowledge of meaning.

[62] I shall call "designator" an expression that is either a singular term or a predicate.

internalist ones of competence a subject $\mathcal{S}$ has of his own I-language, of meaning in $\mathcal{S}'$ own I-language, of synonymy-for-$\mathcal{S}$; their theoretical *explicantia* will be used in Chap. 7 to tackle the problems of belief reports.

4.6.1.1 Singular Terms

(a) We consider as semantically incompetent about a singular term ν a subject who does not know that with ν (s)he cannot refer to more than one object.[63]
(b) We consider possible that a subject is semantically competent about ν (i.e. knows the meaning of ν) without knowing its reference (i.e. without associating to ν information sufficient to distinguish its reference from any other object). For example, we would not say, of a subject who only knows of Carl Hempel that he was a philosopher, that (s)he ignores the meaning of "Carl Hempel".
(c) On the other hand, we do consider as semantically incompetent a speaker who ignores that "Carl Hempel" is a personal name, or that it is a masculine name—features that of course are not sufficient to individuate an object, but that permit to ascribe it to some maximally general category.
(d) Another intuitive datum is that when a subject semantically competent about two names does not know their references, (s)he does however assume that their meanings are different. For example, a subject who associates to "Carl Hempel" and "Rudolf Carnap" only the features that Carl Hempel was a philosopher and that Rudolf Carnap was a philosopher, respectively, does however believe that the meaning of "Rudolf Carnap" is different from the meaning of "Carl Hempel".
(e) We would not deem semantically incompetent about two different singular terms a subject who ignores that they denote the same thing. For example, we would not consider semantically incompetent about "Hesperus" or about "Phosphorus" a subject who only knows that Hesperus and Phosphorus are celestial bodies, ignoring that they are the same.
(f) However, a subject may acquire the piece of information that Hesperus and Phosphorus are the same celestial body, and, when (s)he does, (s)he normally updates her/his cognitive domain by merging the two objects into one.
(g) In many cases it is possible that information associated to a singular term varies as time passes, and that its meaning remains the same; for example, we may first associate to "Carl Hempel" the piece of information that Carl Hempel is a philosopher, then come to know that *he* is the author of *Aspects of Scientific Explanation*, that *he* had the face shown by a photo, and so on.
(h) However, in some cases it is *not* possible that information associated to a singular term varies and that its meaning remains the same. Imagine that Mary associates to "Chomsky" the piece of information that Chomsky is a postman of Brooklyn because Paul, a friend of hers, once told her "That is Chomsky" while pointing at a postman in Brooklyn; later on, in a bookshop, Mary finds a book with

[63] I call therefore "reference" of a singular term ν the (mental) entity one is speaking of when one uses ν in an atomic sentence.

the name "Noam Chomsky" on its cover; it is probable that at this point she glances through the book, presumably because it is not common, but possible, that a postman of Brooklyn writes a book; in that case it is probable that it is an autobiography or something like that; turning over the pages of the book Mary makes a test, and the outcome is negative: the book deals with linguistics. Let us say that in such a case the new piece of information that Chomsky is a linguist is 'incompatible' with the old one, in the sense that it consists in a property that cannot, according to Mary, be ascribed to the same object Mary referred before with that name. Mary has several options: she may refuse new information, or discard old information, or she may update her cognitive domain by splitting one object into two; in this last case, however, the meaning of "Chomsky" changes as well (maybe by being replaced by two different meanings for two different names).

Summing up, we can know the meaning of ν without knowing its reference; hence the meaning of ν cannot be equated with its reference. On the other hand, it may happen that information associated by us to two different singular terms ν and ν' is the same, while their meanings are different; hence the meaning of ν cannot be equated with information associated to it (which I call its epistemic content). This suggest that, for our computational apparatus, the meaning of a singular term ν—what one knows when one has semantic competence of ν—is neither its epistemic content nor its reference, but a method or a procedure for determining the reference, i.e. a function associating to every atomic cognitive state the reference of ν in that state. At this point it is natural to articulate the intuitive notion of meaning of a singular term into two notions—sense and denotation—and to propose $\mathcal{S}$*-intension* as the *explicans* of sense-for-$\mathcal{S}$ of ν and $\mathcal{S}$*-equi-intensionality* as the *explicans* of synonymy-for-$\mathcal{S}$ between singular terms:

Definition 16 Two singular terms ν and ν' are synonymous-for-$\mathcal{S}$ iff they are $\mathcal{S}$-equi-intensional.

In other terms, a subject $\mathcal{S}$ who grasps the sense of ν—who is semantically competent about it—is capable to compute the function $[\![\nu]\!]_{\mathcal{S}}$.

It seems to me that this theoretical choice accounts for the intuitive data. First, a general assumption made in Sect. 4.4.1 is that the faculty of language has access to **CIS**; this means that lexical information associated to any singular term ν is part of ec_{ν}, hence part of what is known by any subject who is semantically competent about ν; this accounts for (c). Second, $[\![\nu]\!]_{\mathcal{S}}$ is the function $\lambda x\, den(\nu,x)$, hence a method for determining the denotation of ν. A function has at most one value; this accounts for (a). Moreover, the method $\lambda x\, den(\nu,x)$ consists in is not variable, and not contextual: it is ideally the same for every time and context (i.e. for every atomic cognitive state), and its difference from the sense of every other singular term rests upon the syntactical identity of ν; this accounts for (d) and for (b). Third, it is plausible to assume that $den(\nu, \sigma)$ is *known* to be a $\mathcal{C}$-object, in the sense that our computational apparatus treats it as a $\mathcal{C}$-object. As I remarked in commenting on the definition of *den* in Sect. 4.2.1, situations of two kinds should be distinguished, according as the epistemic content associated to ν is insufficient or sufficient to univocally identify its denotation with a $\mathcal{C}$-object o already present in the cognitive domain. In situations of the former kind the identity criterion of the denotation of ν is grounded in the syntactical identity of ν; hence if ν is syntactically different from ν', its denotation will be treated by our computational apparatus as different from the denotation of ν'; this accounts for (e). However, new information may be added to the epistemic content of ν—even the piece of information that the denotation of ν is identical to the denotation of ν'; this accounts for (f).[64] In situations of the latter kind the identity criterion of the denotation of ν is grounded in the identity criterion of o; therefore, in situations of this kind lexical information associated to a singular term ν must 'harmonize', or be 'compatible', with our prelinguistic intuitions about objects, in the sense that it cannot attribute to the denotation of ν properties that cannot be enjoyed by o. This accounts for (g) and (h): in (g) the features that Carl Hempel is the author of *Aspects of Scientific Explanation* and that he had a certain face are 'compatible' with the feature that Carl Hempel is a philosopher; in (h) the feature that Chomsky is a linguist is 'incompatible' (for Mary) with the the feature that Chomsky is a postman of Brooklyn.

There are some categories of singular terms about which our intuitions are different from (a) to (h). One example is *numerals*. We would say that someone who doesn't know that 4 comes after 3 and before 5 is incompetent about the meaning of "4"; whereas we would not deem incompetent about the meaning of "the smallest prime number" someone who doesn't know that it denotes 2. Parallely we do deem semantically incompetent about "IV", in Roman notation, or "100", in binary notation, a subject who ignores that they denote the same number as "4"; moreover, we judge all these numerals synonymous. Numerals seem therefore to have the peculiar property that knowing their meaning is sufficient in order to recognize their denotation.

While only numerals have that property among singular terms for numbers, it seems that *all* proper *names of fictitious entities* have that property; for example, we would deem incompetent about the meaning of "Charlus" someone who is not capable to univocally identify that character of the *Recherche* among all the literary characters; and incompetent about the meaning of "Superman" someone who doesn't know that Superman is Clark Kent, and so on. Again, to account for this it is sufficient that the lexical epistemic content of these names includes relevant information.

Analogous is the case of *pseudonyms*: intuitively, someone who doesn't know that "Álvaro de Campos" is a pseudonym of Fernando Pessoa is deemed incompetent about its *meaning*. A related but 'inverse' example: someone who doesn't know that "Amandine-Lucie-Aurore Dupin" is the *true name* of George Sand is deemed incompetent about its meaning.

In order to account for these intuitions it is sufficient to postulate, in the case of a numeral, that its lexical epistemic content contains information about its position in

[64] In the situation (or atomic cognitive state) described in (f) the subject acquires a piece of 'linguistic' information about the names "Hesperus" and "Phosphorus". Different would be a situation in which the subject *discovers* that Hesperus is Phosphorus; in this case the subject would be in a situation of the second kind, and the best explanation of the data available to her/him would be that the denotations of the two names are one and the same $\mathcal{C}$-object.

the chain of numerals—an information available to whoever is able to count; in the case of the name of a character of a literary work, that its lexical epistemic content contains all features associated to it in that literary work; in the case of a pseudonym, that its lexical epistemic content contains the relevant identity.

A final remark. I have not spoken about definite descriptions, which I shall not pay any special attention to in this book; I'll just say that I treat them as singular terms.

4.6.1.2 Predicates

(i) We consider as semantically incompetent about a predicate π a subject who does not know that with π (s)he can apply one concept to n-tuples of objects.[65]
(j) We are typically inclined to consider semantically incompetent about a predicate a speaker who does not associate to it information sufficient to apply its reference to n-tuples of objects of the appropriate category; for example we would say, of someone who associates to "red" only features associated to numbers, and to "even" only features associated to colours, that (s)he ignores the meanings of "red" and "even".
(k) Information associated to a predicate may vary as time passes, and its meaning varies accordingly; for example, if Mary first associated to "prime" the feature that a prime number is divisible only by itself and by 1, and then added the feature that it is greater than 2, we are inclined to say that the concept denoted by "prime" for Mary has changed.
(l) It is possible that a subject associates to two predicates different features but the same reference. For example, if $\mathcal{S}$ associates to "physician" the feature that a physician is a person skilled in the art of healing, and to "doctor" the feature that a doctor is a person engaged in the practice of medicine, $\mathcal{S}$ will apply the two predicates to the same objects.

Summing up, we cannot know the meaning of a predicate π without knowing its reference (i.e. without being capable to apply its reference to objects); on the other hand, the meaning of π cannot be equated with information associated to it (which I call its epistemic content), since it is possible that two predicates to which a subject associates two different epistemic contents have the same meaning. This suggests that, for our computational apparatus, the meaning of a predicate π—what one knows when one has semantic competence of π—is its denotation, i.e. the concept defined by its lexical epistemic content. At this point it is natural to propose the concept denoted by π in σ as the *explicans* of its meaning-for-$\mathcal{S}$:

Definition 17 Two predicates π and π' are *synonymous-for-$\mathcal{S}$ in σ* iff they are $\mathcal{S}$-equi-extensional in σ.

[65] I call therefore "reference" of a predicate π the concept one is applying to objects when one uses π in an atomic sentence.

This may seem to conflict with our intuitions, because the denotation of an expression for a subject normally depends upon the atomic cognitive state occupied by the subject, whereas the meaning of an expression is felt as independent of cognitive states; but it should be noticed that if two predicates are $\mathcal{S}$-equi-extensional in σ, then they are $\mathcal{S}$-equi-extensional in σ', for all σ' of $\mathcal{S}$,[66] hence they are $\mathcal{S}$-equi-intensional.[67] At this point someone might wonder why not to adopt directly $\mathcal{S}$-equi-intensionality as the *explicans* of synonymy-for-$\mathcal{S}$ between predicates. The reason is that I shall *not* articulate the intuitive notion of meaning of a predicate for a subject into the two notions of sense and denotation, and that the computation of extensions is in general quicker than the computation of intensions; but the other choice would be legitimate.

It seems to me that this theoretical choice accounts for the intuitive data. First, the denotation of a predicate in an atomic cognitive state is a $\mathcal{C}$-concept, i.e. a function taking as arguments n-tuples of $\mathcal{C}$-objects and giving as values 1 or 0; this accounts for (i). Second, the denotation of π in σ is an algorithm computing the discrimination function defined by the lexical content of π; hence it is the the lexical content of π that rules the subject's discriminating behaviour: if no epistemic lexical content is associated to π, no algorithm is stated, no function is defined, no $\mathcal{C}$-concept is denoted, and no discriminating behaviour is determined: the subject ignores the meaning of π; this accounts for (j). Third, a change in the epistemic lexical content associated to π may determine a change in the function defined, hence in the $\mathcal{C}$-concept denoted, hence in the subject's discriminating behaviour: the meaning of π has changed; this accounts for (k). Finally, not *every* change in the epistemic lexical content associated to π determines a change in the function defined: it is possible that different lexical contents define the same function, hence the same denotation of the predicate; this accounts for (l).

Predicates of natural language are traditionally divided into two categories—sortal and adjectival predicates[68]—which correspond to different intuitions we have

[66] This is a consequence of two facts: (i) that the denotation of a predicate is determined by its lexical epistemic content, and (ii) that the lexical epistemic content of a predicate does not vary across atomic cognitive states.

[67] A possible counterexample to the claim is the pair of predicates "Greek" and "Hellenic", since there are atomic cognitive states in which they are $\mathcal{S}$-equi-extensional, for some $\mathcal{S}$, and state in which they are not. However, it should be noticed that they are defined predicates ("inhabitant of Greece" and "inhabitant of Hellas"), and that their possible non-$\mathcal{S}$-equi-extensionality depends on the possible non-$\mathcal{S}$-equi-extensionality of the names "Greece" and "Hellas". The claim can therefore be restricted to primitive predicates, or to 'pure' predicates (predicates whose definition does not contain names).

[68] For the sake of definiteness I use here the characterization of sortals adopted by Guarino, Carrara and Giaretta (1994), according to which a sortal is a predicate that provides countability and is temporally stable, where a predicate provides countability if it does not apply to any proper part of what it applies to, and a predicate is temporally stable provided that, if it applies to something at one time then it must at some other times as well. According to this definition "student" is a sortal, "red" is not because it does not provide countability, and "kitten" is not because it is not temporally stable. I shall call "adjectival" predicates that are not sortal.

about their meaning; (i)–(l) concern adjectival predicates. About sortals at least the following must be mentioned:

(m) Information associated to a sortal may vary as time passes, but its meaning doesn't vary accordingly: it remains the same as long as new information is 'compatible' with a sort of stereotype; for example, when we accept the piece of information that a whale is not a fish, we are not inclined to say that the species denoted by "whale" has changed, as we are when we accept the piece of information that a whale does not live in the sea. It seems that, for our computational apparatus, LIVES IN THE SEA is part of the stereotype associated to "whale", while FISH is not.

(n) It is well known that many sortals can plausibly be conceived as proper names, of natural kinds, of species and so on.[69] Of course there is the predicative use as well, and under this respect there is an obvious analogy between adjectivals and sortals; but even in this case there is an intuitive difference between the two kinds of predicates: in order to be granted to assert that an object is red it is sufficient to verify that the feature associated to "red" matches a feature of the object, but in order to be granted to assert that an object is a horse this is not sufficient: we must also verify that it has a sufficient number of features belonging to the stereotype of HORSE.

(o) Our intuitions about synonymy between sortals seem to be oscillating. In the case of "doctor" and "physician" most of us are probably inclined to deem semantically incompetent a subject who ignores that they are synonymous; on the other hand, there are pairs, like "Greek" and "Hellene", in which it seems perfectly possible that a subject is competent about both sortals even if (s)he ignores that they are synonymous.[70]

In order to account for (m) it is sufficient to postulate that the lexical epistemic content of a sortal contains information sufficient to individuate a stereotype: the denotation of the sortal will be the $\mathcal{C}$-concept defined by the stereotype. As a consequence, if the lexical content is modified with features compatible with the stereotype, the $\mathcal{C}$-concept denoted by the sortal will remain the same; if the lexical content is modified with features incompatible with the stereotype, the $\mathcal{C}$-concept denoted by the sortal will be different. In order to account for (n) it seems plausible to assume that a sortal predication like "Bucephalus is a horse" can be analysed along lines suggested by the following paraphrase: "Bucephalus belongs to the species Horse",[71] where "Horse" is the proper name of a species. However, I shall not try to implement this suggestion here. As for (o), I will come back to these intuitions in Sect. 7.2.5.1 of Chap. 7.

[69] See also Grandy (2016).

[70] It should be remembered that we are speaking here of intuitions about synonymy, hence of the intuitive notion of synonymy in an E-language, not of the theoretical notion of synonymy in an I-language.

[71] For a development of this suggestion see Beyssade and Dobrovie-Sorin (2005).

4.6.1.3 (Atomic) Sentences

In consequence of compositionality, if δ and δ' are two designators and α and α' are two atomic sentences such that α' is $\alpha[\delta'//\delta]$, then α and α' are $\mathcal{S}$-equi-intensional iff δ and δ' are $\mathcal{S}$-equi-intensional.[72] This remark motivates the choice of *$\mathcal{S}$-equi-intensionality* as the *explicans* of synonymy for $\mathcal{S}$ between atomic sentences:

Definition 18 Two atomic sentences α and α' are *synonymous for $\mathcal{S}$* iff they are $\mathcal{S}$-equi-intensional.

This definition points to an analogy between singular terms and sentences: knowledge of their meaning-for-$\mathcal{S}$ amounts to being capable to compute their $\mathcal{S}$-intension. But the analogy stops here; in particular, I think no articulation of the intuitive notion of meaning of sentences into sense and denotation is appropriate. The rationale for such an articulation, in the case of singular terms, is the fact that subjects have, already at a prelinguistic level, an innate notion of object, of which a crucial aspect is that one and the same object can be given in several ways, or through a multiplicity of perspectives; this presupposes that we are capable to cognitively distinguish the object, as invariant, from the varying presentations of it. When language comes in, it is therefore natural to look for linguistic correlates of objects and perspectives or presentations, and the notions of denotation and sense, respectively, of a singular term are the natural candidates.[73] In the case of sentences there is nothing analogous to the prelinguistic articulation into objects and perspectives, which a distinction between there sense and denotation could be grounded on; Frege's proposal of the dichotomy truth-values/thoughts sounds quite unnatural.

This does not mean that Frege's distinction between understanding a sentence (grasping the thought it expresses) and judging it (recognizing that thought as true) has to be given up; on the contrary, it corresponds to what is perhaps the fundamental intuition we have about the meaning of sentences, according to which it is quite possible to understand a sentence, i.e. to know its meaning, without knowing whether it is true/evident; what I am proposing is simply not to equate a sentence's being evident (or judged true) with its denotation.

4.7 A Comparison with the Neo-Verificationist Model of Sense

It may be useful to compare the theoretical *explicantia* I have proposed for the intuitive notion of meaning of singular terms and predicates with the explanation

[72] Remember that two predicates are $\mathcal{S}$-equi-intensional iff they are $\mathcal{S}$-equi-extensional in σ, for some atomic cognitive state σ of $\mathcal{S}$.

[73] This is, roughly, Frege's idea. I have suggested a different, but related, account of sense. See the next section.

proposed by Dummett, in his book *Frege. Philosophy of Language*, of the Fregean notion of sense:

> To know the sense of a proper name is to have a criterion for recognizing, for any given object, whether or not it is the bearer (referent) of that name; to know the sense of a predicate is to have a criterion for deciding, for any given object, whether or not the predicate applies to that object [...]. (Dummett, 1973: 229)

A crucial role is played, in this explanation, by the phrase "any given object",[74] for

> An object cannot be recognized as the referent of a proper name, or as something to which a particular predicate applies, [...] unless it has first been singled out in some definite way. [...] Thus, in order to construe the interpretation of sense which we are considering, we have to suppose that, for each categoy of objects, there is some favoured method of being 'given' an object of that category to which the criteria of identification of application relate. (Dummett, 1973: 232)

In the case of ostensible objects (objects that can be pointed to), the favoured method is demonstrative identification, i.e. the use of a demonstrative accompanied by a pointing gesture:

> Thus the sense of a personal proper name would consist in the criterion for identifying a man pointed to as the bearer of the name, and, in general, the sense of any proper name '*a*' of an ostensible object (an object that can be pointed to) will consist in the criterion for the truth of what we may call 'recognition statements' of the form 'That is *a*'. Likewise the sense of a predicate 'ξ is P' whose range of definition comprises only ostensible objects would consist in the criterion for the truth of what we may call 'crude predication' of the form 'That is P'. (*Ibid.*)

In the case of abstract objects,

---

[74] The importance of the role played by the way an object is given is a hallmark of Dummett's interpretation of Fregean sense, as opposed to the 'mode of presentation' interpretation:

> the conception of the mode of presentation of an object, the way in which it is given to us, conflates two distinguishable things: our grasp of the linguistic significance of a term; and our apprehension of an object which we can presently perceive or discern. The most natural way to understand the conception of the sense of a […] singular term is […] as the grasp of whatever principle governs the correct identification of an object as the referent of the term. If, now, we add that we can never be given an object save under some mode of presentation, and identify the mode of presentation with a sense as thus conceived, an apparent paradox is created: we seem to be unable to break out of the circle of senses to the real object beyond. To speak of the identification of an object as the referent of a term presupposes a means of picking out the object for which no further question of identification arises: our problem then was how we could arrive at such a terminus. But the sense in which an object may be said to be given to us in a particular way when it is observed is altogether different from that in which it is said to be 'given' by our grasp of the sense of a name or description used to speak about it. (Dummett 1981b: 142–143).

The explanation of "a given object" has been the starting point of my approach to the semantics of singular terms as well.

> We have to find [...] some preferred range of names for them: e.g., in the case of natural numbers, we might select the numerals from some particular system of notation [...]. Thus, on such an account, the sense of an arbitrary numerical term n would consist in the criterion for deciding the truth-value of any sentence of the form $\ulcorner \nu = \kappa \urcorner$, where κ is a numeral [...]. (Dummett, 1973: 499)

The assumption of the existence, for each category or type of objects, of a 'favoured' or 'direct' method of identification of objects of that type is a fundamental component not only of Dummett's interpretation of Frege, but also of neo-verificationist theories of meaning and of mathematics[75]; it is therefore worthwhile to discuss it in some detail.

A preliminary remark is in order. The assumption I am going to discuss ought to be carefully distinguished from another neo-verificationist assumption discussed in Chap. 2: that it is necessary to introduce into the realm of intuitionistic proofs a distinction between canonical and non-canonical ones. The distinction between the two assumptions is blurred within the neo-verificationist theories of meaning, because the canonical/non-canonical distinction is extended from logically complex sentence to atomic ones, in which case the canonical/non-canonical distinction is equated with a distinction between direct and indirect proofs. However, the canonical/non-canonical distinction and the direct/indirect distinction are intuitively different: the former concerns the linguistic structure of arguments, while the latter is of an epistemological nature, concerning the way in which knowledge of atomic propositions is acquired.[76] As a consequence, the rationale for the canonical/non-canonical distinction is intuitively different from the rationale for the direct/indirect distinction. I have discussed the former in Sect. 2.4.2. of Chap. 2; here I shall discuss the latter, after having discussed the assumption of the existence of a 'favoured' or 'direct' method of identification of objects of different sorts.

4.7.1 On Direct Methods of Object Identification

There are at least three theoretical reasons, according to neo-verificationists, to make such an assumption. The first has already been quoted in footnote 74:

> To speak of the identification of an object as the referent of a term presupposes a means of picking out the object for which no further question of identification arises. (Dummett, 1981b: 143)

In other terms, as the name "recognition statement" suggests, the preferred way of giving physical objects is intended to ensure *recognition* of the given object by the subject whom it is given to; and this because only if recognition of the *given* object is ensured is it plausible to say that to know the sense of a proper name is to have «a

[75] See for instance Dummett (1981b), Chap. 6, Prawitz (1994a), Martin-Löf (1985).

[76] This does not exclude the possibility of seeing the direct/indirect distinction as a particular case of canonical/non-canonical distinction, as in the neo-verificationist approaches.

criterion for recognizing, for any *given* object, whether or not it is the bearer of that name», as Dummett's interpretation of Fregean sense requires.[77]

The second reason is related to the necessity of satisfying the requirement of compositionality:

> Given the way in which a sentence is constructed out of its component expressions, and given the senses of these expressions, we have, then, an understanding of the means whereby that sentence can be recognized as having one or other truth-value. [...] For any given sentence, there will always be something which we may regard as being the most *direct* means of recognizing it as having one or other truth-value; not in the sense of that means which involves the least exprenditure of effort, nor the most practicable or the most certain; but that which corresponds, step by step, with the way in which the sense of the sentence is determined from those of its components. (Dummett, 1981b: 236–237)

Consider for instance the atomic sentence

(23) Frege is a great philosopher:

in order to directly recognize—or verify—its truth-value we need two things: a verification of "That is Frege" and a verification of the truth-value of "That is a great philosopher", where the former (the sense of "Frege") consists in

(24) a function *f* from men pointed to to $\{1, 0\}$ according as the man pointed to is or is not identified as Frege

and the latter (the sense of "is a great philosopher") consists in

(25) a function *g* from men pointed to to $\{1, 0\}$ according as the man pointed to is or is not a great philosopher.

A third reason has been pointed out by Prawitz in response to Quine's inextricability thesis[78]:

> What counts as evidence, as a sufficient condition for asserting a sentence, depends not only on the meaning of the sentence, but also on our theories about the world. [...] In *Word and Object*, Quine discusses in detail how problematic the endeavour [is] to separate meaning from information already in the case of a simple sentence such as 'Rabbit' [...]. (Prawitz, 1994a: 135)
>
> What Quine's [...] examples show quite clearly is that in a verificationist theory of meaning one must distinguish between what I shall call *direct* and *indirect* evidence, and that the meaning of a sentence is to be identified only with what is counted as direct evidence. (Prawitz, 1994a: 137)
>
> As stressed by Føllesdal, the inextricability thesis is not epistemological but metaphysical; the point is not that we may not always know how to separate meaning from information but that there are not two such items to separate. (Prawitz, 1994a: 138)

[77] See also Evans (1982), 98:

> the mode of identification is that in which the subject, given the object in that way, is then able to determine the applicability to the object of the properties that may be ascribed to him.

[78] The name "inextricability thesis" is due to Dummett; see Dummett (1978: 387).

I hold that none of these reasons is cogent. Starting from the first, the question is whether demonstrative identification is such as to ensure recognition. The answer is clearly negative. For example, if Paul is introduced to his next-door neighbour with the words "This is Mr. John Smith", and Paul mistakes the man pointed to for his baker (maybe because his baker and John Smith are identical twins), there is no recognition on Paul's part, and the recognition statement would not transmit to him the (Fregean) sense of "John Smith".[79]

Another difficulty concerns the very notion of recognition statement. The ability to determine the truth-value of the recognition statement "This is John" can be plausibly suggested as an explanation of the knowledge of the meaning of the name "John" only if the statement is interpreted as expressing the identity of an unknown entity named "John" with the known man that is indicated, and which is known just because he is perceptually present. But it is not always the case that a statement of the form "This is John" is, or even *can*, be interpreted in this way. Suppose for example that both I and Paul know, of a certain boy b we have never met, that he got lost in the park, that his name is "John", that he is four years old, that he wears a red apron, and that he is fair-haired. Suppose I walk through the park and I meet a boy who corresponds to what I know about him; in this situation I am entitled to say to Paul, who is walking with me: "This is John". It is true that I use a sentence of the form "This is n", but I use it in a completely different way from the one assumed by Dummett as standard when he speaks of recognition statements: what we know is the referent of "John", and the boy pointed to becomes known only in virtue of his identity with the referent of "John". As a consequence, we are left with an unexplained notion of recognition sentence, since the purely syntactic characterization given by Dummett in terms of the use of a demonstrative is too large.

A further, crucial, difficulty is introduced by Dummett himself (Dummett, 1973: 238). Consider (23): as we have seen, the most direct means of recognizing its truth-value includes the functions *f* and *g* described in (24) and (25), whose argument is … what? A wad of bones? It is impossible to believe that someone presented with it we would be capable to recognize it as Frege. Someone we would meet by returning to the past? The idea that such an extraordinary close encounter is both *possible* and in any sense *relevant* to verify the truth-value of (23) is hard to swallow. Dummett replies that, although (24) and (25) describe the *direct* verification of (23), it may happen that the most frequent, or even the only actually available, verification of (23) is *indirect*; for example, the current verification of (23) consists in reading Frege's works and judging their philosophical value. However, his remark does not answer a new objection strictly related to the original one. The trouble is the following: a direct verification is a verification because it «proceeds in accordance with what

---

[79] Dummett recognizes an analogous difficulty:

> There is, indeed, no ground to claim even the process of ostension, considered as involving appeal to a sortal concept, as infallible: there is often room for a mistake over whether an object of a given sort is present at all, and, if so, only one. (Dummett 1981b: 143)

the sentence means» (Prawitz, 1994b: 491); but why is an indirect verification a verification? The neo-verificationist's answer consists in defining it as something that «can be reduced to a direct, canonical one, which can be seen as a justification of these indirect means of verification.» (*Ibid.*) In the present case, reading Frege's works is justified as a verification of (23) because it is a *shortcut* (the expression is Dummett's), i.e. a shorter way to get the same result as the *direct* way. Hence, if we want to justify the use of an indirect verification of (23) we must assume (i) that a direct verification exists[80]; (ii) that the given verification is reducible, in some sense, to the direct one; (iii) that reducibility of the given verification to the direct one is relevant in order to establish its status of verification of (23); and all these assumptions are highly questionable.

Dummett explicitly concedes that

> There is, ideed, nothing mandatory about selecting, as this preferred method, the use of ostension, in the case of concrete objects - and there is nothing in Frege to support this selection; (Dummett, 1973: 238-239)

but the preceding discussion suggests that what is wrong in Dummett's assumption is not the choice of ostension instead of another canonical way of giving concrete objects, but the more general assumption that the definition of sense ought to make reference to a fixed way of giving concrete objects that ensures recognition. As a matter of fact, such an assumption is not necessary, because another way is open to implement Dummett's interpretation of Fregean sense: to let the way objects are given vary, and conceive their recognition as depending not only on the way they are given, but also on information associated with the name. More precisely, the (Fregean) sense of "Frege" would not be the function *f* described in (24) but

> (24') a function *f'* from presentations of men to {1, 0} according as the presentation matches or not the information associated to "Frege" (thereby granting recognition of the presented object as the denotation of "Frege").

Analogously, the (Fregean) sense of "is a great philosopher" would not be the function g described in (25) but

> (25') a function *g'* from presentations of men to {1, 0} according as the man recognized as presented is or is not a great philosopher.

Once Fregean senses are reinterpreted in this—to my view more plausible—way, the second theoretical reason for assuming the existence of a 'favoured' method of identification of concrete objects vanishes, because it becomes possible to meet the requirements of compositionality *without* the assumption of a 'favoured' method of identification. A compositional verification of (23) will result from the composition of (25') with (24'); if someone wanted to call such a compositional verification "canonical", I would have no objection: the important thing is that it is defined without making any mention of any favoured method of identification.

[80] The assumption that, if there is a verification of α, then there a direct/canonical verification of it is called "The Fundamental Assumption" by the neo-verificationists. For a discussion of it see for example Dummett (1991a, 1991b), Chap. 12.

As for the third theoretical reason, I cannot see how reference to ostensive identification could grant such a distinction. Ostensive identification was proposed by Dummett as a favoured method of identification because of its (alleged) capacity to ensure recognition of the given object, but what is needed to ensure 'extricability' is that all information extracted from an object pointed to is part of the meaning of its name, and not collateral information: and this is not the case. Suppose we meet Kripke and we see that he wears spectacles: it does not follow that wearing spectacles is part of the sense of "Kripke", nor that it is relevant for the truth-value of

(26) Kripke is a great philosopher.

Prawitz's answer to Quine's inextricability thesis seems to me to suffer from a further inadequacy. Prawitz takes into consideration the meaning of sentences, and his examples ("Gavagai", "It is raining") give no suggestion about whether and how he would treat names differently from predicates: since he speaks only «about the possibility of sorting out meaning and [collateral] information» (Prawitz, 1994a: 136), he seems to imply that the sense of both names and predicates consists in some amount of information connected to them; I have explained above why I consider this idea adequate in the case of predicates, inadequate in the case of names.

Owing to the clear intuitive differences stressed above between names of ostensible objects and numerals, Dummett's assumption of a 'favoured' way of giving objects seems plausible in the case of natural numbers; when a number is given through a numeral its recognition *is* guaranteed, so that it is quite natural to consider numerals as the canonical presentations of natural numbers; and it seems plausible to assume that there are canonical ways of giving other mathematical entities.[81]

4.7.2 The Direct/Indirect Distinction

I have criticized Prawitz's answer to Quine's inextricability thesis, but I agree on the possibility of distinguishing between knowledge of meaning and knowledge of collateral information, and on the theoretical necessity of doing it, because Quine's argument for it seems to me unconvincing. Here is Quine's argument in *Word and object*:

> Intuitively the ideal would be to accord to the affirmative meaning of "Gavagai" just those stimulations that would prompt assent to "Gavagai?" on the strength purely of an understanding of "Gavagai", unaided by collateral information [...]. (Quine, 1960: 33-34)

But—he remarks—this approach runs into the difficulty

[81] Gareth Evans objected to Dummett's idea of the numerals as providing a privileged mode of identification of natural numbers that it «cannot be generalized, and this leaves the choice of preferred or standard means simply arbitrary.» (Evans 1982: 98) I do not agree; Martin-Löf's intuitionistic theory of types can be seen as just a systematic attempt to generalize that idea to a vast variety of mathematical objects.

> that we have made no general experimental sense of a distinction between what goes into a native's learning to apply an expression and what goes into a his learning supplementary matters about the objects concerned. (Quine, 1960: 34)

Quine's example is the famous one about the rabbit-fly, in which the natives give their assent to the utterance "Gavagai" both when they see a rabbit and when they see, in particular circumstances, a fly whose presence they know to be regularly associated with the presence of a rabbit; in a situation of this sort we are not capable to answer in a principled way the question whether this happens because "gavagai" means rabbit, for the natives, and they have access to the piece of collateral information that wherever there is a rabbit-fly there is a rabbit, or because "gavagai" means for them something as "rabbit-or-rabbit-fly".

It seems to me that this example does not warrant Quine's conclusion that the distinction between knowledge of meaning and knowledge of collateral information is *illusory*. At most it warrants the conclusion that the distiction is *difficult,* or even *very* difficult to draw; but of course a distinction may be very difficult to draw without being illusory at all. For Quine's conclusion to be legitimate his example should suggest some clear reason for the *impossibility* of the distinction; but it does not. On the contrary, it is very natural to see it as a simple case of *insufficiency* of the empirical data to settle the question. After all, natives eat rabbits, while presumably they don't eat rabbit flies; therefore, things that are rabbit-or rabbit flies are not *always* eaten; and it seems not too difficult to devise experimental situations which permit to discriminate the two possible meanings of «gavagai»; for example, it is plausible that, when a hunter announces to his wife that he has caught a gavagai, his wife's reaction is of enthusiasm if "gavagai" means rabbit, but only of suspense if it means rabbit-or-rabbit-fly and before she sees what is coming out of the game bag.[82]

Davidson seems to offer a reason for the inextricability thesis. It is the fact that, if we want to know, for example, whether someone *believes* that there is the greatest prime number, the most natural thing to do is to ask him; but for his answer to be informative we must grasp the *meaning* of his words; therefore—Davidson concludes—«in interpreting utterances from scratch—in *radical* interpretation—we must somehow deliver simultaneously a theory of belief and a theory of meaning» (Davidson, 1974: 144).

Maybe Davidson is right, but no doubt it is possible to build a theory of meaning without placing it on the background of a program of radical interpretation. The main (although not the only) goal of a theory of meaning, according to Davidson, is to enable someone who does not yet know a language to arrive at an interpretation of it. On the contrary, a theory of meaning, as I conceive it here, aims at explaining what the meaning of each expression is, and what knowledge of it consists in. An important difference between the two approaches comes to light if we assume that such an explantion is addressed to a subject who already knows the language in question. From the point of view of the second approach the theory of meaning does

---

[82] Of course this does not mean that Quine does not point out an important problem: the problem of formulating a *general* criterion for applying the distinction to particular cases. For an answer to Quine along similar lines see Peacocke (1986), 104.

not become useless, since it is perfectly possible that the subject knows *which* is the meaning of each expression of the language without knowing *what* it is and what knowledge of it consists in; whereas from the point of view of the first approach the theory of meaning (or better of interpretation) either becomes useless or the strategies elaborated by it to arrive at a correct interpretation turn out to be inefficacious in explaining understanding, i.e. knowledge of meaning. If we adopt the second approach Davidson's argument is no longer compelling: the problem of what the meaning of an expression is and of what it is to know it is logically independent of the problem of how to come to associate to each expression of an unknown language its meaning.

The inextricability thesis appears to be unacceptable because of some of its consequences too. Since there is no a priori determinable limit to the amount of collateral information (as we may continue to call it) that it is necessary to know in order to be able to recognize a justification for α, an obvious consequence of the inextricability thesis, *when it is taken in conjunction with our starting idea*, is that if a subject does not know *the whole amount* of potential collateral information, then (s)he does not know the meaning of α; and since such collateral information includes of course information about the meanings of all the other expressions of the same language, the ultimate consequence is radical holism.

What is unacceptable in radical holism? First of all that the holistic notion of knowledge of meaning is very different from the intuitive notion. For example, it seems intuitively plausible to say that someone understands the sentence "Tomorrow it will rain" even if he does not know all that is necessary to know in order to distinguish a reliable weathercast from an unreliable one. To this the holist might reply that intuition is not a good guide, and that in several cognitive domains, if we follow the only reliable guide, i.e. the scientific theories of those domains, we realize that a holistic reform of the intuitive notion of knowledge of meaning becomes necessary. For example, it is plausible that a sentence about black holes is not intelligible to someone who does not know a great deal about physical notions intuitively very remote from the one of black hole.

However, there are also not purely intuitive reasons for refusing radical holism. An interesting example has been proposed by H. Gaifman (Gaifman, 1996: 406–407). Let us consider Fermat's last theorem; although it concerns natural numbers, the proof of it given by A. Wiles contains as an essential part a proof of the Taniyama-Shimura conjecture about elliptic curves—a result belonging to the theory of real numbers. The thesis that someone who is incapable to understand Taniyama-Shimura conjecture does not understand Fermat's last theorem would be unacceptable, and this for a reason that is not purely intuitive: such a thesis would entail the impossibility to account for a fact that no theory of the foundations of mathematics would disregard, namely that the system of natural numbers has a clear conceptual priority over the system of real numbers, in the sense that the understanding of reals presupposes an understanding of natural numbers but not vice versa; it is precisely this asymmetric dependency that radical holism cannot explain.

Summing up, I do not share the neo-verificationist assumption of a priviledged way of giving concrete objects, but I agree about the theoretical necessity of distinguishing between knowledge of meaning and knowledge of collateral information. The way I tried to do it above consisted in a sharp distinction between the answer to the Application Question and the answer to the Justification Question: the former exclusively depends on our knowledge of meaning of names and predictes (as it has been analyzed in Sect. 4.6), the latter also on our knowledge of collateral information.

It should be added that refusing the neo-verificationist assumption of a priviledged way of giving concrete objects does not amount to refusing the existence of an intuitive distinction between 'direct' and 'indirect' cases of justification of sentences about such objects. Within an externalist semantics the distinction can be described as one between justifications 'in presence' and 'in absence' of the relevant objects; within the present, internalist, semantics I have characterized it in terms of the sufficiency/insufficiency of information attached to the situational epistemic content of the predicate and of the singular terms to answer the Application Question.

In conclusion, the analogies between Frege's characterization of the sense of singular terms for concrete objects, as I have suggested to interpret it in (22′), and my own characterization, which I proposed in Sect. 4.6.1.1, should be evident; here I want to stress the most substantial difference: Fregean senses are proper for an externalist semantics, mine for an internalist one. More precisely, the Fregean sense of a name (in my interpretation) is a function taking as arguments presentations of objects of the external world, whereas I propose to conceive the sense of a name as a function taking as arguments (atomic cognitive states, but more specifically) activated terms of **IRS**, which can be understood as presentations of $\mathcal{C}$-objects; and a crucial difference between objects of the external world and $\mathcal{C}$-objects is that recognition of an object of the external world is never warranted, for the reasons explained above, while a $\mathcal{C}$-object cannot be given without recognition (i.e. without the possibility for the subject to establish, by means of a computation, whether it is identical or not with any $\mathcal{C}$-object of her/his cognitive domain); as I said in Sect. 4.1.1, the notion of $\mathcal{C}$-object accounts for the subject's *seeing* something *as* a glass of water, not for the subject's seeing something that is in fact a glass of water.

4.8 Concluding Remarks

I conclude with two general remarks.

4.8.1 Testimony

The definition of justification in terms of explanation provides a computational account of the epistemic value of testimony. Consider the following case. Smith has been informed that John is Louisa's husband by a friend of his; since Smith believes that his friend is well-informed and trustworthy, he is intuitively justified to believe that

(27) John is married.

Smith's justification for (27) is obtained through an abductive inference. Smith's atomic cognitive state σ is characterized by the following facts: (i) that sc_{John} contains the piece of information that Smith's friend asserted that John is Louisa's husband; (ii) that $sc_{\text{Smith'sfriend}}$ contains the piece of information that Smith's friend is well informed and trustworthy; (iii) that the question expressed by "Why did Smith's friend assert that John is Louisa's husband in contrast to not asserting that?" arises in σ; (iv) that the topic "Smith's friend asserted that John is Louisa's husband" bears relation R to such sentences as "Smith's friend knows that John is Louisa's husband", "Smith's friend believes that John is Louisa's husband", "Smith's friend was joking", etc. The sentence "Smith's friend knows that John is Louisa's husband" belongs therefore to the class of potential answers to a question arising in σ; this fact and (ii) may trigger a further computation ending with that sentence as the best answer to the question. Since lec_{know} contains the piece of information that "$\mathcal{S}$ knows that α" entails that α,[83] Smith infers that John is Louisa's husband,[84] and then (27), by an inference using the lexical information that "x is a husband" entails "x is married". By Definition 6, $j_\sigma(\text{is married(john)}) = 1$.

4.8.2 *The Frame Problem*

A serious problem for computational approaches to mental processes is the so-called frame problem. Fodor formulates it in the following way:

> Some of the cognitive role of a thought is plausibly determined by essential (specifically, syntactic) properties of the corresponding mental representation; the effects of the logical form of a thought on its role in demonstrative inferences is para- digmatic, and Turing's story about cognition being computational works best in this kind of case. But it seems that some determinants of the role a thought plays in mental processes may not fit this paradigm; in particular, the properties of a thought that are sensitive to *which belief systems* it's embedded in don't seem to.
>
> Inferences in which features of an embedding theory affect the inferential-cum-causal roles of their constituent beliefs are what philosophers sometimes call "global" or "abductive" or "holistic" or "inferences to the best explanation." [...] What they have in common, from the

[83] "α entails β" means here that every justification for α can be transformed into a justification for β, according to the intuitionistic explanation of implication. See Chap. 5. That lexical competence about "know" includes the piece of information that "$\mathcal{S}$ knows that α" entails α is an assumption justified by the analysis of knowledge proposed in Chap. 8.

[84] That the act of inference can be seen as guided by the computation of justifications will be argued in Chap. 5.

> point of view of E(CTM),[85] is that they are presumptive examples where the determinants of the computational role of a mental representational can shift from context to context; hence where the computational role of a mental repre- sentation is not determined by its individuating properties; hence where the computational role of a mental representation is not determined by its syntax. That is: what they have in common, from the point of view of E(CTM), is that they are all presumptive counterexamples. (Fodor, 2000a: 28)

The problem is relevant to the notion of $\mathcal{S}$-justification for an atomic sentence defined above, because it makes reference to the notion of best explanation. Does such reference compromise the the possibility of conceiving an $\mathcal{S}$-justification for an atomic sentence α as the result of a computation? I think not. Although in order to answer the Justification Question it is necessary to perform inferences to the best explanation, the way explanations have been characterized above—as answers to (contrastive) why-questions—prevents such inferences from being 'global' or 'holistic'. They would be global only if the range of data to be scanned to generate the class of potential answers to the why-question arising in the given cognitive state were the totality of the data stored in the belief system(s) of the subject; but this is not the case: the range of data to be scanned is drastically limited by the why-question and by relation R, once the constraints I have suggested above are imposed on it.

Fodor elaborates on his former objection by remarking that

> There is, however, another, weaker way of reading "syntactic determination" compatible with retaining the basic idea that mental processes are computations. Consider, therefore, the [*sic*] what I'll call the Minimal Computational Theory of Mind, M(CTM):
>
> M(CTM): The role of a mental representation in cognitive processes supervenes on some syntactic facts or other. (Fodor, 2000: 29)

Then he criticizes M(CTM). Now, in a sense, the relation between "puddle" and all the other nodes of its semantic field is a syntactic relation between words: not in the strict sense of being deducible from the constituent structure of the why-question arising in σ, but in the sense that the relata ("washing", "rain", etc.) are possible inputs of a computing algorithm; hence my account of the computational answer to the Justification Question is an example of M(CTM), since it assigns to puddles a role that supervenes on the syntactic fact that "puddle" is connected to other words in a semantic field. It is therefore necessary to examine Fodor's objections to M(CTM):

> The effects that global features of belief systems appear to have on cognitive processes is a problem for the Classical computational account of mental architecture — that remains true *even if it's assumed that all of the global features of belief systems that have such effects are syntactic*. M(CTM) (unlike E(CTM)), allows *in principle* for abductive inferences to be computations, that is, for abductive inferences to be exhaustively syntactically driven. [...] But as far as anybody knows, Classical psychological theorizing can exploit this loophole only at the price of a ruinous holism; that is, by assuming that the units of thought are much bigger than in fact they could possibly be. (Fodor, 2000: 33)
>
> In the general case, it appears that the properties of a representation that determine its causal-cum-inferential role, though they may be exhaustively syntactic, needn't be either

[85] «I'll use E(CTM) as a name for the doctrine you get when you do read the Computational Theory of Mind as entailing principle E»; «Principle E: Only essential properties of a mental representation can determine its causal role in a mental life.» (Fodor 2000: 24).

> local or insensitive to context. As things now stand, Classical architectures know of no reliable way to recognize such properties short of exhaustive searches of the background [...] of epistemic commitments. (Fodor, 2000: 38)

Again, the objection has to do with the incompatibility between the necessity of scanning the totality of available data in order to get the best explanation and the possibility of conceiving such a scan as a feasible computation. In response I would stress two points. First, M(CTM) is not a single theory but a whole class of theories sharing the general principle quoted above, and I cannot see an *argument* to the effect that all possible computational accounts of inference to the best explanation belonging to that class require the scanning of the whole range of accessible data; my account, for one, does not: the why-question and the facts (i)–(iii) orient the search for an answer along a rather restricted path, even if other features may considerably extend the variety of possible paths.

Second, it should be observed that, within the framework of an internalist semantics, "best", in "best explanation", cannot mean "most rational": the best explanation simply is the one chosen by the subject's computational system; whether the chosen explanation is the most rational is an independent question. It is possible that the true explanation of the puddles is that a band of drunkards pissed in the street: this doesn't contradict the fact that, in **Case 4**, the hypothesis that (15) is true is the best explanation of the available data. This simply shows that in some cases the computation of the best explanation produces false results, *not* that the search for the best explanation is not a computational process. As I observed in Sect. 4.4.1, it is even possible that the subject doesn't get an answer to the Justification Question: again, this doesn't contradict the computational nature of the search for the best explanation.

It should be added that Fodor's insistence that computational accounts of inference to the best explanation require the scanning of the whole range of accessible data would be justified if "best explanation" were understood as the best of actual explanations.[86] Under that interpretation the piss explanation of the puddles, in **Case 4**, would be an actual explanation, hence it should be inserted into the class of possible explanations, in spite of the fact that John had no cue about its being a possible explanation; more generally, the actual explanations might be so heterogeneous as to make their class impossibile to be described otherwise than by simple enumeration; and this would require indeed the scanning of the whole range of accessible data. However, "best explanation" *cannot* mean the best of actual explanations, as we have seen in Chap. 3:

> Inference to the Best Explanation cannot be understood as inference to the best of the *actual* explanations. [...] The [...] most important reason why Inference to the Best Actual Explanation could not describe our inductive practices is that it would not characterize the process of inference in a way we could follow, since we can only tell whether something is an actual explanation *after* we have settled the inferential question. It does not give us what we want, which is an account of the way explanatory considerations can serve as a guide to the truth. [...]

[86] «For something to be an actual explanation, it must be (at least approximately) true.» (Lipton 2004: 57).

> The obvious solution is to distinguish actual from *potential* explanations, and to construe Inference to the Best Explanation as Inference to the Best Potential Explanation. [...] According to Inference to the Best Explanation, then, we do not infer the best actual explanation; rather we infer that the best of the *available* potential explanations is an actual explanation. (Lipton, 2004: 57-58. *Last italics mine.*)

References

Beyssade, C., & Dobrovie-Sorin, C. (2005). A syntax-based analysis of predication. In E. Georgala, & J. Howell (Eds.). *SALT XV* (pp. 44–63). Cornell University.

Bierwisch, M. (1992). From concepts to lexical items. *Cognition, 42*, 23–60.

Chomsky, N. (1980). *Rules and representations.* Columbia University Press.

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge University Press.

Davidson, D. (1974). Belief and the basis of meaning. *Synthese, 27*, 309–323. (Now in *Inquiries into Meaning and Truth* (pp. 141–154). Clarendon Press (1984)

Dummett, M. (1975). Frege. *Teorema* V, pp. 149–188. (Reprinted with slight variations, under the title "Frege's Distinction Between Sense and Reference", in Dummett (1978), pp. 116–144, Harvard University Press.

Dummett, M. (1978). *Truth and other enigmas*. Duckworth.

Dummett, M. (1981a). *Frege. Philosophy of language*. Second Edition. Duckworth. (First Edition 1973).

Dummett, M. (1981b). *The interpretation of Frege's philosophy*. Duckworth.

Dummett, M. (1991a). *Frege. Philosophy of mathematics.* Duckworth.

Dummett, M. (1991b). *The logical basis of metaphysics.* Duckworth.

Evans, G. (1982). *The varieties of reference.* Oxford University Press.

Fodor, J. (2000a). *The mind doesn't work that way.* The MIT Press.

Fodor, J. (2000b). It's all in the mind. *Times literary supplement*, 23 June, 3–4.

Gaifman, H. (1996). Is the 'bottom-up' approach from the theory of meaning to metaphysics possible? *The Journal of Philosophy, XCII, I*(8), 373–407.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis, 23*, 121–123.

Grandy, R. E. (2016). Sortals. *Stanford Encyclopedia of Philosophy*.

Guarino, N., Carrara, M., & Giaretta, P. (1994). An ontology of meta-level categories. In J. Doyle, E. Sandewall, & P. Torasso (Eds.), *Principles of knowledge representation and reasoning: Proceedings of the fourth international conference* (KR '94) (pp. 270–280). Morgan Kaufmann.

Hauser, M., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Blackwell.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*, 175–219.

Kitcher, P., & Salmon, W. (1987). Van Fraassen on explanation. *The Journal of Philosophy, 84*(6), 315–330.

Lewis, D. (1979). Attitudes *De Dicto* and *De Se*. *The Philosophical Review, 88*, 513–543.

Lipton, P. (2004). *Inference to the best explanation* . Second Edition. Routledge. (First Edition 1991).

Marr, D. (1982). *Vision.* Freeman & Company.

Martin-Löf, P. (1985). On the meaning and justification of logical laws. In C. Bernardi, & P. Pagli (Eds.), *Atti degli incontri di logica matematica* (vol. 2, pp. 291–340). Università di Siena. (Reprinted in *Nordic Journal of Philosophical Logic, 1*(1), 1996, 11–60.)

Martin-Löf, P. (2001). *The sense/reference distinction in constructive semantics.* Transcription of a lecture read at a conference on Frege organised by G. Sundholm at Leiden, 25 August 2001.

McGilvray, J. (1998). Meanings are syntactically individuated and found in the head. *Mind & Language, 13*(2), 225–280.

McNamara, T. P. (2005). *Semantic priming*. Psychology Press, Taylor & Francis Group.

Millson, J., & Straßer, C. (2019). A logic for best explanations. *Journal of Applied Non-Classical Logics, 29*(2), 184–231.

Moschovakis, Y. (1993). Sense and denotation as algorithm and value. In J. Väänänen, & J. Oikkonen (Eds.), *Logic colloquium '90* (vol. 2, pp. 210–249). Springer.

Peacocke, C. (1986). *Thoughts: An essay on content*. Blackwell.

Prawitz, D. (1994a). Meaning and experience. *Synthese, 98*, 131–141.

Prawitz, D. (1994b). Quine and verificationism. *Inquiry, 37*(1995), 487–494.

Quine, W. V. O. (1960). *Word and object*. The MIT Press (New edition 2013).

Scott, D. (1970). Advice on modal logic. In K. Lambert (Ed.), *Philosophical problems in logic* (pp. 143–173). Reidel.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14*, 29–56.

Sperber, D., & Wilson D. (1995). *Relevance: Communication and cognition.* Second Edition. Blackwell. (First Edition 1986).

Stovall, P. (2019). Proof-theoretic semantics and the interpretation of atomic sentences. *Logica yearbook 2019* (pp. 163–177). College publications.

Tichy, P. (1969). Intension in terms of turing machines. *Studia Logica, 24*, 7–25.

Ullman, S. (1984). Visual routines. *Cognition, 18*, 97–159.

Ullman, S. (1996). *High level vision.* The MIT Press.

van Fraassen, B. (1980). *The scientific image.* Oxford University Press.

Chapter 5
$\mathcal{C}$-Justifications for Logically Complex Sentences

Abstract In this chapter the notion of $\mathcal{C}$-justification for logically complex sentences of $\mathcal{L}$ is defined; the definition is inspired to Heyting's inductive definition of the notion of proof of a mathematical sentence α, once that definition is conveniently reinterpreted. In Sect. 5.1 it is argued for the necessity of a reinterpretation; in Sect. 5.2 the notion of evidential factor for α is characterized, and a notion of empirical negation is introduced; in Sect. 5.3 the definition proper is given; Sect. 5.4 contains a characterization of the logical concepts as they are denoted by the logical constants and a definition of $\mathcal{S}$-validity and logical validity; in Sect. 5.5 it is shown that possession of an $\mathcal{S}$-justification for an arbitrary sentence is epistemically transparent by outlining a 'non-objectual' model of evidence possession; as a consequence a way of justifying inference alternative to the neo-verificationist one is suggested.

Keywords BHK-explanation · Justificationist semantics · Evidential factors · Emprirical negation · Justification · Logical concepts · Constructive validity · Intuitionism · Inference · Intuitionism · Theory of grounds · Prawitz · Transparency · Heyting

In this chapter I will be concerned with defining the notion of $\mathcal{C}$-justification for logically complex sentences of $\mathcal{L}$; the definition is inspired to Heyting's inductive definition of the notion of proof of a mathematical sentence α, once that definition is conveniently reinterpreted. In Sect. 5.1 I shall argue for the necessity of a reinterpretation; in Sect. 5.2 the notion of evidential factor for α is characterized, and a notion of empirical negation is introduced; in Sect. 5.3 the definition proper is given; Sect. 5.4 contains a characterization of the logical concepts as they are denoted by the logical constants and a definition of $\mathcal{S}$-validity and logical validity; in Sect. 5.5 it is shown that possession of an $\mathcal{S}$-justification for an arbitrary sentence is epistemically transparent by outlining a 'non-objectual' model of evidence possession; as a consequence a way of justifying inference alternative to the neo-verificationist one is suggested.

G. Usberti, *Meaning and Justification. An Internalist Theory of Meaning*,
Logic, Epistemology, and the Unity of Science 59,
https://doi.org/10.1007/978-3-031-24605-0_5

5.1 Reasons for a Reinterpretation

We have seen in Chap. 3 how the adoption of a defeasible key notion requires that justifications are conceived as cognitive states. In this section I argue that there are reasons to conceive also mathematical proofs as cognitive states, if Heyting's explanation of the meaning of negation is taken seriously; this will be the rationale for a reinterpretation of Heyting's explanation.

In Chap. 2 we have pointed out a discrepancy between Heyting's Explanation and the Proof Explanation of the meaning of $\neg\alpha$.[1] According to the latter «a proof of $\neg\alpha$ is a construction which transforms any hypothetical proof of α into a proof of a contradiction» (Troelstra-van Dalen, 1988: 9), whereas the former requires that a proof of $\neg\alpha$ transforms any hypothetical proof of α into a contradiction. The difference is conspicuous: proofs of contradictions do not exist, according to the Proof Explanation, while contradictions do exist. The rationale for the clause of the Proof Explanation is the fact that $\neg\alpha$ is defined as an abbreviation of $\alpha \to \bot$ and that the clause for implication requires that a proof of $\alpha \to \beta$ is a construction which permits us to transform any proof of α into a proof of β. On the other hand, Heyting's formulation appears not only in his early writings, but in all later ones (e.g. Heyting, 1974: 82; Heyting, 1979: 851).

If we stick at Heyting's Explanation and we identify contradictions with proofs of $\bot$, Heyting's explanation does not satisfy the principle that

(1) Proofs of contradictions do not exist.

It might be tempting to give up (1); but it would be a mistake, from the intuitionist point of view: if proofs of contradictions exist, it could not be excluded that, for some formula α, there is both a proof of α and a function transforming every proof of α into a a proof of $\bot$, i.e. a proof of $\alpha \to \bot$, namely a proof of $\neg\alpha$; if we equate the falsity of a formula with the truth of its negation, we would have that α is both true and false, i.e. is a dialetheia. Now, it has been convincingly argued by many that the language of a logic admitting dialetheias cannot contain an *exclusive* negation, i.e. a negation excluding that both a sentence and its negation are true.[2] Therefore, if we want that in our semantics intuitionistic negation preserves the intuitive exclusive meaning of negation, we must stick to principle (1). Concluding: since contradictions do exist and proofs of contradictions do not, they are different entities. By "contradiction" I do not mean here a linguistic entity such as the sentence "$\bot$" or the sentence "$1 = 2$", but what is meant by such a sentence; and the best characterization of what is meant seems to me contained in the following passage, in which Brouwer responds to an imaginary logician who claims that contradiction is a 'logical figure':

> To this we can reply: 'The words of your mathematical demonstration merely accompany a mathematical *construction* that is effected without words. At the point where you enounce the contradiction, I simply perceive that the construction no longer *goes*, that the required structure cannot be imbedded in the given basic structure. (Brouwer, 1907: 73)

[1] See Chap. 2, (17) and (18).

[2] Cp. Batens (1990), Sect. 4; Quine (1986), 81.

In other terms, a contradiction, in Brouwer's sense, is simply the perception, or the observation, of an impossibility. I will articulate this intuitionin the following section.

5.2 Evidential Factors

Let us come back to the idea, expounded in Chap. 3, that justifications are *cognitive states* in which sentences are evident; from this point of view, a question that naturally arises is: given a cognitive state σ and a sentence α, what does the evidence of α in σ depend on? As a matter of fact, we have already been confronted with this question in Chap. 4, in relation with atomic sentences. The central idea was that a (linguistic) atomic state σ is a mental state in which a knowing subject $\mathcal{S}$ (i.e. a subject endowed with certain computational capacities) knows the *evidential factors* of a given atomic sentence α, i.e. the data and algorithms on which $\mathcal{S}$'s answer (1 or 0) to the Justification Question for α is based; and that when α is evident for $\mathcal{S}$, $\mathcal{S}$'s answer is 1, and σ is an $\mathcal{S}$-justification for α. Now, if we keep in mind that ⊥ is a particular atomic sentence, we may apply the same idea to it: the evidential factors of ⊥ are observations of impossibilities (i.e., contradictions in the sense of Brouwer); when ⊥ is evident for S, $\mathcal{S}$'s answer to the Justification Question for ⊥ is 1, and σ is an $\mathcal{S}$-justification for ⊥. The difference between ⊥ and other atomic sentences is that evidential factors for the latter exist in all cognitive states, while evidential factors for ⊥ exist only in some; and that atomic sentences are evident in some cognitive states, while ⊥ is evident in no cognitive state.

It seems to me that, at this point, it becomes possible to reinterpret Heyting's Explanation as defining not the (evidential) *proofs*, but the evidential *factors* of a logically complex sentence α in an arbitrary cognitive state σ, in such a way that, when certain conditions to be specified are met, α is evident σ; and that a proof of α is such a cognitive state σ. For instance, an evidential factor of β∧γ will be a pair <b,c> such that b is an evidential factor of β and c is an evidential factor of γ; an evidential factor of β → γ is a function *f* such that, if b is any evidential factor of β, *f*(β) is an evidential factor of γ; an evidential factor for ⊥ is the observation of an impossibility.

Reinterpreting Heyting's definition in this way would entail some discrepancies with respect to Heyting's original definition, concerning in particular the clauses for ⊥, ∨ and ∃, and the clause for atomic sentences, which is absent from Heyting's own definition. Because of this I present the definition I shall give of the set $EF_{\alpha,\sigma}$ of the evidential factors of α in a cognitive state σ as an autonomous proposal inspired by Heyting's explanation rather than as an extension of Heyting's very explanation to empirical sentences. In any case, conceiving proofs as cognitive states permits to tackle the problems of a theory of meaning of mathematical and empirical sentences from a unified point of view, in the sense that justifications of empirical sentences and proofs of mathematical sentences belong now to the same category; from this point of view, proofs are 'mathematical justifications' and, equivalently, justifications are 'empirical proofs'. I have introduced the notion of cognitive state in Chap. 3, and in

Chap. 4. I have defined it in relation to atomic sentences; now I shall try to extend it to logically complex sentences. However, before giving the definition it is necessary to discuss a question concerning negation.

5.2.1 Empirical Negation

There are important classes of atomic empirical sentences whose negations cannot be conceived as intuitionistic. Let us come back to **Case 1** described in Chap. 4: when Mary is sitting in position p_1, and sees the disk as round, she has, besides an intuitive justification for the statement.

(2) That disk is round,

also an intuitive justification for

(3) That disk is not elliptical;

what does this justification consist in? Suppose it is a function *f* that associates to each justification for

(4) That disk is elliptical

a contradiction, as a Heyting-style explanation would require; then, when Mary moves to the position p_2 (from which she see that disk as elliptical), she would have no reason to be uncertain about the shape of the disk, since *f* would enable her to associate a contradiction to *every* potential justification for (4), among which there is the visual experience from position p_2. But this is not what happens in fact: Mary hesitates and tries to acquire new relevant information—a clear indication that the two visual experiences are for her on a par as potential justifications for (2) and for (4), respectively.

I take this case as a clear indication that the intuitionistic explanation of negation must be abandoned in the case of atomic empirical statements of this sort. The question therefore arises of their evidential factors. An interesting suggestion is contained in the following passage from Nelson (1959):

> [I]t might be maintained that every significant observation must be an observation of some property, and further that the absence of a property P if it may be established empirically at all, must be established by the observation of (another) property N which is taken as a token for the absence of P. (Nelson, 1959: 208)

In our example, the token for the absence of ELLIPTICAL could be the presence of ROUND; and a way to substantiate this suggestion would be to define an evidential factor for (3) as a pair $<a_1,a_2>$, where a_1 is an evidential factor for (2) and a_2 an evidential factor for something like “ELLIPTICAL and ROUND are incompatible”. The problem with this solution is that the choice of a_1 as the first component of the pair is not more motivated than the choice of an evidential factor for any other sentence of the form “That disk is P”, where P is a property incompatible with being

ELLIPTICAL. Of course, given a specific evidential factor for "That disk is P", it would be natural to choose just it as the first component of the pair; but how should the first component be defined in general? Since there seems to be no way of regimenting the class of properties incompatible with ELLIPTICAL, and since no effective rule for choosing P can apparently be given, the adoption of such a definition of the first component would entail the loss of the epistemic transparency of the relation "x is an evidential factor for the negation of α", even when α is atomic.

The solution I propose is to introduce a symbol " ~ " for the negation of an atomic sentence $\pi(\nu_1,\ldots,\nu_n)$ of this kind, and to define the evidential factors for ~ $\pi(\nu_1,\ldots,\nu_n)$ as the ones that determine the answer 0 to the Justification Question for $\pi(\nu_1,\ldots,\nu_n)$.[3] What about the negations of logically complex sentences? Once the intuitionistic strategy has been abandoned and the idea has been accepted that the negation of an empirical sentence is in many cases an operation deeply different from implication, a natural proposal is to define the notion of evidential factor for negative sentences by a recursion on their logical complexity parallel to the one for positive sentences.

5.2.2 *Evidential Factors Defined*

I shall unify the two recursions into a unified definition:

Definition 1 Let σ be an arbitrary cognitive state such that $[\![\pi]\!]_\sigma=f_C$, $[\![\nu_1]\!]_\sigma=o_1,\ldots,$ $[\![\nu_n]\!]_\sigma=o_n$,; then the set $EF_{\alpha,\sigma}$ of the *Evidential Factors of α in σ* is defined by the following induction on α:

1.	(p)	$EF_{\pi(\nu_1,\ldots,\nu_n),\sigma}=\{f_c(o_1...o_n)\}\cup\mathcal{R}=\{<Q, \pi(\nu_1,\ldots,\nu_n)>\mid Q$ is a why-question arising in σ and $f_c(o_1...o_n)=_\sigma 1$ is a potential answer to Q};
	(n)	$EF_{\sim\pi(\nu_1,\ldots,\nu_k),\sigma}=\{f_c(o_1...o_n)\}\cup\mathcal{R}=\{<Q, \pi(\nu_1,\ldots,\nu_n)>\mid Q$ is a why-question arising in σ and $f_c(o_1...o_n)=_\sigma 0$ is a potential answer to Q};
2.	$EF_{\perp,\sigma}=\{*\mid *$ is a contradiction in σ};	
3.	(p)	$EF_{\beta\wedge\gamma,\sigma}=\{<b,c>\mid b\in EF_{\beta,\sigma}$ and $c\in EF_{\gamma,\sigma}\}$;
	(n)	$EF_{\sim(\beta\wedge\gamma),\sigma}=EF_{\sim\beta\vee\sim\gamma,\sigma}$
4.	(p)	$EF_{\beta\vee\gamma,\sigma}=\{p\mid p$ is a procedure whose execution yields an a such that, for every cognitive state $\sigma'\geq\sigma$, either $a\in EF_{\beta,\sigma'}$ or $a\in EF_{\gamma,\sigma'}\}$;
	(n)	$EF_{\sim(\beta\vee\gamma),\sigma}=EF_{\sim\beta\wedge\sim\gamma,\sigma}$
5.	(p)	$EF_{\beta\rightarrow\gamma,\sigma}=\{f\mid f$ is a function such that, for every cognitive state σ' and for every $b\in EF_{\beta,\sigma'}$, $f(b)\in EF_{\gamma,\sigma'}\}$;
	(n)	$EF_{\sim(\beta\rightarrow\gamma),\sigma}=EF_{\beta\wedge\sim\gamma,\sigma}$

(continued)

[3] See Chap. 4, Definition 10.

(continued)

6.	(p)	$EF_{\forall x\beta,\sigma}$={*g*|*g* is a function such that, for every cognitive state σ' and for every d∈$\mathcal{D}_{\sigma'}$, *g*(d)∈$EF_{\beta[\underline{d}/x],\sigma'}$};
	(n)	$EF_{\sim\forall x\beta,\sigma}=EF_{\exists x\sim\beta,\sigma}$
7.	(p)	$EF_{\exists x\beta,\sigma}$={ q | q is a procedure whose execution yields a pair <d,b> such that, for every cognitive state
		σ'≥σ, d∈$\mathcal{D}_{\sigma'}$ and b∈$EF_{\beta[\underline{d}/x],\sigma'}$};
	(n)	$EF_{\sim\exists x\beta,\sigma}=EF_{\forall x\sim\beta,\sigma}$

Some remarks about this definition.

Clause 1. (p) coincides with Definition 6 of Chap. 4.

Clauses 4. (p) and 7. (p) are a consequence of the remark, made in Chap. 2, Sect. 2.3.2, that, even in the mathematical domain, clauses (H∨) and (H∃) of Heyting's Explanation[4] seem to be too restrictive. Having argued against the neo-verificationist canonical/non-canonical distinction,[5] I have incorporated into 4.(p) and 7.(p) the amendments to Heyting's Explanation suggested by Dummett.[6] Notice that the requirement that the execution of the procedure yields the desired outcome in *every* (and not simply in some) cognitive state σ' ≥ σ is a consequence of the following intuitive facts: (i) that the computing subject knows already in σ that the procedure has that property; (ii) that the outcome of the execution exclusively depends on how the procedure is defined.

The negative clauses are supported by the observation that, in empirical contexts, the most natural way of justifying the negation of a sentence is to exhibit a counterexample to that sentence. For example, the most natural way to justify "Not all men are good" is to exhibit a bad man; by the way, to bring to contradiction the assumption that all men are good would be much more complicated, in spite of the fact that the justified proposition would be, in a sense, weaker. Analogously, the most natural way to justify "It is not the case that John is away and Mary is at home" is to justify either "John is not away" or "Mary is not at home"; and so on.

Negation defined in this way turns out to coincide with Nelson's strong negation (Nelson, 1949); as the negative clauses make clear, ~ does not denote an autonomous logical concept, but a (constructive) concept of falsity inductively defined in terms of the logical concepts.

To sum up, my suggestion is to distinguish two kinds of negation: intuitionistic negation "¬", to be applied to mathematical and to certain classes of empirical

[4] See Chap. 2, (20).

[5] Which was intended to remedy this defect of Heyting's definition, besides others. See Chap. 2, Sect. 2.4.2 and fn. 81.

[6] See Chap. 2, fn. 80. The difference between (H ∨) and the amendment suggested by Dummett can be expressed by introducing the operator $\vdash_t$, to be read as "It is proved at time t that". If we adopt (H ∨) we have that $\forall t\ (\vdash_t \alpha \vee \beta \Rightarrow \vdash_t \alpha \vee \vdash_t \beta)$; if we adopt clause 4.(p) we only have that $\forall t\ (\vdash_t \alpha \vee \beta \Rightarrow \exists\, v(t \leq v \wedge \vdash_v \alpha \vee \vdash_v \beta))$. Until explicit reference to time in mathematics is admitted, the difference is only virtual.

statements, and strong negation " ~ ", to be applied to the vast majority of empirical statements. It may be a question of discussion which negation is involved in a given natural language statement, but in most cases a motivated choice is possible. For example consider the following sentence[7]:

(5) Not all prehistoric men were black-eyed;

probably we will never be able to say, of a specific prehistoric man, that he was not black-eyed, as the explanation of " ~ " would require; at the same time, it is quite plausible to say that we have justifications to believe that (5) is true; and if we reflect on the nature of these justifications, we realize that each of them can be verbalized as a *reductio ad absurdum* of the assumption that all prehistoric men were black-eyed, as the intuitionistic explanation of negation requires. Hence (5) seems to be an example of intuitionistic negation of an empirical sentence. On the other hand, there are cases of negations concerning mathematics that seem to be examples of strong negation. It may be interesting to observe that Heyting himself recognizes the existence (and the necessity) of a "negation of ordinary speech" in the (empirical) metalanguage in which we speak about intuitionistic mathematics:

> the *principle of excluded middle* is not valid. [...] It would be false to pretend that the principle of excluded middle is false. The difference between "not valid" and "false" can be clarified as follows. By "false" we mean the mathematical negation based on contradiction [...]; by "not valid" we mean the negation of ordinary speech, which does not at all imply a contradiction. In mathematical statements only the former occurs, but in statements about mathematics the latter cannot be avoided. The mathematical statement "The principle of excluded middle is false" is false, but the non-mathematical statement "The principle of excluded middle has never been proved" is true. (Heyting, 1958: 108)[8]

5.3 Justifications for Logically Complex Sentences

Knowledge of an evidential factor a of α makes α evident, provided that certain *evidence conditions* are met, specified by the following definition:

Definition 2 Let $\sigma \in \Sigma$ and $a \in EF_{\alpha,\sigma}$; then the relation "a makes α evident in σ for $\mathcal{S}$" (in symbols, $a \vDash^{\mathcal{S}}_{\sigma} \alpha$)[9] is defined by the following induction:

1. (p) $a \vDash_{\sigma} \pi(\nu_1,...,\nu_n)$ iff either $a = f_c(o_1...o_n)$ and $f_c(o_1...o_n) =_{\sigma} 1$, or $a \in \mathcal{R}$ and there is a question Q such that $f_c(o_1...o_n) =_{\sigma} 1$ is the best answer to Q in σ;
   (n) $a \vDash_{\sigma} \sim\pi(\nu_1,...,\nu_n)$ iff either $a = f_c(o_1...o_n)$ and $f_c(o_1...o_n) =_{\sigma} 0$, or $a \in \mathcal{R}$ and there is a question Q such that $f_c(o_1...o_n) =_{\sigma} 0$ is the best answer to Q in σ.

---

[7] I owe this example to Paolo Casalegno.

[8] See also Heyting (1956b): 232. It may be wondered whether Heyting's "negation of ordinary speech" is to be identified with strong negation or with bivalent negation; since «we may reasonably view it as decidable whether or not a given statement has been proved at a given time» (Dummett, 2000: 235), the difference seems to be immaterial.

[9] I will omit the superscript "$\mathcal{S}$" when it is not strictly necessary.

2. not $*\vDash_\sigma\bot$.
3. $<b,c>\vDash_\sigma\beta\wedge\gamma$ iff $b\vDash_\sigma\beta$ and $c\vDash_\sigma\gamma$.
4. $p\vDash_\sigma\beta\vee\gamma$ iff there is an a such that the execution of p yields a and for every cognitive state $\sigma'\geq\sigma$ either $a\vDash_{\sigma'}\beta$ or $a\vDash_{\sigma'}\gamma$.
5. $f\vDash_\sigma\beta\rightarrow\gamma$ iff, for any b such that $b\vDash_\sigma\beta$, $f(b)\vDash_\sigma\gamma$.
6. $g\vDash_\sigma\forall x\beta$ iff, for any $d\in\mathcal{D}_\sigma$, $g(d)\vDash_\sigma\beta[\underline{d}/x]$.
7. $q\vDash_\sigma\exists x\beta$ iff there is pair $<d,b>$ such that the execution of p yields $<d,b>$ and, for every cognitive state $\sigma'\geq\sigma$, $d\in\mathcal{D}_{\sigma'}$ and $b\vDash_{\sigma'}\beta[\underline{d}/x]$.

Definition 3 $J_\sigma(\alpha) = 1$ iff an $a \in EF_{\alpha,\sigma}$ is known such that $a\vDash_\sigma\alpha$.

Definition 4 An $\mathcal{S}$*-justification* for α is cognitive state $\sigma \in \Sigma$ such that $j_\sigma(\alpha) = 1$.

Some remarks concerning these definitions.

1. An immediate consequence of definitions 5.1 and 5.2 is that if $a\in EF_{\pi(\nu 1,\ldots,\nu n),\sigma}$, then $a\vDash_\sigma \sim\sim\pi(\nu_1,\ldots,\nu_n)$ iff $a\vDash_\sigma\pi(\nu_1,\ldots,\nu_n)$; an easy induction shows that this holds for all α.

2. Having an evidential factor does not amount to being evident: a may be an evidential factor of α in σ without α being evident in σ; what evidence of α in σ amounts to is the existence of the relation "a makes evident α in σ".

3. The distinction between evidential factors of α and justifications for α, neglected by Heyting's explanation, allows to account for a significant phenomenon: the fact that our mind can adopt different attitudes towards one and the same entity, because of the fact that it occupies different cognitive states. We have seen two important examples, one in Chap. 4, concerning empirical sentences, the other in the present chapter, concerning sentences of any kind.

Consider first the empirical case. John is looking at a stick half-dipped into a glass of water; in the cognitive state σ_1 he has a (visual) presentation p_1 of a crooked stick and he, justifiedly, believes that the stick is crooked; later on, in the cognitive state σ_2, John touches the stick, has a (tactile) presentation p_2 of a straight stick, and he convinces himself that

(6) The stick is straight.

In σ_2 there are *two* evidential factors of (6), but only one makes (6) evident; within the theoretical framework introduced in Chap. 4, what makes (6) evident is the very choice of (6) as the best answer to the why-question arising in σ_2. In conclusion, in σ_1 and σ_2 John has two different attitudes towards p_1.

The second example is ⊥. Consider a sentence α and a cognitive state σ_1 in which Mary assumes α: in σ_1 there is an $a \in EF_{\alpha,\sigma}$ such that $a\vDash_\sigma\alpha$. Now suppose that in a subsequent state σ_2 Mary derives a contradiction from α, i.e. observes an impossibility in σ_1: by Definition 5.2, clause 2., in σ_2 it does not hold that $a\vDash_\sigma\alpha$. In conclusion, in σ_1 and σ_2 Mary has two different attitudes towards a. As for clause 2. (or, equivalently, (1)) itself, it seems to have the status of a law of thought: it seems to be a constitutive principle of thinking that a contradiction can be observed, but that a sentence expressing it cannot be evident.

4. An obvious analogy between the semantics set out in this chapter and Kripke's semantics for intuitionistic logic lies in the fact that worlds are understood as points in time, or 'evidential situations' at which, intuitively, the subject has information to verify (or prove) sentences. However, Definitions 5.1 and 5.2 point out a significant difference; consider, for instance, implication: Kripke's semantics requires that «to assert A → B in a situation H, we need to know that in any later situation H' where we get a proof of A, we also get a proof of B» (Kripke, 1965: 99), but does not specify what enables the subject to obtain that knowledge; Definitions 5.1 and 5.2 answer this question: what enables the subject to obtain that knowledge are the evidential factors of implication, together with the relation of making evident. The reference to evidential factors will be the essential ingredient of the solution I will propose, in Chap. 8, for Gettier's problems.

5.4 Logical Concepts and Logical Validity

The logical concepts are $\mathcal{C}$-concepts, hence the internalist analysis proposed in Chap. 4, according to which $\mathcal{C}$-concepts are programs for the computation of discrimination functions, applies to logical concepts as well. In particular, I distinguish between prelinguistic and linguistic logical $\mathcal{C}$-concepts: in the case of the former the individuation of each function is of competence of the psychologist, not of the subject, while in the case of the latter the possibility arises that a prelinguistic $\mathcal{C}$-concept is denoted by an operator, hence that the subject himself is in a position to consciously manage it. Take for example the case of conjunction: it is possible that a non-human subject $\mathcal{S}$ computes it, in the sense that $\mathcal{S}$ follows a routine consisting, as a matter of fact, in discriminating between cognitive states in which $\beta \wedge \gamma$ is evident and others in which it is not, without having any capacity to individuate the function, i.e. the goal of the routine; in such a case it would of the psychologist's (or the ethologist's) competence to say that $\mathcal{S}$ manages the concept of conjunction, but $\mathcal{S}$ could not recognize it, since it has not a linguistic concept of conjunction, i.e. no name for the goal of the routine it is following. Quite different is the situation of a human subject, whose language contains the logical constants; in this case two conditions are satisfied: (i) the subject is in a position to individuate, by means of the logical constant, the function computed by the routine (s)he is following, the goal of the routine; (ii) the subject is in a position to recognize the results of the application of the algorithms (s)he is following, i.e. to give an explicit answer to the justification question for conjunction.

5.4.1 The Logical Concepts

I am not concerned with the logical concepts of non-human subjects; I only have mentioned them because it seems to me interesting to leave open the possibility of discrepancies between the logical concepts of a non-linguistic mind and the ones of a mind endowed with a language faculty—a possibility that is guaranteed only within the framework of a theory of meaning in which concepts are not simply the denotations of predicates. However, my actual concern is with the logical concepts as they are denoted by the logical constants. The following seems to be an adequate characterization of the usual logical concepts, if conditions (i) and (ii) are to be met.

Definition 5 The *logical $\mathcal{C}$-concepts* are:

1. The concept of *conceptual application*, i.e. the function f_{app} such that, if $a \in EF_{\pi(\nu_1,\ldots,\nu_k),\sigma}$, $f_{\text{app}}(a,\sigma)=1$ iff $a \vDash_\sigma \pi(\nu_1,\ldots,\nu_k)$.
2. The concept of *contradiction*, i.e. the function $f_\perp$ such that, if $a \in EF_{\perp,\sigma}$, $f_\perp(a,\sigma)=1$ iff $a \vDash_\sigma \perp$.
3. The concept of *conjunction*, i.e. the function $f_\wedge$ such that, if $a \in EF_{\beta\wedge\gamma,\sigma}$ $f_\wedge(a,\sigma)=1$ iff $a \vDash_\sigma \beta\wedge\gamma$;
4. The concept of *disjunction*, i.e. the function $f_\vee$ such that, if $a \in EF_{\beta\vee\gamma,\sigma}$, $f_\vee(a,\sigma)=1$ iff $a \vDash_\sigma \beta\vee\gamma$.
5. The concept of *implication*, i.e. the function $f_\rightarrow$ such that, if $a \in EF_{\beta\rightarrow\gamma,\sigma}$, $f_\rightarrow(a,\sigma)=1$ iff $a \vDash_\sigma \beta\rightarrow\gamma$.
6. The concept of *universal quantification*, i.e. the function $f_\forall$ such that, if $a \in EF_{\forall x\beta,\sigma}$, $f_\forall(a,\sigma)=1$ iff $a \vDash_\sigma \forall x\beta$.
7. The concept of *existential quantification*, i.e. the function $f_\exists$ such that, if $a \in EF_{\exists x\beta,\sigma}$, $f_\exists(a,\sigma)=1$ iff $a \vDash_\sigma \exists x\beta$.

Definition 6 A *logically competent subject* is a (knowing) *subject* $\mathcal{S} = <\sum, \mathcal{CA}, [\![]\!]>$ such that

- $\mathcal{S}$ manages the logical $\mathcal{C}$-concepts;
- the concept of conceptual application is denoted for $\mathcal{S}$ by syntactical concatenation,[10] and $[\![\perp]\!]_\mathcal{S} = f_\perp$, $[\![\wedge]\!]_\mathcal{S} = f_\wedge$, $[\![\vee]\!]_\mathcal{S} = f_\vee$, $[\![\rightarrow]\!]_\mathcal{S} = f_\rightarrow$, $[\![\forall]\!]_\mathcal{S} = f_\forall$, $[\![\exists]\!]_\mathcal{S} = f_\exists$.

Definition 7 A *cognitive state* is an atomic cognitive state of a logically competent subject.

The definitions incorporate two assumptions. The first is that logical competence is invariant across the cognitive states of a subject, hence that it is a constitutive characteristics of subjects rather than of their cognitive states; it is motivated by the hypothesis that, while many $\mathcal{C}$-concepts are learned, logical $\mathcal{C}$-concepts are innate; in

[10] As I said in the Introduction, I am assuming that $\mathcal{L}$ is a good ‘skeleton’ of a natural language.

its turn, this hypothesis is corroborated by what is known about the logical competences of other species. A consequence of this first assumption is that every I-language contains the logical constants, i.e. symbols denoting the logical concepts (while it is possible to imagine I-languages not containing predicates for other concepts).

The second, simplifying, assumption is that each logical constant denotes, for any subject, the same logical concept. The rationale for this is that, although it is possible that in some 'italian' I-language, for example, the logical constant "e" denotes the concept of disjunction instead of conjunction, this would be a case of what Chomsky calls "Saussurean arbitrariness" (Chomsky, 2000: 120), not significant from our actual point of view.

5.4.2 *Validity and Logical Validity. Relativized Substitutivity*

Definition 8 of Chap. 4 defined the relation "σ' assigns to σ the status of $\mathcal{S}$-truth-ground of α" (in symbols $\vDash^{\mathcal{S}}_{\sigma,\sigma'}\alpha$) for α atomic. The extension to logically complex sentences is straightforward:

Definition 8[11]

$\vDash^{\mathcal{S}}_{\sigma,\sigma'}\beta\wedge\gamma$ iff $\vDash^{\mathcal{S}}_{\sigma,\sigma'}\beta$ and $\vDash^{\mathcal{S}}_{\sigma,\sigma'}\gamma$;
$\vDash^{\mathcal{S}}_{\sigma,\sigma'}\beta\vee\gamma$ iff $\vDash^{\mathcal{S}}_{\sigma,\sigma'}\beta$ or $\vDash^{\mathcal{S}}_{\sigma,\sigma'}\gamma$;
$\vDash^{\mathcal{S}}_{\sigma,\sigma'}\beta\rightarrow\gamma$ iff, if $\vDash^{\mathcal{S}}_{\sigma,\sigma'}\beta$, then $\vDash^{\mathcal{S}}_{\sigma,\sigma'}\gamma$;
$\vDash^{\mathcal{S}}_{\sigma,\sigma'}\forall x\beta$ iff, for any $d\in\mathcal{D}_\sigma$, $\vDash^{\mathcal{S}}_{\sigma,\sigma'}\beta[\underline{d}/x]$;
$\vDash^{\mathcal{S}}_{\sigma,\sigma'}\exists x\beta$ iff, for some $d\in\mathcal{D}_\sigma$, $\vDash^{\mathcal{S}}_{\sigma,\sigma'}\beta[\underline{d}/x]$.

The following definitions of $\mathcal{S}$-validity and (logical) validity should be natural in light of the analysis of empirical knowledge and the approach to Gettier problems sketched in Chap. 4:

Definition 9 The relation *σ' is better than σ with respect to α* (in symbols $\sigma' \geq_{\alpha}\sigma$) is defined through the following induction:

- $\sigma' \geq_{\pi(v1,\ldots,vn)}\sigma$ is defined like in Definition 7 of Chap. 4;
- $\sigma' \geq_{\perp}\sigma$ iff, if a contradiction is observed in σ, then it is observed in σ';
- $\sigma' \geq_{\beta\mathbf{C}\gamma}\sigma$, for any binary connective **C**, iff $\sigma' \geq_{\beta}\sigma$ and $\sigma' \geq_{\gamma}\sigma$;
- $\sigma' \geq_{Qx\beta}\sigma$, for any quantifier Q, iff $\sigma' \geq_{\beta[\tau/x]}\sigma$, for every term τ.

Definition 10 α is $\mathcal{S}$-valid (in symbols, $\vDash^{\mathcal{S}}\alpha$) iff there is a $\sigma\in\Sigma$ such that, for all $\sigma'\in\Sigma$, if $\sigma' \geq_{\alpha}\sigma$ then $\vDash^{\mathcal{S}}_{\sigma,\sigma'}\alpha$.

Definition 11 α is valid (in symbols, $\vDash\alpha$) iff, for all subjects $\mathcal{S}$, $\vDash^{\mathcal{S}}\alpha$.)

Definition 12 α is a logical consequence of Γ (in symbols, $\Gamma\vDash\alpha$) iff, for all subjects $\mathcal{S}$, if for all $\alpha\in\Gamma$ $\vDash^{\mathcal{S}}\alpha$, then $\vDash^{\mathcal{S}}\beta$.

[11] The metalinguistic logical constants are to be understood intuitionistically. There is no clause for ⊥ because, according to Definition 2, there is no justification for ⊥ .

Remark 1 (on susbstitutivity)
For every logically competent subject S = <$\sum$, $\mathcal{CA}$, [[]]>, for every sentence α, let $\mathbf{J}_{\mathcal{S}}(\alpha)$ be $\{\sigma|\sigma\in\sum$ and $[\![\alpha]\!]_\sigma=1\}$, the set of $\mathcal{S}$-justifications for α; then the following principles are valid:
(RSPST) If $\sigma\in\mathbf{J}_{\mathcal{S}}(\alpha[\tau])$ and $\sigma\in\mathbf{J}_{\mathcal{S}}(\tau=\tau')$, then $\sigma\in\mathbf{J}_{\mathcal{S}}(\alpha[\tau'//\tau])$;
(RSPP) If $\sigma\in\mathbf{J}_{\mathcal{S}}(\alpha[\pi])$ and $\sigma\in\mathbf{J}_{\mathcal{S}}(\pi\equiv\pi')$, then $\sigma\in\mathbf{J}_{\mathcal{S}}(\alpha[\pi'//\pi])$.

The proof is by induction on α; the base clauses follow from clauses 3.1.–3.4. of Definition 10 of Chap. 4.

When it is not necessary to distinguish these two principles, I shall call them collectively *the Relativized Substitutivity Principle* (RSP).

If α is a mathematical sentence, a useful criterion of $\mathcal{S}$-validity can be stated. I have said above that characterizing (evidential) proofs as cognitive states permits to conceive proofs as 'mathematical justifications' and, equivalently, justifications as 'empirical proofs'. From this point of view, the distinguishing feature of proofs as opposed to justifications, their indefeasibility, is due to the following property of the evidential factors of mathematical sentences:

(7) For every mathematical sentence α, if $a \in EF_{\alpha,\sigma}$ and $\sigma' \geq \sigma$, then $a \in EF_{\alpha,\sigma'}$.[12]

Consider for example an atomic sentence α concerning natural numbers. Its evidential factors are essentially algorithms to compute functions over natural numbers, about which one can have only information inferred from their position in the chain of numbers—information available to whoever has the capacity to count, which is assumed to be universal; hence about numbers it is not possible to have pieces of information conflicting with each other, which means that an acquired piece of information persists through time. Analogous considerations may be extended to other mathematical domains. If α is logically complex, the clauses defining the set of its evidential factors satisfy the subformula property, hence the monotonicity of the function assigning sets of evidential factors may be shown by induction.

For the same reasons, also the following property holds:

(8) For every mathematical sentence α, if $a\vDash^{\mathcal{S}}_{\sigma}\alpha$ and $\sigma' \geq \sigma$, then $a\vDash^{\mathcal{S}}_{\sigma'}\alpha$.

From (7) and (8) it follows that

(9) For every mathematical sentence α, if $j_\sigma(\alpha) = 1$ and $\sigma' \geq \sigma$, then $j_{\sigma'}(\alpha) = 1$.

If α is an empirical sentence, some of its evidential factors in σ may disappear in a subsequent state σ', or may cease to make α evident in σ'; and it may happen that, in σ', there is no other factor that makes α evident. This is exactly what (7) and (8) ensure *not* to happen with mathematical sentences.

[12] In other terms, the function *f* assigning sets of evidential factors to mathematical sentences is monotonic. The stronger property that *f* is constant does not hold. For example, an evidential factor in σ of "n + m = m + n" may be the algorithm for addition; in a subsequent cognitive state σ', in which the subject has learned some theorems of number theory, a new evidential factor of the same sentence will be a function associating to every pair <d,d' > of natural numbers an evidential factor of "$\underline{d} + \underline{d'} = \underline{d'} + \underline{d}$".

According to my analysis of Gettier problems sketched in Chap. 4, for a justification σ for $\pi(\nu_1,\ldots,\nu_n)$ to be a truth-ground of $\pi(\nu_1,\ldots,\nu_n)$ relative to σ' it is necessary that the evidential factor that makes $\pi(\nu_1,\ldots,\nu_n)$ evident in σ and σ' is the same: if it is not, it may happen that a subject justifiedly believes the proposition that $\pi(\nu_1,\ldots,\nu_n)$ in σ, that that proposition is even justified in σ', but it is not *known* in σ relative to σ'. This, however cannot happen if $\pi(\nu_1,\ldots,\nu_n)$ is a mathematical sentence, because, even if new evidential factors of it are present in σ', the old ones, present in σ, are still present in σ'. As a consequence, it is not possible that a mathematical sentence is justified and not known, as it is in the case of empirical sentences.[13] This justifies the following criterion of $\mathcal{S}$-validity for mathematical sentences:

Remark 2 If α is a mathematical sentence, α is $\mathcal{S}$-valid iff there is a $\sigma \in \Sigma$ such that $j_\sigma(\alpha) = 1$.

Which formal system is adequate to the definition of logical validity just given? I don't know. An obvious candidate for the fragment containing only mathematical sentences and not containing ~ is intuitionistic logic, owing to the analogy between evidential factors as defined above and proofs of Heyting's explanation. But there are several crucial differences. I shall conclude with some examples of valid and invalid sentences.

(10) $(\alpha\wedge(\alpha \rightarrow \bot)) \rightarrow \bot$ VALID

We must define a $g \in EF_{((\alpha\wedge(\alpha \rightarrow \bot)) \rightarrow \bot),\sigma}$ such that $g\vDash_\sigma((\alpha\wedge(\alpha \rightarrow \bot)) \rightarrow \bot)$. By definition, g must be such that, for every $a \in EF_{\alpha\wedge(\alpha \rightarrow \bot),\sigma}$, $g(a) \in EF_{\bot,\sigma}$ and, if $a\vDash_\sigma(\alpha\wedge(\alpha \rightarrow \bot))$, then $g(a)\vDash_\sigma\bot$. Since no evidential factor x can satisfy the condition that $x\vDash_\sigma\bot$, g exists (and is the null function) iff there is no $a \in EF_{\alpha\wedge(\alpha \rightarrow \bot),\sigma}$ such that $g(a)\vDash_\sigma\alpha\wedge(\alpha \rightarrow \bot)$. In fact, no such a exists: it should be pair $<a_1, a_2>$ such that $a_1\vDash_\sigma\alpha$ and $a_2\vDash_\sigma\alpha \rightarrow \bot$; hence a function h would exist such that $h(a_1) = *$ and $h(a_1)\vDash_\sigma\bot$, against condition 2. of Definition 2.

(11) $\alpha\vee\neg\alpha$: INVALID.

Proof If there is no $a \in EF_{\alpha,\sigma}$ such that $a\vDash_\sigma\alpha$, and no $a \in EF_{\neg\alpha,\sigma}$ such that $a\vDash_\sigma\neg\alpha$, then $j_\sigma(\alpha) = 0$ and $j_\sigma(\neg\alpha) = 0$, hence $j_\sigma(\alpha\vee\neg\alpha) = 0$.

(12) $\bot \rightarrow \alpha$: INVALID.

Proof The usual argument for the validity of (12) exploits the fact that there is no proof of $\bot$, so that *any* function vacuously satisfies the condition of transforming every proof of $\bot$ into a proof of α. Analogously, in the present semantics *any* function f vacuously satisfies the condition (i) if $a \in EF_{\bot,\sigma}$ and $a\vDash_\sigma\bot$ then $f(a) \in EF_{\alpha,\sigma}$ and $f(a)\vDash_\sigma\alpha$, because there is no a such that $a \in EF_{\bot,\sigma}$ and $a\vDash_\sigma\bot$. However, (i) is *not* the condition a function f must meet for (12) to be valid: the condition is that (ii) if $a \in EF_{\bot,\sigma}$, then $f(a) \in EF_{\alpha,\sigma}$, and (iii) if $a\vDash_\sigma\bot$, then $f(a)\vDash_\sigma\alpha$; and, as we have seen, for some cognitive state σ there is some a such that $a \in EF_{\bot,\sigma}$: in such a case there is no reason to assert the existence of a function f such that $f(a) \in EF_{\alpha,\sigma}$.

[13] In other terms, my sketched analysis of Gettier problems predicts that significant Gettier problems do not arise with mathematical sentences.

(13) $((\alpha\vee\beta)\wedge\neg\alpha) \to \beta$: VALID.

Proof We must define an f such that, for any a $\in EF_{((\alpha\vee\beta)\wedge\neg\alpha),\sigma}$, f(a) $\in EF_{\beta,\sigma}$, and if a$\vDash_\sigma(\alpha\vee\beta)\wedge\neg\alpha$ then f(a)$\vDash_\sigma\beta$. By definition, a is a pair <a_1, a_2> such that $a_1 \in EF_{\alpha\vee\beta,\sigma}$, $a_1\vDash_\sigma\alpha\vee\beta$, $a_2 \in EF_{\neg\alpha,\sigma}$ and $a_2\vDash_\sigma\neg\alpha$. Let us extract a_1: it is a procedure p whose execution yields an a_3 such that, for every cognitive state σ' ≥ σ, either $a_3 \in EF_{\alpha,\sigma'}$ and $a_3\vDash_{\sigma'}\alpha$ or $a_3 \in EF_{\beta,\sigma'}$ and $a_3\vDash_{\sigma'}\beta$. Suppose that a_3 satisfies the first alternative; then $a_2(a_3)\vDash_{\sigma'}\bot$; but by definition $a_2(a_3) = *$, hence we would have that $*\vDash_{\sigma'}\bot$, against condition 2. of Definition 2; hence the execution of p will yield an $a_4 \in EF_{\beta,\sigma}$ such that $a_4\vDash_{\sigma'}\beta$; put f(a) = a_4.

It is worth noting that (13), the disjunctive syllogism, is not a theorem of minimal logic.

(14) $((\alpha\vee\beta) \to \gamma) \to ((\alpha \to \gamma)\wedge(\beta \to \gamma))$: VALID

Proof Let $f \in EF_{(\alpha\vee\beta) \to \gamma,\sigma}$: and suppose that $f\vDash_\sigma(\alpha\vee\beta) \to \gamma$; by definition f transforms every p $\in EF_{(\alpha\vee\beta),\sigma}$ such that p$\vDash_\sigma\alpha\vee\beta$ into f(p) $\in EF_{\gamma,\sigma}$ such that f(p)$\vDash_\sigma\gamma$; p is a procedure whose execution yields an a such that, for every cognitive state σ' ≥ σ, either a $\in EF_{\alpha,\sigma}$, and a$\vDash_{\sigma'}\alpha$, or a $\in EF_{\beta,\sigma'}$, and a$\vDash_{\sigma'}\beta$; by hypothesis, in both cases f(p) $\in EF\gamma_{,\sigma}$ and f(p)$\vDash_\sigma\gamma$; hence f transforms all justifications for α into justifications for γ, and all justifications for β into justifications for γ, i.e. f is a justification both for α → γ and for β → γ. We can therefore define the following function g from justifications for $(\alpha\vee\beta) \to \gamma$ into justifications for $(\alpha \to \gamma)\wedge(\beta \to \gamma)$: $g(f)$ = <f_1,f_2>, where $f_1 = f_2 = f$.

(15) $((\alpha \to \gamma)\wedge(\beta \to \gamma)) \to ((\alpha\vee\beta) \to \gamma)$: VALID

Proof Let $f \in EF_{(\alpha \to \gamma)\wedge(\beta \to \gamma),\sigma}$: it is a pair <$f_1$, f_2> , where

for every a $\in EF_{\alpha,\sigma}$, f_1(a) $\in EF_{\gamma,\sigma}$ and if a$\vDash_\sigma\alpha$ then f_1(a)$\vDash_\sigma\gamma$;

for every b $\in EF_{\beta,\sigma}$, f_2(b) $\in EF_{\gamma,\sigma}$ and if b$\vDash_\sigma\alpha$ then f_2(b)$\vDash_\sigma\gamma$.

Consider an arbitrary procedure p $\in EF_{(\alpha\vee\beta),\sigma}$ whose execution yields a c such that, for every cognitive state σ' ≥ σ, either c$\in EF_{\alpha,\sigma}$, and c$\vDash_{\sigma'}\alpha$ or c$\in EF_{\beta,\sigma'}$ and c$\vDash_{\sigma'}\beta$; define a function f' as follows:

- if c$\in EF_{\alpha,\sigma}$, and c$\vDash_{\sigma'}\alpha$, then f'(p) = f_1(c).
- if c$\in EF_{\beta,\sigma}$, and c$\vDash_{\sigma'}\beta$, then f'(p) = f_2(c).

Hence f' $\in EF_{(\alpha\vee\beta) \to \gamma,\sigma}$ and f'$\vDash_\sigma(\alpha\vee\beta) \to \gamma$.

(16) $\neg\neg(\alpha\vee\neg\alpha)$, [i.e. $((\alpha\vee(\alpha \to \bot)) \to \bot) \to \bot$]: VALID.

Proof We must define a $g \in EF_{((\alpha\vee(\alpha \to \bot)) \to \bot) \to \bot,\sigma}$ such that $g\vDash_\sigma((\alpha\vee(\alpha \to \bot)) \to \bot) \to \bot$. By definition, g must be such that, for every $f \in EF_{(\alpha\vee(\alpha \to \bot)) \to \bot,\sigma}$, $g(f) \in EF_{\bot,\sigma}$ and, if $f\vDash_\sigma((\alpha\vee(\alpha \to \bot)) \to \bot)$, then $g(f)\vDash_\sigma\bot$. By definition g(f) = *, hence it is impossible that $g(f)\vDash_\sigma\bot$; hence g exists (and is the null function) iff no f satisfying the conditions specified exists. In fact, no such f exists: if one did, then $g(f)$ = *, and $g(f)\vDash_\sigma\bot$, against condition 2 of Definition 2.

5.5 Epistemic Transparency Revisited

In Chaps. 2 and 3 I have argued for the necessity of a theoretical notion of justification whose possession is epistemically transparent, and in Chap. 4 I have shown that possession of an $\mathcal{S}$-justification for an atomic sentence is epistemically transparent for $\mathcal{S}$; here I shall try to show the same for possession of an $\mathcal{S}$-justification for an arbitrary sentence.

5.5.1 A 'Non-objectual' Way of Possessing Evidence

In Chap. 2, Sect. 2.4.3.1, I remarked, at the end of my discussion of some of Prawitz's ideas about possession of intuitionistic proofs, that «the lack of epistemic transparency of the possession of an intuitionistic proof of α is—according to Prawitz—a consequence of equating such possession with knowing a term that denotes that proof», and I concluded that, therefore, «it becomes interesting to explore the possibility of a direct access to evidence, not mediated by terms or descriptions». It is interesting to observe that there is a passage of Prawitz (2015) in which Prawitz concedes, or seems to concede, that there is a transparent way of possessing evidence; he writes:

> One finds something to be evident by performing a mental act, for instance, by making an observation, a computation, or an inference. After having made such an act, one is in an epistemic state, a state of mind, where the truth of a certain sentence is evident, or as I have usually put it, one is in possession of evidence for a certain assertion. (Prawitz, 2015: 88)

From this we can extract the suggestion that having evidence for α could be analyzed as being in the *mental state* resulting from having performed the (mental) act that confers evidence to α. This suggestion is crucially different from the idea that having evidence for α amounts to having a term denoting an abstract object of some sort: what is suggested is that having evidence amounts to occupying a mental state. Prawitz obliterates this difference when, shortly after the passage just quoted, he adds:

> it is convenient to think of evidence states as states where the subject is in possession of certain objects. [...] I am so to say reifying evidence and am replacing evidence states with states where the subject is in possession of grounds. (Prawitz, 2015: 88-89)

This gives the impression that he is developing the same idea,[14] but we know that conceiving the possession of evidence as having a term denoting an abstract object entails the loss of the epistemic transparency of the possession of evidence; it is therefore worth to stress the difference and to explore the possibility of a 'non-objectual' way of conceiving evidence and its possession, in the hope that it opens the way to a transparent notion of having evidence.

[14] As we have seen in Chap. 2, there is an important tradition behind this obliteration of the difference between evidence as a mental state and evidence as an object: Curry-Howard isomorphism and the consequent idea that «the question whether a term denotes a ground for an assertion of a sentence A coincides with the question of the type of the term» (Prawitz, 2015: 89).

Why should the execution of a (mental) act give rise to a mental state in which one has evidence for a sentence? Assuming, provisionally and for the sake of the argument, that the execution of an act is an essential condition for acquiring evidence for a sentence, it is clear that it cannot be the *only* condition. Consider for instance the Italian sentence "Piove"; the mental state σ resulting from having performed the act of observing that it is raining would not make it evident to one that "Piove" is true, if at least the following two further conditions were not satisfied: (i) that, when one is in the state σ, one knows that "Piove" means that it is raining,[15] and (ii) that, when one is in σ, one knows that it is the very act of observing that it is raining that makes evident the proposition that it is raining. If one did not know that "Piove" means that it is raining, there would be no reason for one to consider evident *that very sentence*, rather than any other, after having observed that it is raining; and if one did not know that it is the very act of observing that it is raining that makes evident the proposition that it is raining, there would be no reason for one to consider that proposition evident *after having observed that it is raining*, rather than after having performed any other act. Generalizing, the problem is therefore to characterize a theoretical notion of mental state in such a way that it can simultaneously account for knowledge of the meaning of sentences and evidence of propositions.

According to my approach, the meaning of a sentence α for a subject $\mathcal{S}$ is its $\mathcal{S}$-intension, a function from $\mathcal{S}$'s cognitive states to {1,0}, thought of as answers (yes, no) of the subject's cognitive apparatus to the question "Is α evident in the present state?" (or, equivalently: "Is the present state a justification for α?"), and (i) knowing the meaning of α, or the proposition expressed for $\mathcal{S}$ by α, amounts to being capable to individuate such function, i.e. the class of its arguments, the class of its values, and the rule to obtain one value for each argument; the statement of such a rule requires in particular the specification, for every cognitive state σ, of the class of the evidential factors of α in σ. On the other hand, (ii) having evidence in σ for the proposition that α is to have effected two operations: (a) having chosen one of the evidential factors of α in σ; and (b) having verified that that factor makes α evident in σ.

This is exactly what definition 5 requires. Consider for example a sentence of the form "$\beta \wedge \gamma$". According to condition (i), one knows that that sentence expresses the proposition that $\beta \wedge \gamma$; this happens when one knows that "β" expresses the proposition that β, that "γ" expresses the proposition that γ, and one associates to "$\wedge$" the logical *$\mathcal{C}$-concept* of conjunction. In general, managing a $\mathcal{C}$-concept, whether logical or extra-logical, amounts to following an algorithm checking the presence of the appropriate feature; in the case of the $\mathcal{C}$-concept of conjunction the feature to be checked is the property of being an evidential factor of $\beta \wedge \gamma$ in σ, i.e. of being a pair <b,c> such that b is an evidential factor of β in σ and c is an evidential factor of γ in σ. Under the (inductive) hypothesis that the evidential factors of β and of γ are epistemically transparent, the feature of being such a pair is clearly epistemically transparent, for it is obviously legitimate to require that the subject is capable to recognize pairs; and

[15] This is clearly recognized by Prawitz: «It can be assumed to be part of what it is to make an inference that the agent knows the meanings of the involved sentences» (Prawitz, 2015: 96).

if β and γ are atomic, their evidential factors in σ are epistemically transparent, as I have argued in Chap. 4.

According to condition (ii), one has evidence in σ for the proposition that β∧γ when (a) one has selected one element <b,c> of $EF_{\beta\wedge\gamma,\sigma}$, and (b) one has verified that $b\vDash_\sigma\beta$ and $c\vDash_\sigma\gamma$. Selection and verification required by (a) and (b) may be difficult operations, because they require intervention of memory; but limitations of memory and attention are precisely what we make abstraction from when we appeal to an idealized subject (see Chap. 3).

Here the fact is crucial that the subject who knows the meaning of (the proposition expressed by) β∧γ has to do with the *concept* of conjunction, not with an *object* like a construction of a conjunction: an object should be *given* by some name, whereas a concept is not given: being a function, we know it simply when we are capable to compute it.[16] On the other hand, being in possession of evidence for "β∧γ" does not amount to having a term that denotes it, but to *being* (knowingly) *in* a mental state that satisfies the condition for being evidence for "β∧γ"; there is no need of a term or description 'giving' us the evidence because, when we occupy that mental state, the proposition that β∧γ is directly evident to us.

Let us consider, as a second example, the case of implication. According to condition (i), managing the $\mathcal{C}$-concept of implication requires being capable to check the property of being an evidential factor of $\beta \rightarrow \gamma$ in σ, i.e. of being a function *f* such that, for any $b \in EF_{\beta,\sigma}$, $f(b) \in EF_{\gamma,\sigma}$. The question is whether, under the hypothesis that the evidential factors of β are epistemically transparent, the property of being such a function is epistemically transparent. Concededly, it is not *obviously* legitimate to require that the subject is capable to recognize such a function (as it is in the case of the property of being a pair of evidential factors); the subject must have recognized the 'functional dependence' of evidential factors of γ on evidential factors of β as due to the existence of a function *f* associating, to every evidential factor b of β, one evidential factor *f*(b) of γ; it is therefore necessary to postulate that our mind is capable to perform the mental operation of *abstracting* a function from a relation of functional dependence. I see at least two reasons for considering the transparency assumption plausible in this respect. First, it seems to be a matter of fact that humans are in general capable of recognizing relations of functional dependence. Second, it is important to stress that the subject I have made reference to in the preceding pages is an idealized one, essentially in the same sense as a language user is idealized in linguistics, i.e. having no limits of memory, attention, and so on: a subject whose cognitive capacities and performances in any given occasion are taken as representative of the ones of an arbitrary member of the same species. Now, if such an idealized subject is not able, when acquainted with a function *f*, to recognize it as a function with such and such properties, the sole conclusion it is natural to draw is that *f* is *not* such a function. How could it be a function with such and such properties if *nobody* were capable to acknowledge that it is? Of course, it is possible that *I* am not capable

[16] See Chap. 4, Sect. 4.2.2, and Dummett's remarks quoted there: «we can make no suggestion for what it would be to be given a concept» (Dummett, 1973, p. 241); «the notion of identifying a concept […] seems quite inappropriate» (Dummett, 1973: 408).

to realize that something is such a function, because of the limits of my IQ, memory, attention, and so on; but these are precisely the factors which we make abstraction from when we make reference to an idealized subject.

According to condition (ii), one has evidence in σ for the proposition that $\beta \rightarrow \gamma$ when (a) one has selected one element f of $EF_{\beta \rightarrow \gamma,\sigma}$, and (b) one has verified that for any b such that $b \vDash_{\sigma} \beta$, $f(b) \vDash_{\sigma} \gamma$. When f has been selected, (b) amounts to a simple computation.

Again, the fact is crucial that the subject who knows the proposition expressed by $\beta \rightarrow \gamma$ has to do with the *concept* of implication, not with an *object*: an object should be *given* by some name, whereas a concept is not given: being a function, we know it simply when we are capable to compute it. On the other hand, being in possession of evidence for $\beta \rightarrow \gamma$ does not amount to having a term that denotes it, but to *being* (knowingly) *in* a mental state that satisfies the condition for being evidence for $\beta \rightarrow \gamma$; there is no need of a term or description 'giving' us the evidence because, when we occupy that mental state, the proposition that $\beta \rightarrow \gamma$ is directly evident to us.

This approach can be extended to all logical constants; as a whole, it offers the 'non-objectual' way of conceiving evidence and its possession we were looking for.

According to Prawitz, to be in possession of an intuitionistic construction *c* requires that *c* is known under some description, hence that a term is used to denote it (Prawitz, 2015: 85); on the contrary, if the reinterpretation of Heyting's explanation I have proposed is accepted, proofs are no longer given: as mental states, they are occupied, and their properties (in particular, the property of being a proof of a sentence α) are verified by occupying them. For instance, one verifies that one has a proof of β and a proof of γ through the fact that both β and γ are evident, and one verifies that one has a *pair* made of a proof of β and a proof of γ through the fact that one is capable to compute the *concept* of conjunction and the fact that one has executed the operation of pairing; there is no need that that operation is 'given'.

Finally, it is true that proofs, conceived as cognitive states, are quantified over, but this does not mean that terms are needed to denote them: variables are enough; analogously, it is perfectly possible to develop a theory of quantification over objects without having names for them.

5.5.2 *The Justification of Inference*

In Chap. 2, at the beginning of Sect. 2.4.3.1., I mentioned Prawitz's distinction between two ways of conceiving proofs: proofs as chains of legitimate inferences, and proofs as something to be defined on independent grounds. If we relate it to Heyting's explanation of the meaning of logical constants, this distinction corresponds to the one I have suggested, in Chap. 2, between evidential and inferential proofs.[17]

It is important to realize that the definition of a legitimate inference as an inference that gives rise to a proof when attached to a proof, which is usually adopted in PTS and which Prawitz (2015) charges with «turn[ing] the usual conceptual order between inferences and proofs upside down», is not made necessary by the choice of taking as fundamental the notion of evidential proof, but by the idea that a proof is what is presented by a valid argument,[18] so that the global notion of argument receives a conceptual priority over the local notion of inferential transition. If we give up this idea, and still retain as fundamental the notion of evidential proof, we are free to define a valid argument as a chain of valid inferences, along lines analogous to ToG. Now, Heyting's proofs *are* evidential proofs; on the other hand, the idea that proofs are necessarily presented by valid arguments is extraneous to the intuitionistic conceptual framework; hence it is possible to give an intuitionistic account of inferential proofs as chains of valid inferences. To show how, I will use the reinterpretation of Heyting's explanation suggested above.

First, it seems plausible to postulate that our mind is capable to perform the mental operations necessary to enter into the cognitive states required by Heyting's explanation of the logical constants. For example, as we have seen above, a proof of the sentence $\beta \wedge \gamma$ is a cognitive state σ in which, among other things, one is capable to check the property of being a pair $<b,c>$ such that b is an evidential factor of β in σ and c is an evidential factor of γ in σ; it is therefore necessary to postulate that our mind is capable to perform the mental operation of *pairing* two evidential factors. Analogously, a proof of the sentence $\beta \rightarrow \gamma$ is a cognitive state σ in which, among other things, one is capable to check the property of being a function *f* such that, for any $b \in EF_{\beta,\sigma}$, $f(b) \in EF_{\gamma,\sigma}$; it is therefore necessary to postulate that our mind is capable to perform the mental operation of *abstracting* a function from a relation of functional dependence.

Second, let us define the validity of inferences in the following way, which seems to be perfectly orthodox from the intuitionistic point of view:

---

[17] Remember that the criterion to distinguish the two notions is that the former is defined by induction over the logical complexity of proved sentences, the latter by induction over the number of inferential steps. Prawitz remarks that

> the constructions are defined by recursion over sentences, while proofs as we usually know them are generated inductively by adding inferences. (Prawitz, 2015: 86).

[18] Here is a typical expression of this idea:

> By a *proof* of A from Γ we may understand either a valid argument for A from Γ or, more abstractly, what such an argument represents; I shall here reserve it for latter use. (Prawitz, 2005: 678).

Definition 13 An inference from $\alpha_1,\ldots,\alpha_n$ to β is valid (in symbols: $\alpha_1,\ldots,\alpha_n \vDash \beta$) iff a function f from $EF_{\alpha 1,\sigma} \cup \ldots \cup EF_{\alpha n,\sigma}$ to $EF_{\beta,\sigma}$, is known such that if $a_1 \vDash_\sigma \alpha_1,\ldots, a_n \vDash_\sigma \alpha_n$, then $f(a_1,\ldots,a_n) \vDash_\sigma \beta$.

Third, let us observe that, if our mind is capable to recognize and to define the functions required to manage the logical concepts, then the inferences corresponding to Gentzen's introduction rules are valid in the sense of Definition 13, since we obviously know a function associating to every proof of the premisses a proof of the conclusion. For example, an application of $\wedge$-introduction is valid because we associate to the evidential factor b of β and the evidential factor c of γ the pair $<b,c>$; an application of $\rightarrow$ -introduction is valid because we associate, to the functional dependence of $c \in EF_{\gamma,\sigma}$ on $b \in EF_{\beta,\sigma}$, existing in virtue of our having a derivation of γ from β, the function $\lambda x^\beta y^\gamma$, which is a proof of $\beta \rightarrow \gamma$. And so on. Notice that introduction rules are not meaning-giving; the meaning of the logical constants consists in the logical concepts they denote; the validity of the rules is a consequence of the 'faithfulness' to such logical concepts of the results of the mental operations postulated for each constant.

Fourth, in connection with each logical concept C it seems plausible to postulate that our mind is capable to take evidential factors for C-propositions as inputs for certain other mental operations; for instance, extraction of one component from a pair of evidential factors; application of a function to its arguments, and so on. In this way also inferences corresponding to Gentzen's elimination rules are valid in the sense of (7). For example, consider $\vee$-elimination.

$$\cfrac{\beta\vee\gamma \qquad \begin{matrix}[\beta]\\ |\\ \delta\end{matrix} \qquad \begin{matrix}[\gamma]\\ |\\ \delta\end{matrix}}{\delta}.$$

Let σ be a cognitive state which is a justification for $\beta\vee\gamma$ and in which two proofs are known: a proof π_1 of δ from β as premiss and a proof π_2 of δ from γ as premiss. Then in σ two functions f_1 and f_2 are known such that

- for any $b \in EF_{\beta,\sigma}, f_1(b) \in EF_{\delta,\sigma}$ and, if $b \vDash_\sigma \beta$, then $f_1(b) \vDash_{\sigma\delta}$;
- for any $c \in EF_{\gamma,\sigma}, f_2(c) \in EF_{\delta,\sigma}$ and, if $c \vDash_\sigma \gamma$, then $f_2(c) \vDash_\sigma \delta$.

Moreover, in in σ a procedure p is known whose execution yields an a such that, for every cognitive state $\sigma' \geq \sigma$, either $a \in EF_{\beta,\sigma'}$ and $a \vDash_{\sigma'} \beta$, or $a \in EF_{\gamma,\sigma'}$ and $a \vDash_{\sigma'} \gamma$.

In the first case, let us apply f_1 to a: $f_1(a) \in EF_{\delta,\sigma}$ and $f_1(a) \vDash_\sigma \delta$; in the second case, let us apply f_2 to a: $f_2(a) \in EF_{\delta,\sigma}$ and $f_2(a) \vDash_\sigma \delta$; in both cases σ is a justification for δ.

Also other operations can be postulated. Consider for example Aristotle's syllogism Barbara, where β, γ and δ are mathematical formulas:

(17) $\forall x(\beta \rightarrow \gamma)$
(18) $\forall x(\gamma \rightarrow \delta)$
(19) $\therefore \forall x(\beta \rightarrow \delta)$.

Consider a cognitive state $\sigma \in \Sigma$ which is a justification for both (18) and (19); by Definition 10 and Remark 2, this means that

(i) in σ a function $g \in EF_{\forall x(\beta \to \gamma),\sigma}$ is known such that $g \vDash_\sigma \forall x(\beta \to \gamma)$, i.e. such that for any $d \in D_\sigma$, $g_d \vDash_\sigma \beta \to \gamma[\underline{d}/x]$, i.e. $g_d \in EF_{\beta \to \gamma[\underline{d}/x],\sigma}$ and for any $b \in EF_{\beta[\underline{d}/x],\sigma}$, $g_d(b) \in EF_{\gamma[\underline{d}/x],\sigma}$ and, if $b \vDash_\sigma \beta[\underline{d}/x]$, then $g_d(b) \vDash_\sigma \gamma[\underline{d}/x]$;

(ii) in σ a function $h \in EF_{\forall x(\gamma \to \delta),\sigma}$ is known such that $h \vDash_\sigma \forall x(\gamma \to \delta)$, i.e. such that for any $d \in D_\sigma$, $h_d \vDash_\sigma \gamma \to \delta[\underline{d}/x]$, i.e. $h_d \in EF_{\gamma \to \delta[\underline{d}/x],\sigma}$ and, for any $c \in EF\gamma_{[\underline{d}/x],\sigma}$, $h_d(c) \in EF_{\delta[\underline{d}/x],\sigma}$ and, if $c \vDash_\sigma \gamma[\underline{d}/x]$, then $h_d(c) \vDash_\sigma \delta[\underline{d}/x]$.

Let us stipulate that $h_d =_{def} \lambda y(g_d(h_d(y))$: by (i) and (ii), if $b \in EF_{\beta[\underline{d}/x],\sigma}$, then $h_d(b) \in EF_{\delta[\underline{d}/x],\sigma}$ and, if $b \vDash_\sigma \beta[\underline{d}/x]$, then $h_d(b) \vDash_\sigma \delta[\underline{d}/x]$; hence

(iii) $h_d \in EF_{\beta \to \delta[\underline{d}/x],\sigma}$ and $h_d \vDash_\sigma \beta \to \delta[\underline{d}/x]$.

Now let us stipulate that $h =_{def} \lambda x \lambda y(g_x(h_x(y))$: by (iii), if $d \in D_\sigma$, $h_d \in EF_{\beta \to \delta[\underline{d}/x],\sigma}$ and $h_d \vDash_\sigma \beta \to \delta[\underline{d}/x]$; hence $h \in EF_{\forall x(\beta \to \delta),\sigma}$ and $h \vDash_\sigma \forall x(\beta \to \delta)$. Hence, the mental operations that give evidence to (20), when (18) and (19) are evident, are functional composition and functional abstraction.

Finally, let us stipulate that a valid argument is a chain of valid inferences, and that an inferential proof is what is presented by a valid argument.

I have said that it is *possible* to give an intuitionistic account of inferential proofs as chains of valid inferences, and I have sketched how it can be given. I have *not* said that this is the only possible account, nor that it is the best one. Prawitz restricts himself to the analysis of what he calls reflective inference, characterized by the fact that, when we perform it, we are «aware of passing to a conclusion from a number of premisses that are explicitly taken to support the conclusion» (Prawitz, 2015: 66); but he also acknowledges that «most inferences that we make are not reflective but intuitive, that is, we are not aware of making them» (*Ibid.*). I do not agree with the presupposition that the opposition between reflective and intuitive inference is the same as the opposition between conscious and unconscious inference: much evidence is known today for the existence of *reflective* unconscious mental phenomena,[19] and inferential processes might well belong to this kind. On the other hand, I do agree with Prawitz that most inferences are unconscious; since they can also be reflective,[20] an account must be given of them that cannot be the same as the one given above of conscious inferences.

I am thinking of processes of the sort described by Jacques Hadamard in his beautiful booklet *The Psychology of Invention in the Mathematical Field*. Hadamard individuates four stages in the process of mathematical invention: preparation, incubation, illumination, and 'precising'/verifying. The first is entirely conscious:

> Sudden inspirations [...] never happen except after some days of voluntary effort which has appeared absolutely fruitless and whence nothing good seem to have come [...]. These efforts then have not been as sterile as one thinks. They have set going the unconscious machine

[19] There is evidence as well of intuitive conscious phenomena; see Evans (2010).

[20] Of course here "reflective" means no longer conscious, as in Prawitz' use, but belonging to System 2 as characterized for instance by Evans (2010).

> and without them it would not have moved and would have produced nothing. (Hadamard, 1945: 45)[21]

Incubation is on the contrary a period of unconscious work by the mind, triggered by the conscious effort of the preceding stage. At the conscious level «the study seems to be completely interrupted and the subject dropped» (Hadamard, 1945: 16); but several clues indicate that two kinds of action are performed by the mind at the unconscious level: combination and selection of ideas.

> Indeed it is obvious that invention or discovery, be it in mathematics or anywhere else, takes place by combining ideas. Now, there is an extremely great number of such combinations, most of which are devoid of interest, while, on the contrary, very few of them can be fruitful. Which ones does our mind - I mean our conscious mind - perceive? Only the fruitful ones, or exceptionally, some which could be fruitful.
>
> However, to find these, it has been necessary to construct the very numerous possible combinations, among which the useful ones are to be found. [...] As Poincaré observes, to create consists precisely in not making useless combinations and in examining only those which are useful and which are only a small minority. Invention is discernment, choice. (Hadamard, 1945: 29–30)

Illumination is again conscious. Gauss describes in the following way a relevant episode:

> Finally, two days ago, I succeeded, not on account of my painful efforts, but by the grace of God. Like a sudden flash of lightning, the riddle happened to be solved. I myself cannot say what was the conducting thread which connected what I previously knew with what made my success possible. (Hadamard, 1945: 15)

It is plausible to assume that an analogous alternation of conscious and unconscious stages takes place with inference, hence in particular that the computational processes responsible of inference are not accessible to consciousness, but their results are. As a consequence of this assumption, it is possible that we are aware of the fact that a certain procedure provides a general method transforming every proof of the premisses into a proof of the conclusion, hence of the fact that the conclusion becomes evident when the premisses are; and this *independently* of the fact that we are in a position to produce a chain of (legitimate) inferences. From this standpoint, conscious inferences are not *constitutive* of a proof, although they are necessary in order to justify the inferential transitions an argument representing that proof is made of. The two processes of generation of a proof and of justification of the argument representing it are independent of one another. When one is looking for, or generating, or even understanding, a proof, one's object is a whole pattern together with its defining property.[22] For example, when we read Euclid's proof of the proposition that there are infinitely many prime numbers, we are requested to consider, for an arbitrary natural number n, the number $n! + 1$ together with certain operations to be effected on it; when we grasp those operations as a general method to associate to

[21] Hadamard is quoting in turn Poincaré (1910: 329).

[22] According to Hadamard, «The true process of thought in building up a mathematical argument is certainly [...] to be compared with [...] the act of recognizing a person.» (Hadamard 1945: 15).

n a prime number larger than n, we have understood the proof, and the proposition becomes evident to us.[23]

If we look at human inference as an activity of our reflective mind whose results are accessible to consciousness but whose processes are largely unconscious, a strong analogy emerges between inference and linguistic knowledge, as it is described for example by L. Rizzi in the following passage:

> Mental computations extend well beyond what is accessible to consciousness and introspection. We have conscious access to the representations which are computed, and this allows us to produce and understand new structures, and to express metalinguistic judgments of well-formedness and interpretation. But we have no introspective access to the underlying computational mechanisms. (Rizzi, 2016: 347)

However, there seems to be an important difference between linguistic and inferential computations: the former are not as subject to error as the latter are[24]; this is perhaps a sign of the fact that, while linguistic computations are the result of a dedicated computational device, inferential ones either are not ruled by a dedicated device, or are ruled by a dedicated device exposed to interferences with other components—interferences which might be responsible for error. In any case, the non-immunity from error of inferential computations makes necessary a check of the correctness of proofs.[25] It seems to me that Gentzen's analysis of the aforementioned Euclid's proof (Gentzen, 1936: 144–9) could be seen, also, as a good example of a procedure that convinces us of the correctness of the proof, just because the proof can be analysed into a chain of legitimate inferences. Be that as it may, what is directly relevant to my argument is that generation of a proof and justification of the argument representing it are two distinct things.[26] Proof generation processes are guided by the explanation of the logical constants[27] (with no mention of inferential steps), while proof justification processes precisely aim to analyse proofs into chains of legitimate inferences—in my terminology, to analyse evidential proofs into inferential proofs.

This latter account of the validity of inference might be suspected of being exposed, like proof-theoretic semantics' definition of proof, to Prawitz's objection of turning the usual conceptual order between inferences and proofs upside down. But this would be an erroneous impression. When PTS defines a legitimate inference as an inference that gives rise to a proof when attached to a proof, "proof" must be understood as meaning *inferential* proof,[28] so that there is a real inversion of the usual conceptual order between inferences and proofs. The same does not hold in

[23] Cp. Hadamard's remarks on the process of understanding Euclid's proof (Hadamard 1945: 76–7).

[24] As a consequence, grammaticality judgements of native speakers are an important source of data for the linguist, whereas correctness judgements of subjects about inferences are not a source of data for the logician, although they are for the psychologist.

[25] This corresponds to Hadamard's fourth stage, 'precising'/verifying.

[26] In ToG this distinction is blurred, since a proof consists of a chain of conscious inferences.

[27] The differences between Heyting's explanation and mine are not relevant here.

[28] Within proof-theoretic semantics it is a conceptual necessity to define valid and canonical arguments by means of a *simultaneous* induction; since the definition of canonical argument makes reference to inferential steps, the notion of proof defined in this way is the inferential one.

the case of intuitionism; on the one hand, as we have seen above, the intuitionist is not compelled to adopt PTS's definition of legitimate inference; on the other, even if that definition were adopted, "proof" would be understood as meaning *evidential* proof, and this would entail no inversion: inferential proofs could still be defined as chains of legitimate inferences.

What I have said should not be understood as a disavowal of the epistemic value of inferential proofs; although validity of inference is not *defined* in terms of them, it is often *recognized* by knowing subjects through chains of inferential steps. As I said above, when we are looking for, or generating, or even understanding, a proof, our object is a whole pattern together with its defining properties: when we grasp it together with its properties we have understood the proof, and the proposition becomes evident to us; however, grasping it may be a very long process, during which the analysis of an inferential step into a number of 'atomic' steps may turn out to be necessary; to use a metaphor, it is like to cover an epistemic distance through a number of steps. This seems to be the essential reason why we are not logically omniscient.

5.5.3 Cozzo's Objections to ToG

In a paper devoted to the analysis of Prawitz's Theory of Grounds Cesare Cozzo has observed that if an inference act is defined as involving grounds for the premisses, then the ground-theoretical analysis of deduction clashes with some pre-theoretical convictions, namely: (i) that, besides valid inferences, there are invalid inferences; (ii) that there are inferences with mistaken premisses; (iii) that an inference with mistaken premisses can be valid; (iv) that «the experience of necessity of thought also characterizes the transition from mistaken premisses to their immediate consequences.» (Cozzo, 2015: 114).

It is important to observe that Cozzo's fourth objection, in particular, poses a problem that is independent of the need to extend an anti-realistic theory of meaning to empirical sentences: the necessity of a defeasible key notion (such as ground-candidate, or justification in my approach) arises within the theory of inference, independently of the nature of the sentences that are being inferred.

Prawitz's answer, in Prawitz (2015), consists in relaxing the definition of inference in the sense that the operation involved in a generic inference is no longer defined over grounds for the premisses, but over *alleged* grounds, where in fact «no requirement is put on the alleged grounds and the operation». (Prawitz, 2015: 94).

This maneuver allows him to answer Cozzo's objections (i)-(iii), but not objection (iv); moreover, answers to (i)-(iii) require to classify as mistakes some premisses that, intuitively, are not mistaken at all. Let me explain first this latter remark. In fact, in his paper Cozzo suggested, as a possible solution to his own objection, the definition of a notion of ground-*candidate*, something that can be either a genuine ground, providing a warrant to assert, or a pseudo-ground, not providing a warrant to assert.[29] But Prawitz's alleged grounds have nothing to do with ground-candidates: since no restriction is put on alleged grounds, they are entities of any kind. Now, an assertion based on an entity of any kind may be true or false, but it is difficult to see how it can be rational at all; on the contrary, the rationale for introducing ground-candidates is precisely the necessity to explain how an empirical assertion may be rational even if the asserted proposition is false.

Prawitz is perfectly aware of this essential difference between assertibility conditions of mathematical and empirical propositions. Let us read again a passage already quoted in Chap. 3:

> If a sentence is asserted in mathematics on the basis of what one thinks is a proof of it and it later turns out that the sentence is false, one would ordinarily say that the alleged proof was not a proof and that therefore the sentence was incorrectly asserted. But outside mathematics, one may want to say that a sentence was correctly asserted (on sufficient ground) although it later turned out that the sentence was false, *i.e.*, the grounds on which the sentence was asserted are still regarded as having been sufficient in the situation in question (although they are not so anymore). (Prawitz, 1980: 8)

From this standpoint, our subject's hallucinatory experience of presently falling rain is an adequate observation that it is raining, the subject *has* a ground for asserting that it is raining, and he is in a position to know that he is in possession of that ground: nothing prevents us from assuming that possession of grounds for this sentence, and more generally for observation sentences, is epistemically transparent. Now we can easily see why Prawitz's maneuver does not answer Cozzo's fourth objection, according to which «the experience of necessity of thought also characterizes the transition from mistaken premises to their immediate consequences.» If John justifiedly (but falsely) believes that it is raining, and consequently goes out with an umbrella, we deem his behaviour as perfectly rational in virtue of a valid inference; this inference is valid because it is individuated, in particular, by an operation that, applied to a ground-candidate for "it is raining" and to a ground-candidate for "if it is raining, then one goes out with an umbrella", produces a ground-candidate for "one goes out with an umbrella". But according to Prawitz's definition this is not a valid inference, because operations are defined only over actual grounds.

5.6 Conclusion

I have argued that only an intuitive notion of evidence whose possession is epistemically transparent is capable to play the role of key notion of a theory of meaning, hence also to make possible an explanation of inference on epistemic grounds. As a consequence, any definition of a theoretical notion intended to be an explicans of intuitive evidence should satisfy the condition of material adequacy that its possession is epistemically transparent. In the case of atomic sentences this is recognized

[29] Owing to the analogy between Cozzo's objections and the difficulties about the Proof Explanation of implication and negation stressed in Sect. 5.1, Cozzo's ground-candidates play a role analogous to the role of evidential factors in my approach; but no characterization of them is suggested.

by Prawitz but, as I have argued in Chap. 2, the condition of material adequacy is not satisfied by ToG in the case of empirical sentences. Moreover, one of Cozzo's objections shows that an account of inference on epistemic grounds requires a defeasible key notion also for mathematical sentences.

One of the basic ideas of ToG is that

> the question whether a term denotes a ground for an assertion of a sentence A coincides with the question of the type of the term. (Prawitz, 2015: 89)

The facts mentioned in the preceding paragraph—if they *are* facts—suggest that this idea of understanding as type-checking is untenable. For, on the one hand, the mathematical result underlying the whole idea, Curry-Howard isomorphism, is based on the assumption that grounds are objects—which is responsible for the non-transparency of their possession; on the other hand, the idea is plausible only if at least some terms denote grounds in a transparent way—and this is not ensured if, as I have argued, possession of grounds for atomic empirical sentences is not transparent.

For these reasons I have proposed, in Chap. 4 and in the present one, the definition of a defeasible notion of $\mathcal{C}$-justification for α which, not being factive but only k-factive, is not exposed to the charge of non-transparency I addressed to grounds for atomic empirical statements and such that the assumption of transparency becomes plausible. The leading idea has been to define justifications as mental states resulting from the computation of the evidential factors of sentences; in this way an answer has been suggested also to one of the questions Prawitz (rightly) considers crucial for an adequate explanation of inference:

> Could there be something like *recognizing* the validity of an inference, understood as less demanding than knowing but as something of sufficient substance to imply that one is justified in holding the conclusion true? (Prawitz, 2010: 16)

The answer I have proposed is that one recognizes the validity of an inference when one, on the basis of one's knowledge of the meaning of the premisses and of the conclusion, is capable to compute a function tranforming the evidential factors of the former into evidential factors of the latter.[30]

Finally, let me stress an important advantage of the notion of (empirical) justification I am proposing. A typical criticism to which inferentialist notions of justification are exposed is the sceptical conclusion that no belief is inferentially justified. According to the inferentialist approach, a subject $\mathcal{S}$ is inferentially justified in believing that β if and only if $\mathcal{S}$ is justified in believing that α, α is true, and β is

[30] According to Prawitz,

> to infer a conclusion A from a set Γ of premisses may be experienced not just as making the assertion A giving a set Γ of premisses as one's reason but rather as *seeing* that the proposition asserted by A is true given that the propositions asserted by the premisses of Γ are true. To characterize correct reasoning we may need to give substance to this metaphorical use of seeing. (Prawitz, 2012: 890).

My suggestion has been to give a *computational* substance to the metaphorical use of seeing.

inferred from α. The typical sceptical argument has two main steps: (i) the observation that the inference of β from α is not necessary, i.e. that the possibility cannot be ruled out that α is true and β false; and (ii) an argument to the conclusion that, in order to rule out such a possibility, it is necessary to use the very form of inference that is questioned. The notion of empirical justification I propose is *not* inferentialist: a cognitive state is not something to be believed, nor the premise of an inference. It is rather a notion of justification based on explanation: a subject $\mathcal{S}$ is justified in believing that α, in this sense, if and only if α is the best explanation of the data available to $\mathcal{S}$. In this case the sceptic has no room to make the first step. The only possibility would be to say that something *seems* to explain the data,[31] but does not really explain them; but this is just what cannot be done: there simply is no gap between apparent explanation an real explanation, or, equivalently, between seeming to understand and actually understanding. To recall a passage quoted in Chap. 3,

> We do not appear to know how to make the contrast between understanding and merely seeming to understand in a way that would make sense of the possibility that most of the things that meet all our standards for explanation might nonetheless not really explain. To put the matter another way, we do not see a gap between meeting our standards for the explanation and actually understanding in the way we easily see a gap between meeting our inductive standards and making an inference that is actually correct. (Lipton, 2004: 22)

For the same reason, a notion of justification based on explanation is epistemically transparent: there being no gap between apparent explanation and real explanation, there is not a point of view from which a cognitive state σ can be judged *to be* a $\mathcal{C}$-justification for a sentence α in spite of the fact that no subject in σ recognizes it as a $\mathcal{C}$-justification for α, or from which σ can be *judged not* to be a $\mathcal{C}$-justification for α in spite of the fact that a knowing subject in σ believes that it is. *To be* a $\mathcal{C}$-justification for α, for a cognitive state σ, is *to be conceived* as such by an idealized subject in σ: for cognitive states, as for explanations, *esse est concipi*.

References

Batens, D. (1990). Against global paraconsistency. *Studies in Soviet Thought, 39*(3/4), 209–229.

Brouwer, L. E. J. (1907). *Over de grondslagen der wiskunde*, Thesis, Amsterdam. (Part. Engl. tr. On the foundations of mathematics. In Brouwer 1975: 13–101).

Brouwer, L. E. J. (1975). *Collected works, vol. I: Philosophy and foundations of mathematics*. In A. Heyting (Ed.). North Holland.

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge University Press.

Cozzo, C. (2015). Necessity of thought. In H. Wansing (Ed.), *Dag Prawitz on proofs and meaning* (pp. 101–120). Springer.

Dummett, M. (1973). *Frege: Philosophy of language*. Duckworth.

Dummett, M. (2000). *Elements of intuitionism*. Clarendon Press. (First Edition 1977).

Evans, J. (2010). *Thinking twice. Two minds in one brain*. Oxford University Press.

[31] Notice that there is no possibility of error in taking some piece of information i as a datum: either i is the input of some of the subject's information-processing systems, thereby being a datum, or it is not, in which case it is not a datum.

Gentzen, G. (1936). Die Wiederspruchsfreiheit der Reinen Zahlentheorie. *Mathematische Annalen, 112*, 493–565. (Engl. tr. The consistency of elementary number theory. In Gentzen (1969) (pp. 132–213)).

Gentzen, G. (1969). *Collected papers*. In M. E. Szabo (Ed.). North Holland.

Hadamard, J. (1945). *The psychology of invention in the mathematical field*. Princeton University Press. (Enlarged Edition 1949, reprinted by Dover Publications, 1954.)

Heyting, A. (1956). La conception intuitionniste de la logique. *Les Études Philosophiques, 2*, 226–233.

Heyting, A. (1958). Intuitionism in mathematics. In R. Klibansky (Ed.), *Philosophy in the mid-century. A Survey* (pp. 101–115). La Nuova Italia.

Heyting, A. (1974). Intuitionistic views on the nature of mathematics. *Synthese, 27*, 79–91.

Heyting, A. (1979). Intuizionismo. In *Enciclopedia del Novecento*, vol. 3 (pp. 846–55). Istituto dell'Enciclopedia Italiana.

Kripke, S. (1965). Semantical analysis of intuitionistic logic I. In J. N. Crossley & M. A. E. Dummett (Eds.), *Formal systems and recursive functions* (pp. 92–129). Amsterdam.

Lipton, P. (2004). *Inference to the best explanation.* Second Edition. Routledge. (First Edition 1991).

Nelson, D. (1949). Constructible falsity. *The Journal of Symbolic Logic, 14*, 16–26.

Nelson, D. (1959). Negation and separation of concepts in constructive systems. In A. Heyting (Ed.), *Constructivity in mathematics* (pp. 208–225). North-Holland.

Poincaré, H. (1910). Mathematical creation. *The Monist*, *XX*(3), 321–335.

Prawitz, D. (1980). Intuitionistic logic: A philosophical challenge. In G. H. von Wright (Ed.), *Logic and philosophy* (pp. 1–10). Nijhoff.

Prawitz, D. (2005). Logical consequence from a constructivist point of view. In S. Shapiro (Ed.), *The Oxford handbook of philosophy of mathematics and logic* (pp. 671–695). Oxford University Press.

Prawitz, D. (2012). The epistemic significance of valid inference. *Synthese, 187*(3), 887–898.

Prawitz, D. (2015). Explaining deductive inference. In H. Wansing (Ed.), *Dag Prawitz on proofs and meaning* (pp. 65–100). Springer.

Quine, W. V. O. (1986). *Philosophy of logic* (2nd ed.). Harvard University Press.

Rizzi, L. (2016). Linguistic knowledge and unconscious computations. *Rivista Internazionale di Filosofia e Psicologia, 7*(3), 338–349.

Troelstra, A. S., & van Dalen, D. (1988). *Constructivism in mathematics* (Vol. 1). North-Holland.

Chapter 6
Truth and Truth-Recognition

Abstract This chapter is concerned with the relation between the intuitive notions of truth and evidence or truth-recognition. While the intuitionists grant no space to the (intuitive) notion of truth of a mathematical sentence (Sect. 6.1), according to many supporters of anti-realist theories of meaning, in particular neo-verificationist ones, the intuitionistic attitude is unacceptable because, on the one hand, it is highly counterintuitive, and on the other hand *some* notion of truth, irreducible to proof possession, cannot be avoided even within an anti-realist conceptual framework. In Sect. 6.2 two arguments for the necessity of a distinction between truth and truth-recognition are analyzed and criticized: Dummett's argument based on the Paradox of Inference, And Prawitz's considerations concerning the content of assertions. Sect. 6.3 discusses the neo-verificationist debate between temporalist and atemporalist conceptions of truth.

Keywords Anti-realist truth · Intuitionism · Dummett · Prawitz · Neo-verificationism

The internalist theory of meaning I am outlining confers the role of key notion to the intuitive notion of evidence, from which a theoretical notion of C-justification for a sentence α has been extracted. The intuitive notion of truth has been replaced by evidence for essentially two basic reasons: first, because it is based on a common-sense notion of reference not amenable to a scientific treatment; second, because the truth-conditions of a sentence are in general transcendent, and consequently their knowledge cannot be characterized as a cognitive state.

While the intuitionists grant no space to the (intuitive) notion of truth of a mathematical sentence, according to many supporters of anti-realist theories of meaning, in particular neo-verificationist ones, the intuitionistic attitude is unacceptable because, on the one hand, it is highly counterintuitive, and on the other hand *some* notion of truth cannot be avoided even within an anti-realist conceptual framework. One may certainly agree that the intuitionistic view is counterintuitive; but this may depend on the fact that our intuitive point of view is naively realist, and in any case the conflict

G. Usberti, *Meaning and Justification. An Internalist Theory of Meaning*, Logic, Epistemology, and the Unity of Science 59,
https://doi.org/10.1007/978-3-031-24605-0_6

with intuition cannot be the only reason to refuse a theoretical notion. The real problems, in my opinion, are the following: (i) whether there are compelling *theoretical* reasons for introducing a distinction between being true that α and being evident that α; (ii) whether, in the case of an affirmative answer to (i), there are reasons for a temporal rather than for an atemporal view of (epistemic) truth; (iii) whether a negative answer to (i) is consistent. I shall discuss (i) and (ii) in this chapter, (iii) in Chap. 9.

6.1 Truth from an Intuitionistic Standpoint

We have seen in Chap. 2 that, according to Heyting, «The notion of truth makes no sense […] in intuitionistic mathematics» (Heyting, 1958: 279). However, it should be stressed that Heyting is speaking of the realist, platonic, notion of truth; his assertion cannot therefore be understood as excluding that, within an intuitionistic framework, it is possible to *define* some notion that, on the one hand, can plausibly be proposed as a notion of truth, and, on the other hand, is reducible to others already present within that framework. It seems to me that the only reductive definition of truth acceptable within an intuitionistic framework is the one that equates the truth of a mathematical sentence α with the actual possession of a proof of α[1]; according to it, truth becomes a tensed property of mathematical statements, expressible by means of something like a predicate "$\text{True}_t(\underline{\alpha})$" or an operator "$\vdash_t \alpha$", meaning that α has been (or is, or will be) proved at time t, or that α is known at t. The study of this notion of truth pertains traditionally to the theory of the creative subject, but of course it is potentially interesting also from a more general, or philosophical, point of view. In any case, truth is, from this point of view, merely a *façon de parler*: there is no notion of truth distinct from knowledge.

In the preceding chapters I have neglected the notion of truth, but it should be clear that the spirit of my approach is akin to the intuitionistic attitude. Moreover, the way I have proposed to conceive proofs and justifications permits us to extend the intuitionistic reductive view of truth to empirical sentences. We have seen in Chap. 3 the rationale for conceiving justifications as cognitive states, and in Chap. 5 the reasons for conceiving evidential proofs of mathematical sentences as justifications of a particular kind, i.e. as justifications whose possession is indefeasible - hence, again, as cognitive states. Moreover, I have argued for the epistemic transparency of justifications. If justifications are cognitive states, this means that if one is in a cognitive state that is a justification for α, then one is in a position to know that one is. At this point truth can be reductively defined in the following way:

Definition 1 *α* is *true* at t if and only if t is the temporal index of a cognitive state that is a proof of α.

---

[1] Cp. Brouwer's claim quoted in Chap. 2: «there are no non-experienced truths» (Brouwer, 1949: 488).

On the other hand, a natural characterization of truth-recognition will be like the following:

Definition 2 One recognizes that α is true at t if and only if one knows that one is in a cognitive state that is a proof of α and whose temporal index is t.

Since proofs are epistemically transparent, the truth of α coincides with the recognition that α is true, and a primitive notion of truth is not necessary.[2]

6.2 Two Arguments for a Distinction

In this section will argue that, within the conceptual framework of intuitionism, the distinction between truth and truth-recognition is not necessary by replying to two arguments to the contrary, one based on the paradox of inference, the other on considerations concerning the content of a sentence.

6.2.1 The Paradox of Inference

Here is a formulation of the paradox:

> If in an inference the conclusion is not contained in the premise, it cannot be valid; and if the conclusion is not different from the premises, it is useless; but the conclusion cannot be contained in the premises and also possess novelty; hence inferences cannot be both valid and useful. (Cohen & Nagel, 1934: 173)

The following reformulation, due to Dummett, indicates a possible way out:

> For [an inference] to be legitimate, the process of recognizing the premisses as true must already have accomplished whatever is needed for the recognition of the truth of the conclusion; for it to be useful, a recognition of its truth need not actually have been accorded to the conclusion when it was accorded to the premisses. (Dummett, 1975: 297)

The moral Dummett draws from the paradox is that

> The justifiability of deductive inference - the possibility of displaying it as both valid and useful—requires *some* gap between truth and its recognition; that is, it requires us to travel some distance, however small, along the path to realism, by allowing that a statement may be true when things are such as to make it possible for us to recognize it as true, even though we have not accorded it such recognition. (Dummett, 1975: 314)

[2] I am assuming here that what makes of a predicate a *truth*-predicate is the validity of Tarski's equivalences "True($\underline{N}$) iff t', where N is the name of a sentence of the object language and t the translation of that sentence into the metalanguage. I shall argue for this claim and discuss its significance in Chap. 9.

6.2.1.1 Dummett's Argument

Let us consider the argument more closely. As a matter of fact, Dummett concedes that many forms of argument can be simultaneously valid and useful even if truth is not distinguished from truth-recognition. In a crucial passage of "The justification of deduction", after having said that

> The possibility of representing an epistemic advance as capable of being made by *indirect* means [...] rests upon having a model of meaning which does not equate the truth of a statement with our explicit knowledge of its truth. (Dummett, 1975: 313. *Italics mine*)

Dummett adds:

> It is quite different with a direct demonstration. The truth of a conjunction, for instance, simply consists in the truth of the premisses from which it is inferred by means of and-introduction, and so the recognition that it is true is not the recognition of a property which it had independently of the possibility of inferring it in that way. (*Ibid.*)

I take this passage to mean the following: as the truth of "α and β" is *constituted* by the inference of and-introduction, the truth of the conclusion did not subsist before the inference; as a consequence, an epistemic advance in passing from the premises to the conclusion is granted, and the usefulness of and-introduction is assured. The correctness of this interpretation seems to me confirmed by the following passage:

> On any possible view, it is part of the meaning of 'and' that a conjunction cannot be established save by establishing its two constituents; hence there can be no problem about the essential role of the rule of conjunction introduction in anything serving as conclusive grounds for a conjunctive statement. (Dummett, 1975: 312)

On the other hand—we may add—the rule is obviously valid: since it is constitutive of the meaning of "and", it is self-justifying, in the sense that it is valid by definition.

Since analogous considerations can be applied to all introduction rules, it follows that the simultaneous usefulness and validity of all introductory inferences can be accounted for, even if truth and truth-recognition are not distinguished: they are useful because recognizing the truth of the premisses does not coincide with recognizing the truth of the conclusion; they are valid because the truth (or truth-recognition, for that) of the conclusion is *constituted* by the truth (-recognition) of the premisses and by the execution of the inference.

The problematic case, according to Dummett, is the one in which a sentence is established indirectly, i.e. by means of an inference applying an elimination rule, for instance *Modus Ponens*; it is precisely in this case that a distinction between truth and truth-recognition is *required*. Here are Dummett's words:

> For there to have been an epistemic advance, it is essential that the recognition of the truth of the premisses did not involve an explicit recognition of that of the conclusion [...]. For the demonstration to be cogent, on the other hand, it is necessary that the passage from step to step involve a recognition of *truth* at each line. For the semantic proof of validity to have any force, that is, really to be a justification of the forms of inference used, this recognition of truth, in following out the demonstration, cannot *constitute* the truth of the statements so recognized: it must be a recognition of a property which is in accordance with the content of the statements, as given by the preferred model of meaning. (Dummett, 1975: 313)

It may be helpful to remember how Dummett conceives the validity of an elimination rule. In a nutshell, the idea is that, in the case of elimination rules, "valid" means justifiable, and that "justifiable" means eliminable: an elimination rule is valid if it is possible to obtain its conclusion without using the rule, under the hypothesis that its major premise has been obtained through an argument whose last step is an introduction; and the truth of this hypothesis is the content of what Dummett calls "the Fundamental Assumption". Take for instance $\rightarrow$ -elimination: it is valid because, under the hypothesis that its major premise is proved by introduction, we *could* have obtained its conclusion directly, without using it, by 'supplementing' the open proof of its major premise with the proof of its minor premise. We must therefore distinguish two things: having actually obtained the conclusion, and the possibility of obtaining it; at this point, if we equate the former with the recognition of the truth of the conclusion, and the latter with its own truth, we have an account of how the elimination rule can be simultaneously useful and valid, since having recognized the truth of the two premisses is not yet having recognized the truth of the conclusion, but the truth of the conclusion already subsists when the truth of the premisses does.

6.2.1.2 An Alternative View

Is Dummett's argument compelling? Before trying to answer let me illustrate an alternative solution to the problem of accounting for the simultaneous usefulness and validity of an eliminative inference—a solution that does not postulate a distinction between truth and truth-recognition.

First, let me give an explicit characterization of the usefulness of an inference. I shall say that

(1) A form of argument is *useful* if and only if the cognitive state in which one recognizes the truth of the premises is not the same as the cognitive state in which one recognizes the truth of the conclusion.

Now, the cognitive state in which one recognizes the truth of α can be equated with that in which one has a proof of α; and this state is identical with the state that *is* a proof of α (since cognitive states are epistemically transparent: if a cognitive state is a proof of α, then one is in a position to know that it is, and if one does what one is in a position to do, one knows that it is; hence one has a proof of α). As a consequence, the cognitive state in which one recognizes the truth of α is individuated by asking which cognitive state is a proof of α; and this question is answered by the inductive definition of proof of α given by Heyting. For the proofs of two sentences α and β to be different, it is sufficient that α and β have different logical forms; hence the cognitive state in which one recognizes the truth of a premise α is different, in general, from the cognitive state in which one recognizes the truth of the conclusion β, so the inference from α to β is useful.

On the other hand, validity should be defined in the following way, according to the intuitionist:

(2) A form of argument is valid if and only if a method is known for transforming every proof of the premises into a proof of the conclusion.

With these characterizations, which seem to be perfectly orthodox from the intuitionistic point of view, the simultaneous usefulness and validity of eliminative forms of argument can be explained without distinguishing truth from truth-recognition. Take for instance the rule of $\rightarrow$ -elimination or *Modus Ponens*: it is useful, simply because the logical form of β is different from that of α and of $\alpha \rightarrow \beta$; it is valid, because the major premise assures that we have a constructive function f transforming every proof of α into a proof of β, and the minor premise that we have a proof of α: the method required by (2) is the operation of functional application of f to the given proof of α. And we are not forced to distinguish truth from truth-recognition; for, let us assume that the truth of β coincides with the recognition of its truth: what follows is that, since when we have recognized the truth of α and of $\alpha \rightarrow \beta$ we have not yet recognized the truth of β, β is not yet true before we have applied the function f to the proof of α; but this does not prevent the inference from being already valid, since its validity consists in the fact that the method mentioned in (2) (in this case, the operation of functional application) has an intrinsic property, known to subsist before and independently of any application of the method (in this case, the property of yielding a proof of β when applied to f and to a proof of α).

6.2.1.3 Dummett's Argument Reconsidered

If my proposal is tenable, Dummett's argument cannot be compelling. My hypothesis is that it is valid but not sound, because one of its premises is not true, so that its conclusion is not supported. The premise that, in my view, is not true is that it is necessary to distinguish two kinds of proofs: direct or canonical, and indirect or non-canonical ones. That this premise is false, i.e. that the canonical/non-canonical distinction is not necessary within the intuitionistic conceptual framework, has been argued in Chap. 2, Sect. 2.4.2; here I shall argue that the necessity of the canonical/non-canonical distinction is a premise of the argument.

That the distinction is a premise of Dummett's argument is easily seen from the fact that Dummett concedes that the simultaneous usefulness and validity of direct proofs can be accounted for without distinguishing truth from truth-recognition, whereas he does not concede it for indirect proofs. This depends on the fact that Dummett's very characterizations of usefulness and validity are different in the case of direct and indirect inferences.

(3) The usefulness of an introductive inference is defined essentially in the way I have suggested in (1).[3]
(4) The usefulness of an eliminative inference, on the other hand, is defined (not explicitly) as a sort of 'energetic saving'.

[3] Provided that my interpretation of the passage from Dummett (1975: 313) quoted at the beginning of Sect. 6.2.1.1 is correct.

In "The justification of deduction", when he analyzes an application of *Modus Ponens*, Dummett observes:

> The proof having been accepted, we are willing to proceed from an assertion of the antecedent, however based, to an assertion of the consequent, without necessarily carrying out that operation which the proof supplies which will lead to a direct verification of the statement we are inferring. (Dummett, 1975: 308)

Hence, an inference applying *Modus Ponens* is useful in the sense that it permits to recognize the truth of its conclusion in a more rapid/economic way than the one consisting in obtaining a direct proof of it—with a metaphor Dummett often uses, it offers a *shorthand* to the conclusion; in this rapidity/economy consists the usefulness of *Modus Ponens*, and in general of eliminative inferences.

(5) The validity of an eliminative inference consists, as we have seen, in the fact that we could have arrived at the conclusion without it, by means of introductive rules only (if the Fundamental Assumption holds).
(6) The validity of an introductive inference consists, as we have seen, in the fact that, if its premisses are true, the conclusion is *made* true by the inference itself.

Now, it is just because usefulness and validity are defined in different ways for introductions and eliminations that Dummett can conclude that, in the case of eliminative inferences, truth and truth-recognition must be distinguished. Since a canonical proof is conceived as an argument whose main deduction does not contain eliminative inferences (Dummett, 2000: 272), the distinction canonical/non-canonical is indeed a crucial premise of Dummett's argument.

If we give up the distinction, a general picture emerges of the justification of deduction, quite different from the neo-verificationist one.

Consider for instance the inductive simultaneous definition of canonical proofs and proofs given by Prawitz (2005) along the lines of Martin-Löf (1985). Canonical proofs are conceived as a subclass of proofs. However, the only difference from Heyting's original definition is that the proofs mentioned in the inductive clauses are qualified, in Prawitz's definition, as effective methods for finding canonical proofs, whereas Heyting gives no characterization of them.[4] Now, as I have argued, there is no need of this qualification in order to avoid circularity: in both definitions circularity is avoided, as in any recursive definition, because the sentences whose proofs are mentioned in each clause are less complex than the sentence the clause deals with. Secondly, what is an effective method for finding a canonical proof? It seems to me that there is no other way of having some intuitive grasp of this notion than by appealing to valid *arguments*: an effective method for finding a canonical proof is a sort of 'abstract' analogue of the reduction procedures of non-canonical arguments to canonical ones. But these reductions are needed just for arguments, in order to define their validity (where validity amounts more or less to reducibility to canonical form); making reference to them in the case of mental entities like proofs seems utterly unmotivated: the notion of proof is intended to articulate the notion of evidence, and

[4] Actually, in Prawitz's definition the clauses for the constants different from $\rightarrow$ and $\forall$ require that the subproofs are canonical; but, as Prawitz remarks (fn. 9), this requirement can be left out.

evidence either commits us to recognize truth or is not evidence: a kind of evidence that requires to be *justified* by another kind simply is *not* evidence.[5] Therefore, the substitution of Heyting's definition of proof with the simultaneous definition of canonical proof and categorical proof is not motivated. What is motivated is the distinction between canonical and non-canonical *arguments*; but it seems to be conceived as an epistemological distinction rather than an ontological one, in the sense that canonical and non-canonical arguments do not present two kinds of proofs, but are two *ways* of presenting proofs of one kind. And this view has a consequence concerning the justification of deduction: there are not self-justifying forms of argument; all forms of argument, direct as well as indirect, need to be justified, precisely by the fact that they give methods for transforming the proofs of the premises into proofs of the conclusions. (This is just the definition of validity I have suggested as intuitionistically orthodox.)

To sum up: the distinction between truth and truth-recognition is indeed necessary if the distinction between canonical and non-canonical proofs as two kinds of proofs is necessary; conversely, if we do not introduce the canonical/non-canonical distinction, it is not necessary to distinguish truth from truth-recognition in order to solve the paradox of inference. Shortly, the distinction between truth and truth-recognition is necessary if and only if the canonical/non-canonical distinction is; since, as I have argued, the latter is not conceptually necessary within the framework of intuitionism, neither the former is a conceptual necessity.[6] Moreover, the distinction between canonical and non-canonical proofs as two kinds of proofs seems to clash with the intuitionistic view of proofs as cognitive states; since the very possibility of basing an intuitionistic theory of meaning for empirical sentences on the defeasible notion of justification seems to depend on conceiving justifications (and therefore proofs) as mental entities, there is at least one good reason for giving up the distinction between canonical and non-canonical proofs as two kinds of proofs,[7] i.e. a reason for *not* distinguishing truth from truth-recognition.

6.2.2 The Content of Assertions

The second argument for the necessity of a distinction between truth and truth-recognition is based on the idea that the content of an assertion cannot, in general, be taken to be that the assertion has been or will be verified.[8] Prawitz, for instance, holds that it is

---

[5] Of course I am not speaking of hypothetical evidence, which is recognized by intuitionists, but of indirect evidence.

[6] With "conceptual necessity" I mean a principle whose validity can be extracted from the sole analysis of the concepts involved.

[7] This is not to say that a distinction on different grounds is not possible.

[8] Cp. for instance chap. 13 of Dummett (1973), Brandom (1976), (Prawitz 1987: 137), Prawitz (1998a: 46–47), Prawitz (1998b: 30), Prawitz (1998c: 291–292).

> a misrepresentation of the assertion to think of its content as being that a proof has been found. It is to put too much in the content. (Prawitz, 1998a: 46)

He argues for this thesis in the following way:

> To take an absurd example, it may be true that somebody who proves that there are infinitely many twin primes, must know an extraordinarily great deal about prime numbers, but from this, it does not follow that if there are infinitely many twin primes, then someone knows or will know so much about prime numbers - that absurdity follows however, if the sentence "there are infinitely many twin primes" is taken to mean that someone has proved or will prove that there are infinitely many twin primes. (Prawitz, 1987: 137)

Let me try to reconstruct the argument. Suppose the content of the sentence

(7) There are infinitely many twin primes

is the same as the content of the sentence

(8) Somebody has proved or will prove that there are infinitely many twin primes;

then the sentence

(9) If somebody proved (will prove) that there are infinitely many twin primes, then somebody knows (will know) a great deal about primes;

should entail the sentence

(10) If there are infinitely many twin primes, then somebody knows (will know) a great deal about primes.

But (9) does not entail (10); therefore (7) and (8) cannot have the same content. The positive suggestion is that the content of (7) should rather be equated with the content of

(7') It is true that there are infinitely many twin primes,

provided, of course, that "It is true that α" is understood as not equivalent to "Somebody proved (will prove) that α".

I do not think that the intuitionists (in particular Heyting) have ever equated the content of (7) with the content of (8), nor do I see any reason for equating them. When Heyting writes

> Every theorem has the form (if enunciated without abbreviations): «A construction with such and such properties has been effected by a mathematician» (Heyting 1958: 278)

what he means is that the *assertion* of a sentence α amounts to asserting that a proof of α has been effected, *not* that the *content* of the assertion of α is that a proof of α has been effected. On the question of the relations between assertion and content Heyting's view does not seem substantially different from Frege's; according to Frege (Frege, 1918: 356) the assertion of α is the manifestation of the mental act of judging α, i.e. of recognizing the truth of α, but the content of the assertion of α is the same as the content of other acts, like asking, assuming, and so on: it is the content of α itself, and this content is a thought. Analogously, according to Heyting the assertion of α is the manifestation of the knowledge that a proof of α has been effected, but the content of α is the same as the content of other acts: a thought, or a proposition. For both Frege and Heyting the content of "Snow is white" is the thought (or proposition) that snow is white. Moreover, as we have seen in Chap. 2, Heyting adds that

> The affirmation of a proposition is not itself a proposition; it is the determination of an empirical fact, viz., the fulfillment of the intention expressed by the proposition. (Heyting, 1931: 48)

Therefore, the content of (7) cannot be the same as the content of (8), which is a sentence expressing an empirical fact.

This seems to me a perfectly acceptable view, provided that we can give a plausible account of what a proposition is. Well: my suggestion in Chap. 4 has been to conceive propositions as in a sense analogous to Carnapian intensions, with the difference that while the Carnapian intension of a sentence is a function from possible worlds to truth-values, I suggest to define the proposition (or the thought) expressed by α as a function f from cognitive states σ to $\{0,1\}$, such that $f(\sigma) = 1$ if σ is a proof of $\alpha, f(\sigma) = 0$ otherwise. A subject can be said to grasp the thought expressed by α if (s)he can compute f, i.e. if (s)he can discriminate the cognitive states that are proofs of α from the ones that are not.

Prawitz argues for the necessity of a notion of truth distinct from truth-recognition also by means of an analysis of such mathematical activities as conjecturing and wondering:

> My problem is [...] *what* the intuitionist is conjecturing or wondering. If he is allowed the notion of provability in the sense of potential existence of a proof, then we can take him to conjecture that the sentence is provable and to wonder whether it is. But if truth is analysed in terms of our already having at hand a method for finding a proof [...], then it seems difficult to attach any sense to normal conjecturing or wondering in mathematics - to conjecture is certainly not normally to conjecture that we already have a proof or a method for finding one, and to wonder is not normally to wonder whether we have such a thing. (Prawitz, 1998c: 291)

The answer I suggest is essentially the same as before: to conjecture that, or to wonder whether, or to assume that, there are infinitely many twin primes is not to conjecture/wonder/assume that the sentence "There are infinitely many twin primes" is true, but to conjecture/wonder/assume that there are infinitely many twin primes, i.e. to manifest certain mental acts towards the thought/proposition expressed by that sentence; the thought/proposition expressed is characterized in terms of the notion of proof: no essential role is played by the notion of truth; no notion of content of a sentence as distinct from its meaning is necessary. Concluding, the intuitionist who identifies truth with truth-recognition is not committed to the idea that the content of an assertion is that it has been or will be verified.

It may be interesting to ask why Prawitz is assuming that the intuitionist is committed to that idea. I submit that the reason is an underlying assumption Prawitz himself, like most neo-verificationists, makes: that «the content of the assertion is that the asserted sentence is true.» (Prawitz, 1998c: 46) From this assumption and the thesis that α is true if and only if one knows that α, it follows that the content of the assertion is that the asserted sentence is known. However, the intuitionist does not subscribe the neo-verificationist's underlying assumption: for him, the content of the assertion of "Snow is white" is the content of the asserted sentence "Snow is white", i.e. the thought that snow is white.

It seems to me that the underlying neo-verificationist assumption generates perplexing problems within the neo-verificationist teories of meaning. According to Prawitz,

> you must make a distinction between what is guaranteed and what is said by an assertion. By asserting a sentence you guarantee that there is a proof of it, but that is not what the assertion says; the *content* of the sentence, what you say by asserting it, is simply that the sentence is true, not that you have a proof of it. (Prawitz, 1998a: 46)

The content of α is different from the meaning of α, since knowing the meaning of α consists in knowing its assertibility conditions, while knowing its content amounts to knowing its truth-conditions, and truth-conditions are not assertibility conditions. As a consequence, when we ask: "Are there infinitely many twin primes?", we do not assign to the sentence we are using its normal (intuitionistic) meaning, but a different one. The distinction invoked by Prawitz seems therefore to evoke a sort of double or counterpart of the intuitionistic meaning of sentences: their content, 'what they say'. I surmise that it is what in Chap. 6, Sect. 6.3.2.2 I will call their *potential* meaning.

More to the point, the assumption that the content of the assertion is that the asserted sentence is true is exposed to the very objection Prawitz raises against the intuitionist. Let us consider again (7): as we have seen, its content should be equated with the content of (7'); since, according to the neo-verificationist, "α is true" means that α is knowable, or that there is a proof of α, the content of (7) is the same as the content of

(11) It is knowable that there are infinitely many twin primes

or of

(12) There is a proof that there are infinitely many twin primes.

Then both the sentences

(13) If it is knowable that there are infinitely many twin primes, then the human capacities are truly great

and

(14) If there is a proof that there are infinitely many twin primes, then the human capacities are truly great

should entail the sentence

(15) If there are infinitely many twin primes, then the human capacities are truly great,

which is absurd.

6.3 Neo-Verificationist Conceptions of Truth

Let us consider the second problem mentioned at the beginning. The objections to the intuitionistic identification of truth with truth-recognition I have tried to answer in the preceding section have been raised by neo-verificationists, who find the distinction between canonical and non-canonical proofs a conceptual necessity; as a consequence, they do not accept the intuitionistic *definition* of truth as actual possession of a proof and look for an anti-realistically acceptable but *primitive* notion of truth, not reducible to actual possession of a proof, but of course in need of elucidation. As a matter of fact, essentially two proposals concerning (what might be called) epistemic truth have been formulated: the former conceives truth as a *temporal* property of (propositions expressed by) sentences, the former as an *atemporal* property. In this section I shall introduce the neo-verificationist debate between temporalists and atemporalists, and I shall raise objections against both positions.

Let us observe preliminarily that, while the primary alternative is between temporal and atemporal truth, this intersects with another, secondary, alternative between an existential and a modal characterization of truth. According to the former α is true if and only if a proof (if α is a mathematical sentences) or a verification (if α is an empirical sentence) of α *exists*; according to the latter α is true if and only if α is knowable, i.e. it is *possible* to know it. The question of elucidating the primitive notion of (epistemic) truth leads therefore to the question of elucidating the involved notions of existence and possibility.

6.3.1 Temporal and Atemporal Truth

The central idea of the atemporal notion of epistemic truth is to conceive existence, and therefore truth, as tenseless notions: a proof (or a verification) of α exists, or does not exist, outside time; equivalently, α can, or cannot, be proved (verified) outside time. As a consequence, if α is proved at time t, it is, strictly speaking, senseless to wonder whether α was true before t; but a more liberal way of speaking can also be adopted, according to which α was true before t, and obviously remains true thereafter; otherwise stated, the atemporality of truth can be equated with its eternity. In either case the act of proving α is not what makes it true, but only what allows us to ascertain its truth, which subsists even when no proof act is performed.

The view of truth as an atemporal property is explicitly upheld by Prawitz and Martin-Löf. Prawitz writes for example:

> A statement is true, I suggest, if and only if it can be verified, or in other words, it is verifiable. If a statement is in the past or future tense, we had better say that it is true if and only if it could have been verified (it was verifiable) or it will be possible to verify (it will be verifiable), respectively. More generally, we may say that a statement is true if and only if a verification exists, in an abstract, tenseless sense of exists. (Prawitz, 1998b: 31)
>
> [A] sentence does not become true by our verification of it: the possibility of a direct verification must already have been there, and hence the sentence was already true before

> our verifying it. In other words, if it is asked about a sentence that has not been verified or refuted whether it can be verified, we have an objective question, there is a fact of the matter. (Prawitz, 1998b: 32)

Prawitz explicitly admits that his notion of truth conflicts with orthodox intuitionistic views, for example in the following passage:

> Intuitionistic philosophers sometimes use true as synonymous with the truth as known, but this is clearly a strange and unfortunate use. We need a notion of truth where, without falling into absurdities, we may say, *e.g.*, that there are many truths that are not known today. [...] [T]he two concepts [the non-realistic and the realistic concept of truth (*g.u.*)] agree (in contrast to the intuitionistic one mentioned above) in allowing the existence of truths which in fact will never be known. (Prawitz, 1980: 8–9)

Martin-Löf, instead, accepts the intuitionistic notion as legitimate, identifying it with what he calls actual truth; but he introduces beside it a notion of *potential* truth:

> to say that A is potentially true is to say that A can be proved, that is, that a proof of A can be constructed, which is the same as to say, in usual terminology, simply that A is true. (Martin-Löf, 1991: 142)

Moreover, he holds that the intuitionists should admit potential truth within their conceptual framework:

> it has often been pointed out that it is very counter-intuitive to say that a proposition becomes true when it is proved [...] of course, even an intuitionist cannot fail to understand this objection, and ought to answer it by saying that there is, not only the notion of actual truth, but also the notion of potential truth [...]. (*Ibid.*)

Here is how Martin-Löf elucidates the notion of potential truth:

> We may characterize the difference between the notions of actual truth and potential truth by saying that actual truth is knowledge dependent and tensed, and potential truth is knowledge independent and tenseless, but observe that, even if I say that it is knowledge independent, it does not mean that it is altogether independent of the notion of knowledge, because [...] to say that A is potentially true is to say that A can be actually true, that is, that A can be known to be true. So, clearly, the notion of actuality precedes potentiality in the conceptual order, and hence [...] the notion of potential truth [...] is conceptually dependent upon the notion of knowledge. (*Ibid.*)

The remark on the partial dependence of the notion of potential truth upon the notion of knowledge is presumably intended to make more plausible the intuitionistic acceptability of atemporal truth. However, two points should be stressed. First, the definition of potential truth makes essential reference not only to the notion of knowledge (or proof) but also to the notion of possibility, and possibility is conceived by the intuitionists as an epistemic state, i.e. as a state the knowing subject occupies within time: the state in which one acknowledges one's possession of a method or a procedure to obtain a proof. On the contrary, the notion of possibility introduced by Prawitz and Martin-Löf is not characterised as an epistemic state, but as merely *factual* accessibility, subsisting out of time, to an epistemic state: it is far from clear that such a notion is intuitionistically acceptable. Second, the principle invoked by Martin-Löf in order to justify the introduction of potential truth is a typical modal principle:

> [I]f a proposition has been, is being or will be proved, then certainly it can be proved, that is, it is potentially true [...]
>
> The principle just spelled out is again a principle which had a succinct scholastic formulation: it is the principle, Ab esse ad posse valet consequentia (illatio). (Martin-Löf, 1991: 143)

The principle is certainly acceptable for an intuitionist, but only in its temporal version:

(16) If α is proved at t, then α can be proved at t;

from the acceptability of (16) nothing can be inferred about the acceptability of atemporal potential truth.

Dummett, on the contrary, has proposed a temporal notion of epistemic truth, concisely characterized by Prawitz in the following way:

> Dummett has tried a position somewhere between the one I am suggesting and the identification of truth with 'is verified'. He considers cases in which we have constructed a verification of a statement A but we do not know that it is a verification of A and hence have not verified A, and Dummett's proposal is that in such cases A is true. An illuminating example is the case where we have found a (constructive) verification of A∨B. Such a verification will contain a procedure which, when carried out, will yield a direct verification of A or of B, although we do not know that it is a verification of A or that it is a verification of B, and then, according to Dummett's proposal, either A is true or B is true.
>
> This proposal thus solves the problem about disjunctions,[9] but it still makes truth tensed and depending on whether we are actually in possession of a procedure of the appropriate kind. (Prawitz 1994: 88).

6.3.2 Objections and Discussions

The reason why Dummett supports a temporal notion is that atemporal truth is open, according to him, to a fundamental objection (Dummett, 1998: 127–129).

6.3.2.1 Dummett's Objection to Atemporal Truth

I state Dummett's objection in the form of an argument.

(i) Let P = {α|α is provable}, R = {α|α is refutable}, N = {α|α is neither provable nor refutable}. From the atemporal conception of the existence of proofs the following trichotomy follows:

(17) Either α ∈ P, or α ∈ R, or α ∈ N.

[9] The problem consists in the fact, pointed out by Dummett himself, that, under the identification of truth with actual possession of a proof (or of a verification), α∨β may be sometimes true without either α being true or β being true, i.e. that temporal truth does not commute with disjunction. I shall come back to this point in Chap. 9.

Explanation

For every α, it is determinate whether there is or there is not a proof of α, i.e. whether $\alpha \in P$ or $\alpha \notin P$; likewise, it is determinate whether there is or not a refutation of α, i.e. whether $\alpha \in R$ or $\alpha \notin R$. As a consequence, it is determinate whether $\alpha \in N$ or $\alpha \notin N$.

(ii) asserting α is asserting that $\alpha \in P$

asserting $\neg\alpha$ is asserting that $\alpha \in R$

asserting $\neg\neg\alpha$ is asserting that $\alpha \notin R$

Explanation

$\neg\neg\alpha$ is assertible iff it can be proved that α cannot be refuted.

(iii) asserting $\neg\neg\alpha$ is asserting that either $\alpha \in P$ or $\alpha \in N$

Explanation

Given the trichotomy (17), asserting that $\alpha \notin R$ amounts to asserting that $\alpha \in P$ or $\alpha \in N$.

(iv) It is inconsistent that $\alpha \in N$ is assertible

Explanation

(a) If $\alpha \in N$ is assertible, then $\alpha \notin P$ is assertible and $\alpha \notin R$ is assertible (by definitions of P and R)
(b) If $\alpha \notin P$ is assertible, then $\neg\alpha$ is assertible (by definition of $\neg$)
(c) If $\alpha \notin R$ is assertible, then $\neg\neg\alpha$ is assertible (by definition of $\neg$)
(d) If $\alpha \in N$ is assertible, then $\neg\alpha \wedge \neg\neg\alpha$ is assertible: contradiction.
(v) asserting $\neg\neg\alpha$ is asserting that $\alpha \in P$

Explanation

From (iii) and (iv).

(vi) $\alpha \vee \neg\alpha$ is assertible

Explanation

From (i), (ii) and (v).

(vii) $\alpha \vee \neg\alpha$ is not intuitionistically assertible, in contradiction with (vi).

6.3.2.2 Prawitz's Answer

Prawitz (1998c) calls into question step (i):

> [T]he argument hinges upon the assumption that from my conception it follows that each sentence belongs to one of the three classes considered: then it can be concluded that a sentence which does not belong to the set of refutable sentences belongs to one of the other two sets. However, the assumption is unwarranted as far as I can see. [...]
>
> The crucial question is thus this: can one consistently maintain that the existence of proofs is an objective matter, that if a proof of a sentence is found, then it was already determined that the sentence had a proof, without being committed to holding also that the existence of proofs is determinate in the sense that for each sentence there is either a proof of it or there is no proof of it? (Prawitz, 1998c: 288)

Let us formulate more explicitly the crucial question CQ; let "α is true" abbreviate "There is a π such that π is a proof of α" (where the domain of the quantifier is the class of proofs of sentences); then the question is

(CQ) Can one consistently maintain that

(i) For every α (if α is proved at t, then α is true)
without being committed to holding also that
(ii) For every α (either α is true or α is not true)?

However, (in Prawitz 1998c: 288–289) Prawitz does not answer exactly (CQ), but the following:

(Q) Can one consistently maintain that

(iii) There is not an α such that (α is not true and $\neg\alpha$ is not true)
without maintaining that (ii)?

His answer to (Q) consists in (a) deriving a contradiction from the assumption that there is an α such that (α is not true and $\neg\alpha$ is not true), thereby showing that (iii) is assertible; and (b) using the fact that the implication $\neg\exists x\neg Px \rightarrow \forall xPx$ is not intuitionistically valid to show that (i) may not be assertible.

This clearly shows that he assigns to the metalinguistic logical constants their intuitionistic meanings. So, let us apply the same idea to answer (CQ). If we read the implication (i) intuitionistically, it says that there is a general method M to associate to every verification of the antecedent, i.e. to every observation that a proof of α has been constructed, a verification of the consequent, i.e. of the fact that a proof of α atemporally exists. Does such a method exist? For the orthodox intuitionist, who identifies existence of proofs with their actual possession, surely not: before Andrew Wiles had proved Fermat's Theorem, a proof of it did not exist; and invoking the modal principle *Ab esse ad posse* would be besides the point, as we have seen. Hence, it is not sufficient to assign to implication and other logical constants their intuitionistic meaning to be warranted to assert (i): it is necessary to assign them a *potential* meaning I shall characterize in a moment.

Moreover, the atemporal existence or non-existence of a proof of α in no way *depends* on the fact that, at a certain time, a proof (or a refutation) of α has been found, as the intuitionistic reading of implication would require. Suppose that in five years a nuclear war destroys all life on earth, and that in the meantime no proof nor refutation of G, Goldbach's Conjecture, is found: in five years no proof of G will have been found, but how could this fact have any effect on the atemporal existence of a proof of G? We can conclude that the antecedent of implication (i) plays no role: for every α, (i) is true (under the potential reading of implication to be defined) iff its consequent is true, i.e. iff a proof of α atemporally exists. The same remark can be made about the implication

(i') For every α (if α is refuted at t, then ¬α is true):

for every α, (i') is true iff ¬α is true. Since if ¬α is true α is not true, and, as a matter of fact, the domain of proofs either contains a proof of α or contains no proof of α, the following

(18) For every α, either α is true or α is not true

is true, in the potential sense defined below: one of the two disjuncts is true. This seems to be confirmed by the following passages:

> the question of whether something is a proof is fixed when the meanings are given, that is, when it is given what counts as a canonical proof. From this it is natural to conclude that already, before a proof of a sentence is found, it is determined that there is such a proof. Provability, which I want to identify with truth, becomes in this way something objective. (Prawitz, 1998a: 50)
>
> [I]f it is asked about a sentence that has not been verified or refuted whether it can be verified, we have an objective question, there is a fact of the matter. (Prawitz, 1998b: 32)

These passages are interesting also from another point of view: they give an essential hint about how Prawitz is, in fact, understanding the metalinguistic logical constants: not according to orthodox intuitionism, as we have seen, nor classically, because of the role ascribed to the notion of proof; as anticipated, my suggestion is that he uses them according to what, in Martino & Usberti (1994), has been called their *potential* meaning. The basic idea is that the introduction of the notion of atemporal existence of proofs gives sense to the notion of *objective fact*, which is senseless within an intuitionistic conceptual framework: «the objective fact that A is the fact that the proposition "A" is atemporally provable» (Martino & Usberti, 1994: 103). Once this notion is accepted, «it is quite natural to characterise proofs in terms of the facts they prove» (*Ibid.*),[10] and the following characterization of the potential meaning of the logical constants can be given (where "is true" is to be read according to the stipulation above):

Definition 3

- ⊥ is not true;
- α∧β is true if α is true and β is true;
- α∨β is true if either α is true or β is true;

[10] As I remarked in Sect. 2.2.2.3 of Chap. 2, this is exactly what Heyting could not have done.

- $\alpha \rightarrow \beta$ is true if it is provable that β is true provided α is true;
- $\forall x\alpha$ is true if it is provable that, for every individual d, $\alpha[\underline{d}/x]$ is true;
- $\exists x\alpha$ is true if, for some individual d, $\alpha[\underline{d}/x]$ is true.

Notice that the the metalinguistic logical constants of this definition are to be interpreted classically, since the atemporal existence, for example, of all the objective facts expressed by $\alpha[\underline{d}_1/x]$, $\alpha[\underline{d}_2/x]$,… is an objective fact as well; the potential meaning of the logical constants can therefore be seen as their classical meaning applied to an ontology of proofs of atomic sentences.

Therefore, if we reason in the metalanguage according to the potential meaning of the logical constants, we cannot consistently maintain (i) without holding true (ii); on the other hand, we cannot maintain (i) if we reason in the metalanguage according to the intuitionistic meaning of implication. What we *can* do—which is what, I suspect, Prawitz does—is to understand (i) according to its potential meaning, and (ii) according to its intuitionistic meaning: in this way we can assert (i) and refrain from (ii).

The question is whether this strategy, in turn, is consistent. Prawitz writes:

> I do not see why the disjunction ‘either there exists a proof of A or there does not exist a proof of A’ must be taken in a classical way. Although we think of the proofs as having some kind of existence even before we find them, an intuitionist may still maintain that to assert the disjunction that either there is or there is not a proof of A requires that we know how to find a verification either of the existence of a proof of A or of the non-existence of a proof of A. For an arbitrary A we do not know how to find such a verification, and we should then have no difficulty in resisting the thought that the disjunction in question is true. (Prawitz, 1998a: 48)

It seems to me that, while an orthodox intuitionist may of course *not* take the disjunction in a classical way, a potential intuitionist cannot, at least in the course of a reasoning in which (s)he wants to assert (i): to be legitimated to assert (i) (s)he must introduce the notion of objective fact, thereby assigning the logical constants their potential meaning; at this point it does not seem consistent to come back to the orthodox understanding of disjunction in order to refrain from (ii).

In conclusion, from the fact that possibility is conceived as tenseless it follows that what has been called in Chap. 2 the principle of valence[11] becomes intelligible and valid in its potential reading, in which it simply means that all sentences are atemporally determinate; Prawitz’s insistence that

> in view of our constructive understanding of disjunction [...] we know of no ground for holding for every sentence that there either is or is not a proof of it. (Prawitz, 1998c: 289)

is illegitimate for the reasons just explained.

[11] See Chap. 2, (9).

6.4 Conclusion

I have argued that, when the meaning of the logical constants is explained in terms of the notion of proof, truth and truth-recognition may be equated. Of course truth, defined in this way, is not classical/realistic truth: the class of sentences that are classically true is not the same as the class of sentences whose truth is recognized, in the sense that a proof of them is actually possessed. So the question arises: in which sense is it legitimate to say that truth-recognition is a notion of truth? More generally: under which conditions is a notion a notion *of truth*? The question is important and difficult, and I shall discuss it in detail in Chap. 9.

I conclude with a remark on the principle of valence (18). We have seen in Chap. 2 that Dummett proposes it as a criterion of objectivism in semantics. He writes for example:

> In some cases, the semantic theory advocated by the anti-realist will involve, not merely the rejection of bivalence, but the abandonment of a truth-conditional theory of meaning, because the semantic theory underlying his theory of meaning is not of the objectivist type: he does not admit, for statements of the given class, any notion of truth under which each statement is determinately either true or not true, independently of our knowledge. (Dummett, 1982: 242)

On the basis of this criterion, Dummett classifies intuitionism as a form of non-objectivism; here is why:

> In this semantics [the intuitionistic semantic theory for mathematical statements sketched by Heyting (*g.u.*)], the semantic values of the component expressions of a sentence jointly determine a decidable relation between that sentence and an arbitrary mathematical construction which obtains just in case that construction constitutes a proof of that sentence. A sentence may then be said to be true if and only if there exists a construction that constitutes a proof of it: but, since the phrase "there exists", in this definition, is itself interpreted constructively, we may not assert, for an arbitrary mathematical statement with a well-defined meaning, that there either does or does *not* [Italics mine (*g.u.*)] exist a construction which is a proof of it, nor, therefore, that it either is true or is not true. (Dummett, 1982: 234)

However, in another passage of the same paper he writes:

> If we wish to say that a mathematical statement can be true only if there exists a proof of it, we have [...] only two choices. We can interpret "exists" as meaning concrete existence, our actual possession of a proof; in this case "is true" becomes a tensed predicate of mathematical statements, a statement being able to change from not being true to being true, although not conversely. Each statement is then either true or *not* [Italics mine (*g.u.*)] true at any given time, although it may be neither true nor false, where its falsity involves the existence of a disproof; but there will be no question of its being objectively true, although we (collectively) are unaware of its truth. Alternatively, we may construe "exists", and therefore "is true", as tenseless. We shall, in this case, have to interpret "exists" constructively; we can then rule out the possibility of a statement's being neither true nor false, since its not being true would be tantamount to its being false, but we cannot assert, in advance of a proof or disproof of a statement, or an effective method of finding one, that it is either true or false. Because, on this second interpretation, "exists" is understood constructively, we shall still be unable to conceive of a statement as being true although we shall never know it to be true, although we can suppose true a statement as yet unproved. (Dummett, 1982: 259).

The intuitionist exemplifies the first interpretation, since for him existence of a proof is actual possession of it; hence for him each statement is either true or not true at any given time: according to Dummett's criterion, he is a semantical objectivist, against what is said in the first passage.

Where does the contrast between the two passages come from? In my opinion, from a different construal of the italicized "not" in the two passages: in the former as (intuitionistic) mathematical negation (i.e., as implication of a contradiction), in the latter as empirical negation. The former construal is correct if "exists", and therefore "is true", is understood as tenseless (because tenseless non-existence cannot be empirically observed, but only be mathematically proved), but it is incorrect if "exists" is understood as tensed, as it is understood by intuitionists. In this case "not" must be construed as empirical negation, and the principle of valence holds; this is the content of Heyting's passage quoted in Chap. 5, at the end of Sect. 5.2.2; an analogous passage can be quoted here:

> It would be wrong to say that the principle of the excluded third is false because that would mean that it implies contradiction. Now, it is not contradictory that ["There is an exceptional number" *(g.u.)*] or ["There is no exceptional number" *(g.u.)*] is true; we have only observed that in the present state of science there is no reason to affirm one or the other. This observation does not constitute a theorem of logic, just as the observation that a certain mathematical problem is unsolved does not constitute a mathematical theorem. (Heyting 1956: 232)[12]

Summing up, Dummett's criterion of semantical objectivism is not reliable, because the outcome of its application depends on how the notion of truth, and consequently the notion of negation, are conceived.

References

Brandom, R. (1976). Truth and assertibility. *The Journal of Philosophy, 73*(6), 137–149.

Brouwer, L. E. J. (1949). Consciousness, philosophy and mathematics. In *Proceedings of 10th international congress on philosophy*, Amsterdam, 1235–1249. (Now in Brouwer 1975, 480–494).

Brouwer, L. E. J. (1975). *Collected works, vol. I: Philosophy and foundations of mathematics*. In A. Heyting (Ed.). North Holland.

Cohen, M. R., & Nagel, E. (1934). *An introduction to logic and scientific method*. Routledge and Kegan Paul.

Dales, H. G. & Oliveri, G. (Eds.). (1998). *Truth in mathematics*. Oxford University Press.

Dummett, M. (1973). *Frege: Philosophy of language*. Duckworth.

Dummett, M. (1975). The justification of deduction. In *Proceedings of the British Academy, LIX*, 201–231, (pp. 290–318). (Now in Dummett (1978).)

12 «Il serait erroné de dire que le principe du tiers exclu soit faux car cela signifierait qu'il impliquerait contradiction. Or il n'est pas contradictoire que ["Il existe un nombre exceptionnel" (*g.u.*)] ou ["Il n'existe pas de nombre exceptionnel" (*g.u.*)] soit vrai; nous avons seulement constaté qu'en l'état actuel de la science il n'y a aucune raison pour affirmer l'un ou l'autre. Cette constatation ne constitue pas un théorème de la logique, tout comme la constatation qu'un certain problème mathématique n'est pas résolu, ne constitue pas un théorème mathématique.» Heyting calls "exceptional" a number n such that $n - 1$, n is not prime and n is not the sum of two or three primes.

Dummett, M. (1978). *Truth and other enigmas*. Duckworth.
Dummett, M. (1982). Realism. *Synthese, 52*(1), 55–112. (Reprinted in Dummett (1993) (pp. 230–276)).
Dummett, M. (1993). *The seas of language.* Oxford University Press.
Dummett, M. (1998). Truth from the constructive standpoint. *Theoria, LXIV*, 122–138.
Dummett, M. (2000). *Elements of intuitionism*, Clarendon Press. (First Edition 1977).
Frege, G. (1918). Thoughts. In Frege (1984) (pp. 351–372).
Frege, G. (1984). *Collected papers on mathematics, logic, and philosophy.* In B. Mc Guinness (Ed.). Blackwell.
Heyting, A. (1931). Die intuitionistische Grundlegung der Mathematik. *Erkenntnis, 2*, 106–15. (Engl. tr. The intuitionist foundations of mathematics. In P. Benacerraf & H. Putnam (Eds.), *Philosophy of mathematics* (pp. 52–61). Prentice-Hall, 1983[2].)
Heyting, A. (1956). La conception intuitionniste de la logique. *Les Études Philosophiques, 2*, 226–233.
Heyting, A. (1958). On truth in mathematics. *Verslag van de plechtige viering van het honderdvijftigjarig bestaan der Koninklijke Nederlandse Akademie van Wetenschappen* (pp. 277–279). North Holland.
Martin-Löf, P. (1985). On the meaning and justification of logical laws. In C. Bernardi & P. Pagli (Eds.), *Atti degli incontri di logica matematica*, vol. II (pp. 291–340). Università di Siena. (Reprinted in *Nordic Journal of Philosophical Logic*, *1*(1), 1996, 11–60).
Martin-Löf, P. (1991). A path from logic to metaphysics. In G. Sambin & G. Corsi (Eds.), *Atti del congresso* Nuovi problemi della Logica e della Filosofia della Scienza, vol. 2 (pp. 141–149). CLUEB.
Martínez, C., Rivas, U., & Villegas-Forero, L. (Eds.). (1998). *Truth in perspective.* Ashgate.
Martino, E. & Usberti G. (1994). Temporal and atemporal truth in intuitionistic mathematics. *Topoi, 13*(2), 83–92. (Now in Martino (2018) (pp. 97–111)).
McGuinness, B. & Oliveri, G. (Eds.). (1994). *The philosophy of Michael Dummett*. Kluwer.
Prawitz, D. (1980). Intuitionistic logic: A philosophical challenge. In G. H. von Wright (Ed.), *Logic and philosophy* (pp. 1–10). Nijhoff.
Prawitz, D. (1987). Dummett on a theory of meaning and its impact on logic. In B. Taylor (Ed.), *Michael Dummett: Contributions to philosophy* (pp. 117–165). Nijhoff.
Prawitz, D. (1994). Meaning theory and anti-realism. In McGuinness & Oliveri (1994) (pp. 79–89).
Prawitz, D. (1998a). Truth and objectivity from a verificationist point of view. In Dales & Oliveri (1998) (pp. 41–51).
Prawitz, D. (1998b). Truth from a constructive perspective. In Martínez, Rivas & Villegas-Forero (1998) (pp. 23–35).
Prawitz, D. (1998c). Comments on the papers. *Theoria, LXIV*, 283–337.
Prawitz, D. (2005). Logical consequence from a constructivist point of view. In S. Shapiro (Ed.), *The Oxford handbook of philosophy of mathematics and logic* (pp. 671–695). Oxford University Press.

Chapter 7
Belief, Synonymy, and the *De Dicto/De Re* Distinction

Abstract The purpose of this chapter is to define the notion of justification for doxastic reports, sentences of the form "$\mathcal{S}$ believes that α". What makes the problem particularly complex is the presence, in doxastic reports of natural languages, of a well-known ambiguity tracing back to Aristotle, and called in the Middle Age the *De Dicto/De Re* ambiguity ; it is therefore necessary to analyze preliminarily this ambiguity. In the Introduction (Sect. 7.1) it is argued that the De Dicto/De Re ambiguity conceals in fact two different ambiguities and distinctions: the Transparent/Opaque (TO) one and the Epistemic Specific/Non-specific (ESN) one. Section 7.2 is devoted to the TO ambiguity; in Sects. 7.2.1–7.2.5 it is argued that the foundational puzzles concerning it do not admit an optimal solution within the framework of externalist semantics; in Sects. 7.2.6–7.2.10 the distinction is analyzed and formally represented, within the framework of the internalist semantics outlined in Chaps. 4 and 5, as concerning not two kinds of belief but two different propositions semantically expressed—for the Believer and for the Reporter, respectively—by the subordinate clause of the belief report. In Sect. 7.3 a solution to the Paradox of Analysis is suggested. Sect. 7.4 is devoted to the ESN distinction; in Sect. 7.4.1 it is argued that it cannot be represented in terms of scope; in Sect. 7.4.2 a distinction is introduced between to kinds of cognitive states which can serve as justifications for sentences of the form ∃xα; in Sect. 7.4.3 the distinction is connected to the one between the assertibility conditions, within intuitionistic logic, of ∃xα and ¬∀x¬α; in Sect. 7.4.4 and in Sect. 7.4.5 the distinction is used to account for the ESN ambiguity.

Keywords Belief-reports · Propositional attitudes · De dicto/de re · Specific/non-specific · Synonymy · Mates' puzzle · Frege's puzzle · Kripke's puzzle · Quine · Internalist semantics · Paradox of analysis · Cognitive states · Intuitionism

My purpose in this chapter is to define the notion of justification for doxastic reports, sentences of the form "$\mathcal{S}$ believes that α". What makes the problem particularly complex is the presence, in doxastic reports of natural languages, of a well-known ambiguity tracing back to Aristotle, and called in the Middle Age the *De Dicto/De*

G. Usberti, *Meaning and Justification. An Internalist Theory of Meaning*, Logic, Epistemology, and the Unity of Science 59,
https://doi.org/10.1007/978-3-031-24605-0_7

Re ambiguity; it is therefore necessary to analyze preliminarily this ambiguity. In the Introduction (Sect. 7.1) I argue that it conceals in fact two different ambiguities and makes therefore necessary to introduce two different distinctions: the Transparent/Opaque (TO) one and the Epistemic Specific/Non-specific (ESN) one . Section 7.2 is devoted to the TO distinction; in Sects. 7.2.1–7.2.5 it is argued that the foundational puzzles concerning it do not admit an optimal solution within the framework of externalist semantics; in Sects. 7.2.6–7.2.10 the distinction is analyzed and formally represented, within the framework of the internalist semantics outlined in Chaps. 4 and 5, as concerning not two kinds of belief but two different propositions semantically expressed by the subordinate clause of the belief report, for the reporter and for the believer, respectively. In Sect. 7.3 a solution to the Paradox of Analysis is suggested. Section 7.4 is devoted to the ESN distinction; in Sect. 7.4.1 it is argued that it cannot be represented in terms of scope; in Sect. 7.4.2–7.4.5 it is analyzed and formally represented by taking advantage of the difference, within the present internalist semantics, between the meanings of $\exists x\alpha$ and $\neg\forall x\neg\alpha$.

7.1 Introduction. The *De Dicto/De Re* Distinction

In 1956 Quine discussed sentences containing 'intensional' verbs like "hunt", "want", "seek", "believe", "wish", and exhibiting an intuitive ambiguity between two readings called by him *notional* and *relational*. This notional/relational ambiguity has been soon connected to, or identified with, the much older *De Dicto/De Re* ambiguity of modal propositions, already observed by Aristotle and characterized by Thomas Aquinas in the following terms:

> A modal proposition is either *de dicto* or *de re*. A modal proposition *de dicto* is one in which the whole *dictum* is the subject and the modal is the predicate, as when it is said 'For Socrates to run is possible'. A modal proposition *de re* is one where the modal is interpolated in the *dictum*, as when it is said 'For Socrates it is possible to run'. (Thomas Aquinas, 1976)[1]

7.1.1 The Traditional Formulation

In Quine (1956) Quine distinguishes two intuitive senses of the report

(1) Ralph believes that someone is a spy

according to the first, *notional*, one Ralph simply believes that there are spies, while according to the second, *relational*, one Ralph believes, of some specific person, that

[1] «Propositionum autem modalium quaedam est de dicto, quaedam est de re. Modalis de dicto est, in qua totum dictum subiicitur et modus praedicatur, ut Socrates currere est possibile; modalis de re est, in qua modus interponitur dicto, ut Socratem possibile est currere.» English translation by C. Dutilh Novaes, in Dutilh Novaes (2004).

(s)he is a spy. Quine represents this *notional/relational ambiguity* in terms of the scope of the existential quantifier:

(2) Ralph believes that ∃x (x is a spy)

represents the notional sense of (1)—phrased in English as "Ralph believes that someone is a spy" or as "Ralph believes that there are spies", while

(3) ∃x (Ralph believes that x is a spy)

represents the relational sense—phrased in English as "There is someone Ralph believes to be a spy" or "Ralph believes of someone that (s)he is a spy".

As I said, the notional/relational ambiguity has been connected to the *De Dicto/De Re*ambiguity; for reasons that will become clear in a moment, I shall reserve the name "notional/relational distinction" for the *intuitive* distinction Quine had in mind, and "*De Dicto/De Re*" for the *theoretical* distinction generated by Quine's idea of representing the intuitive distinction in terms of the scope ambiguity of the existential quantifier. With this terminology, we can say that Quine held that the distinction was the *explicans* of the notional/relational distinction.

Quine remarks that «the difference [between the two senses] is vast» (Quine, 1956: 184); on the other hand, it is intuitively clear that between the notional and the relational sense there is some relation; as a consequence the Intuitive Question naturally arises:

(IQ) When does the *notional* formulation imply the *relational* formulation?

If Quine's idea that the *De Dicto/De Re* distinction is the *explicans* of the notional/relational distinction is accepted, (IQ) is to be made precise in the following way, which I shall call the Traditional Formulation of the problem:

(TF) When does the *De Dicto* formulation imply the *de re* formulation?

whose precise formulation should be:

(4) When is the following inference correct[2]:

$$(\exists\text{Exp})\,\frac{\mathcal{S}\text{ believes that }\exists x P(x)\quad \forall x((\mathcal{S}\text{ believes that }P(x)) \rightarrow E(x))}{\exists x(\mathcal{S}\text{ believes that }P(x))}\,?$$

As a matter of fact, however, (4) has been replaced by (5) as a precise formulation of (TF):

(5) When is the following inference correct[3]:

[2] "E(x)" is to be read as "x exists". a definite

[3] See for example Kripke (2011),323. In the case t is a definite description "E(t)" means the the relevant existence and uniqueness conditions are satisfied. Presumably, one reason for this replacement was the reference, in many discussions of (TF), to examples involving definite descriptions (such as "Ralph believes that the man in the brown hat is a spy"), with an implicit presupposition that definite descriptions are singular terms.

$$\frac{\mathcal{S}\text{ believes that P(t)}\quad \text{E(t)}}{\exists \text{x}(\mathcal{S}\text{ believes that P(x))}}?$$

The answer to (5) depends on the answer to (6):

(6) When is the following inference correct:

$$\text{(t Exp)}\frac{\mathcal{S}\text{ believes that P(t)}\quad \text{E(t)}}{\mathcal{S}\text{ believes of t that it is P}}?$$

As a consequence, the Traditional Formulation of the problem has been identified with (6).

7.1.2 A Distinction Between Two Distinctions

I dissent both from the replacement of (4) with (5) and from (TF). First, it is evident that (5) is not equivalent to (4): an answer to (4) implies an answer to (5), but the converse does not hold; it would hold if from

(i) $\mathcal{S}$ believes that $\exists$xP(x)

it could be correctly inferred

(b) $\mathcal{S}$ believes that P(t),

for some specific t; for in such a case E(t) would be inferred from (ii) and the second premise of the inference in (4), hence, if we had an answer to (4), we would have an answer to (5). But (ii) cannot be inferred from (i): it may happen that one believes that someone is a spy (that there are spies) without believing of anyone that (s)he is a spy.

It might be objected that it is impossible that a subject believes the proposition expressed by an existential sentence without believing any of its instances. But this is false. For example, every classical mathematician believes the proposition expressed by (7):

(7) $\exists x(\text{irrational}(x)\wedge\text{rational}(x^{\sqrt{2}}))$

(phrased as "there is an irrational number whose power to $\sqrt{2}$ is rational"), because there is a well known non-constructive proof of it (i.e. a proof based on the law of excluded middle), according to which the desired number is either $\sqrt{2}$ or $\sqrt{2}^{\sqrt{2}}$, but nobody believes any instance of (7), because nobody knows whether $\sqrt{2}^{\sqrt{2}}$ is rational or not.[4]

[4] As a matter of fact, $\sqrt{2}^{\sqrt{2}}$ is known to be irrational because of the Gelfond-Schneider theorem. The example should therefore be qualified by saying that nobody believed either disjunct before Gelfond-Schneider theorem. See Hindley (2015).

In conclusion, (TF) should be articulated into *two* distinct questions: (6) and (8):

(8) When is the following inference correct:

$$\text{(Part)}\frac{\mathcal{S}\text{ believes that }\exists x\alpha(x)\quad \forall x((\mathcal{S}\text{ believes that }\alpha(x))\rightarrow E(x)}{\mathcal{S}\text{ believes that }\alpha(t)}$$

for some term t? For, when we have an answer to (8) and to (6), we have an answer to (5) because we can obviously obtain the conclusion of (∃Exp) from the conclusion of (Exp) by existential generalization (Gen).

The two questions are distinct in the sense that they pose different problems. In order to characterize the former I introduce the following terminology:

> A context in which a term phrase can occur is an *opaque context* (or *referentially opaque context*) if substitution of a coreferential term phrase does not always preserve truth. [...] A context which is not (referentially) opaque is (*referentially*) *transparent* [...] Sentences in which a term phrase occurs in an opaque context are generally ambiguous; there will in general be both a *referential reading* (loosely called the "transparent reading"), on which [...] substitution of a coreferential term phrase does preserve truth, and a *nonreferential reading* (more accurately called a *not-purely-referential reading*; loosely called the "opaque reading"), on which and substitution of a coreferential term phrase may happen but often does not preserve truth. Thus a context is opaque if it *permits* of a not-purely-referential reading; but it will almost always permit a purely referential reading as well. (Partee, 1974: 833–834)

If we consider (tExp), the crucial difference between the (principal) premise and the conclusion is that t occurs in an opaque context in the former, while in the latter it occurs in a transparent context, in which susbstitution is truth-preserving; question (6) asks therefore, in fact, when does substitution of a term occurring in an opaque context with a coreferential term preserve truth; in any case, as I will argue below in this section, it concerns the TO distinction. In the case of (Part), on the contrary, we have not to do with a change of the scope of the existential quantifier in passing from the (principal) premise to the conclusion, and this strongly suggests that we are confronted with a different question, hence presumably with a distinction different from the TO one;[5] which one?

If we look in Quine's writings for an intuitive characterization of the notional/relational ambiguity that does not mention the scope of the existential quantifier, we find a couple of unequivocal indications. The passage quoted above, in which Quine (1956) remarks that the difference between (2) and (3) is vast, continues as follows: «indeed, if Ralph is like most of us, [(2)] is true and [(3)] is false.» The difference is explicitly stated in Quine (1979, 273):

> the seemingly vital contrast […] between merely believing there are spies and suspecting a *specific* person. (*emphasis added*)

[5] We will see in a moment some arguments to the effect that the two distinctions are mutually independent.

Clearly the crucial difference between merely believing that there are spies and suspecting a specific person consists in the fact that in the former case the believer has a specific person in mind, in the former he has not. No doubt Quine would reject the idea that having in mind a specific person can constitute an *explanation* of the difference between the two cases (Quine, 1979: 273); however, my purpose here is to individuate the intuitive phenomenon, not to explain it. No doubt Quine would reject also the idea that having in mind a specific person can constitute a *criterion* of the difference in question; however, in a passage of Quine (2000: 429) he comes near to giving a criterion acceptable for him:

> What was at issue, back then, was the contrast between just believing there are spies and believing of someone that he is a spy. The natural answer is that in the second case we can specify a spy.

Since one can specify a spy if and only if one has that spy in mind, it seems undeniable that the intuitive distinction Quine is referring to is the one between Epistemic Specificity and Non-specificity (ESN), as it is called nowadays.[6]

Are the TO and the ESND really different? Quine (1960: Sect. 31) holds that the two distinctions are connected to ambiguities which are mutually exclusive, and that the specific/non-specific ambiguity is nothing but the manifestation of the transparent/opaque ambiguity when what is inserted in a certain position is not a definite, but an indefinite term. Probably his view that the two ambiguities are mutually exclusive is due, as J. D. Fodor suggests,

> to the fact that in discussing specificity he employs examples containing the indefinite pronoun *someone* rather than indefinite noun phrases with some lexical content. (Fodor, 1970: 228)

but the idea that they are manifestations of the same phenomenon has also a more theoretical origin in his conviction that the failure of the Substitutivity Principle and the failure of (Gen) are two criteria for the same phenomenon of opacity:

> [I]f referential opacity is an infirmity worth worrying about, it must show symptoms in connection with quantification as well as in connection with singular terms. (Quine, 1953: 145)

But there are at least two reasons to hold that the two ambiguities are independent from one another.

First, the ESN ambiguity arises also in sentences containing, besides an indefinite noun phrase, no opaque verbs, no modals or negation, and no other quantifiers with which the indefinite could interact. The classical example is the sentence[7]

(9) A student in the syntax class cheated on the final exam.

---

[6] See Farkas (1996), von Heusinger (2011), Kamp and Bende-Farkas (2019).

[7] Fodor and Sag (1982), 355. Fodor and Sag call this ambiguity "Referential/Quantificational", and they reserve the name "Specific/Non-Specific" for a similar ambiguity in belief contexts that depends on the scope of the existential quantifier.

K. von Heusinger characterizes the difference between the two readings of (9) in the following terms:

> In the specific interpretation […] the speaker "has a referent in mind" and makes an assertion about this referent. In the non-specific reading […], the speaker just makes an assertion that the set of students in the syntax class who cheated on the final exam is not empty. (Von Heusinger, 2011: 1044)[8]

Second, when an indefinite noun phrase occurs within a belief context it is possible to exhibit four kinds of situations, each of which forces one of the following four possible combinations of readings excluding the others*:* S + O, N + O, S + T, N + T; and this possibility of cross-classification shows that the two distinctions are independent.[9] Consider for instance the belief report

(10) Bond believes that a woman entered the room.

Here are the four relevant situations.

S + O The first scene of a film of 007. Bond sees someone entering a room; several signs induce him to believe that it is Mata Hari. Smith, a spectator, explains by means of (10) the situation to his friend Jones who arrived late. Later on Bond realizes that it was his best friend disguised as Mata Hari; but, before, Smith would not have accepted the report "Bond believes that his best friend entered the room".

N + O The first scene of another film of the same cycle. Coming back to his room after a long absence, Bond notices that various objects have been displaced and finds a hairpin on the floor; he persuades himself that a woman has entered the room. Smith explains by means of (10) the situation to Jones who arrived late. Later on Bond realizes that it was a man disguised as a woman.

S + T First film. Smith thinks that the person seen by Bond entering the room is Mata Hari, and explains by means of (10) the situation to Jones. Later on they realize that Bond had at once understood that it was his best friend disguised as Mata Hari.

N + T Second film. Smith is persuaded that a woman entered Bond's room, and explains by means of (10) the situation to Jones. Later on they realize that Bond had at once understood that it was a man disguised as a woman.

If my claim that TO and ESN distinctions are different is accepted,[10] it should be clear why I dissent also from (TF) as a way of making (IQ) precise: it presupposes Quine's idea that the *De Dicto/De Re* distinction is the *explicans* of the notional/relational distinction, where the *De Dicto/De Re* distinction conflates two different distinctions, each of which should be analyzed and represented separately.

---

[8] It might be held that (9) has different uses, to be explained at a *pragmatic* level. In Sect. 7.3.1 I suggest an argument supporting the conclusion that (9) and similar cases are *semantically* ambiguous.

[9] As far as I know this point was made for the first time by Fodor (1970: 226 ff). See also Bonomi (1983), 98–100.

[10] A further reason to distinguish the two distinctions will be given in Chap. 8, Sect. 8.2.4.

An important example of the benefits obtained from a clear distinction between TO and ESN distinctions is offered by the Shortest Spy Problem.[11] Suppose that (2) is true; on the other hand, if Ralph is minimally reasonable, also (11) is true:

(11) Ralph believes that no two spies are of exactly the same height;

from (2) and (11) it follows that

(12) Ralph believes that the shortest spy is a spy,

hence, by (tExp),

(13) Ralph believes of the shortest spy that (s)he is a spy,

and this implies (3) by (Gen); so (2) implies (3), against the intuition, quoted above, that the difference between the notional and the relational sense is vast.

The Shortest Spy Problem is generally interpreted as generated by the doctrine that (tExp) is unrestrictedly valid; but this is correct only under Quine's assumption that the distinction is the *explicans* of the notional/relational distinction; if this assumption is discarded—as it is necessary to do if one accepts the claim that the TOD and the ESN distinctions are different—the possibility emerges of a new interpretation of the Shortest Spy Problem, as related to the conditions under which (Part), instead of (tExp), is valid. This interpretation will be illustrated in Sect. 7.4.5.

Let us start from question (6): when is (tExp) correct?

It is interesting to observe that Quine characterizes (1) as ambiguous with respect to the notional/relational distinction, but does it by reversing the natural order of exposition (Quine, 1956: 184): instead of starting from the intuitive ambiguity and then proposing a logical representation of the two readings, he starts from the two (semi-) formal sentences (2) and (3) and then remarks that «Both may perhaps be ambiguously phrased as [(1)]»; in this way he takes the scopal analysis of the ambiguity for granted (and invites the reader to to the same).

Exactly the same procedure Quine (1956) follows when the subordinated sentence contains not an indefinite noun phrase but a singular term, be it a proper name or a definite description. For example, when he introduces the case of Ortcutt, he asks:

> Can we say of this *man* (Bernard J. Ortcutt, to give him a name) that Ralph believes him to be a spy? (Quine, 1956: 185)

The presupposition is that the notional reading of Ralph's belief is represented by

(14) Ralph believes that Ortcutt (or: the man seen at the beach) is a spy,

and the relational reading by

(15) Ralph believes of Ortcutt (or: the man seen at the beach) that he is a spy.

---

[11] References in Sosa (1970), fn. 11. See also Kaplan (1969: 220).

I shall argue in Sect. 7.4.1 that the scopal representation of the ESN distinction is inadequate; in the case of the scopal representation of the TO distinction exemplified by (14) and (15) I do not have a similar argument, but also in this case the scopal representation is a bit of a stretch. First, the original intuition is that (14) is *ambiguous*, whereas according to the scopal approach it *unambiguously* represents the notional reading[12]; second, while in the case of a (n existential) quantifier the notion of its scope is perfectly meaningful, in the case of a singular term of a first order logical language it is not: we must artificially introduce some scope indicator, playing a role analogous to "of" in the paraphrase (15).

If we give up Quinean regimentations and stick at the original intuition that (14) (like (1)) is ambiguous, we have as a consequence that belief, *all* belief, is a relation between subjects and propositions[13]; how can we account for the TO distinction, under this assumption? This question will be the object of the next section.

7.2 The Transparent/Opaque Distinction

This section is articulated into three parts. The first is of a methodological nature; Sect. 7.2.1 introduces the foundational puzzles: Frege's, Mates', and Kripke's; Sect. 7.2.2 expounds reasons supporting two principles the puzzles show to be incompatible: the Substitutivity Principle (SP) and the Disquotational Principle (DP); Sect. 7.2.3 elucidates the notion of optimal solution of the puzzles; Sect. 7.2.4 explains the tension between the search for an optimal solution and the fundamental intuition (EDS), that in every language there are different synonymous expressions. The second part is constituted by Sect. 7.2.5, where it is argued that, if one wants to comply with (EDS), an optimal solution does not exist within the framework of externalist semantics. In the third part an account of the TO distinction is proposed; more specifically, Sect. 7.2.6 illustrates the internalist consequences of the choice of basing formal semantics on the relative notion of synonymy-for-a-subject ($\mathcal{S}$-synonymy); in Sect. 7.2.7 an internalist semantics for $\mathcal{L}_{\mathrm{Bel}}$ is developed; in Sect. 7.2.8 a correspondence is illustrated between a natural ambiguity in $\mathcal{L}_{\mathrm{Bel}}$ (the reporter/believer ambiguity) and the transparent/opaque ambiguity; in Sect. 7.2.9 a necessary reformulation of question (6) is explained; in Sect. 7.2.10 a solution to the foundational puzzles is proposed.

---

12 This assumption is made explicit by Kripke (2011: 323).

13 An obvious objection is that it is plausible that animals too have beliefs, even if they have not a language to espress propositions; but propositions, as they have been characterized in Chap. 4, are not necessarily expressed by the sentences of a language.

7.2.1 The Foundational Puzzles

I will assume without discussion that belief is a relation between subjects and propositions, meaning by "proposition" what a sentence says, its content; since I shall not take into consideration demonstratives, I will occasionally identify propositions also with the meanings of sentences.

Consider the following arguments:

(16) (a) $\mathcal{S}$ believes that 9 is greater than 7

(b) 9 is the number of planets
(c) $\mathcal{S}$ believes that the number of planets is greater than 7;

(17) (a) $\mathcal{S}$ believes that Sophia Loren is an actress

(b) Sophia Loren is Sofia Scicolone
(c) $\mathcal{S}$ believes that Sofia Scicolone is an actress;

(18) (a) $\mathcal{S}$ believes that John is a man

(b) For every x, x is a man iff x is a featherless biped
(c) $\mathcal{S}$ believes that John is a featherless biped;

(19) (a) $\mathcal{S}$ believes that John is an unmarried man

(b) The concept UNMARRIED MAN is the concept BACHELOR
(c) $\mathcal{S}$ believes that John is a bachelor;

(20) (a) $\mathcal{S}$ believes that Paul is a doctor

(b) The concept DOCTOR is the concept PHYSICIAN
(c) $\mathcal{S}$ believes that Paul is a physician.

They have in common the fact that the conclusion (c) is inferred from the premisses (a) and (b) by means of one of the following *Substitutivity Principles*[14]: the *Substitutivity Principle for Singular Terms*

(SPST) if $\vDash \alpha[\tau]$
and $\vDash \tau = \tau'$
then $\vDash \alpha\left[\tau'//\tau\right]$.

and the *Substitutivity Principle for Predicates*

(SPP) if $\vDash \alpha[\pi]$
and $\vDash \pi \equiv \pi'$
then $\vDash \alpha\left[\pi'//\pi\right]$

When it is not necessary to distinguish these two principles, I shall call them collectively *the Substitutivity Principle* (SP).

---

[14] This is true under the assumption that descriptions are singular terms. If they are treated otherwise (for instance as quantifiers) the analogy should be described in a more complicated way. For an excellent analysis see Neale (2000).

A basic question about arguments (16)–(20) is whether they are valid. Since they result from the application of a single principle to a single kind of context, which I shall call *doxastic*, our question can be rephrased as follows: is (SP) valid in doxastic contexts? *Frege's puzzle* consists in the fact that we face here a dilemma.[15] On the one hand, the arguments seem to be valid since (SP) is a fundamental principle; on the other hand, in the arguments (16)–(20) «it is not permissible—to quote Frege's words—to replace one expression in the subordinate clause by another having the same customary meaning» (Frege, 1892: 166).

If we consider the arguments listed above, we can observe that some of them admit a reformulation in which the premise (b) is much stronger. For example, in (19) and (20) the premises (b) could be replaced by

(19′) (b) Necessarily, the concept UNMARRIED MAN is the concept BACHELOR
(20′) (b) Necessarily, the concept DOCTOR is the concept PHYSICIAN

or even by

(19″) (b) "Unmarried Man" is synonymous with "Bachelor"
(20″) (b) "Doctor" is synonymous with "Physician".

When (19) and (20) are rephrased in the latter way, what we obtain are instances of the central problem related to the notion of synonymy, *Mates' puzzle*.[16]

Neither Frege nor Mates say why the replacement of one expression with another having, respectively, the same (ordinary) reference and the same (ordinary) meaning is not permissible. The question concerns the status of the judgement that, in appropriate circumstances, the (a)- and (b)-sentences in the above arguments are true and the (c)-sentences false: is it simply a commonsense judgement, or can it be defended on principled grounds?

A suggestion comes from Kripke when he introduces his celebrated puzzle about belief. Consider for instance (17); we can suggest that the reason why we consider (17) (a) true and (17) (b) false is that, in the great majority of cases, a subject will assent to "Sophia Loren is an actress" and will not assent to "Sophia Scicolone is an actress". The principle invoked here is the following *Disquotational Principle*[17]:

(DP) A sincere, reflective, rational subject $\mathcal{S}$, who understands "α", believes that α (i.e. believes the proposition expressed by "α") if, and only if, $\mathcal{S}$ is disposed to assent to "α".

We can therefore say that Frege's and Mates' puzzles reveal the existence of a tension between (SP) and (DP): if we accept both principles, we can conclude that the same subject believes and does not believe the same proposition.

[15] See Frege (1892), 166. The dilemma is usually called "Frege's puzzle about doxastic reports", to distinguish it from Frege's puzzle about identity statements. Since I will not be concerned with the latter, I shall call the former simply "Frege's puzzle".

[16] On the relations between Frege's Puzzle and Mates' Puzzle see for instance Soames (1987a), 108, and Moffett (2002).

[17] Soames (2002), 9, Kripke (1979), 248.

Kripke's Puzzle reveals an analogous tension between (DP) and another intuitively valid principle. Pierre is a French speaker who lives in France; he forms in some way the belief that London is pretty, which he expresses by asserting, and accepting, "Londres est jolie"; by (DP) (right to left) for French we can say that

(21) Pierre croit que Londres est jolie.

Later on he moves to London, learns English by 'direct method' and lives in an unattractive part of the city; he therefore forms a belief which he expresses by saying, and accepting, "London is not pretty"; by (DP) (right to left) for English we can say that

(22) Pierre believes that London is not pretty.

It is also true that Pierre has *no* inclination to assent to "London is pretty"; from this, by (DP) (left to right) for English we can infer that

(23) Pierre does not believe that London is pretty.

At this point we can observe that the English translation of (21) is

(24) Pierre believes that London is pretty;

applying the *principle of translation* (Kripke, 1979: 250)

(TP) If a sentence of one language expresses a truth in that language, then any translation of it into any other language also expresses a truth (in that other language),

we obtain that (24) is true, in contradiction with (23).

Since (DP) is incompatible not with one, but with two intuitively evident principles, the obvious solution seems to be to discard it. In fact, Kripke's puzzle has been presented as «an apparent reductio ad absurdum» of (DP) (Soames, 2002: 11). From this standpoint, those who, invoking (SP), consider (16)–(20) valid are the supporters of a scientific theory of belief, whereas those who, invoking (DP), consider (16)–(20) invalid are sceptical about the possibility of a scientific approach.

However, not only (SP), but also (DP) is supported by solid reasons ; moreover, as we will see in the next section,[18] there are serious reasons for questioning the validity of (SP) in doxastic contexts which are independent of (DP).

7.2.2 Reasons for Substitutivity and Disquotation

An important reason supporting (SP), not explicitly stated but often implicitly advocated in many arguments, is the following. (SP) is an immediate consequence of the

[18] See alsofn. 29.

Compositionality Principle (Comp)[19]; for, whatever the semantic value of an expression is, if (Comp) is true then the result of the substitution, in a sentence (a complex expression) of a term (resp., predicate) with a term (resp., predicate) having the same semantic value cannot but have the same semantic value as the original sentence. Hence, the reasons supporting (SP) are essentially the same as the ones for (Comp); and there are several reasons supporting (Comp).[20]

However, there are serious reasons supporting (DP) as well. It is legitimate to assume that the study of belief is a part of psychology, exactly like the study of vision, of object perception and of linguistic competence. It is a matter of fact, or at least a largely shared opinion, as we have seen in Chap. 1, that the best available theories of these domains are computational-representatational theories. In a computational-representatational theory input and output of a computation must be *representations*, complex *symbols* of a representational system,[21] and a computation is a process involving states, where transitions from one state to another are governed by an algorithm or program. For theories of this kind, the subject's answers in controllable experimental conditions are an essential source of data. In many cases such answers consist in expressing assent to, or dissent from, sentences; for instance, in a linguistics paper one typically finds examples of this sort[22]:

(25) * Voici l'homme que je crois t que t pourra nous aider l'année prochaine.
(26) Voici l'homme que je crois qui t pourra nous aider l'année prochaine.

The asterisk in (25) records the fact that a French native speaker dissents from the sentence "(25) is a good sentence", whereas the absence of asterisk in (26) records the fact that the speaker assents to the sentence "(26) is a good sentence"; and those behaviors are taken as indicating, respectively, that the speaker believes that (25) is not a good sentence, and that the speaker believes that (26) is a good sentence. These are considered as linguistic *data*, and the linguist's goal is to build an explanatory theory of them.

So, an important part of scientific practice in psychology and related fields consists in taking the subject's reactions to a sentence as indicating the fact that the subject believes the proposition expressed by that sentence.[23] Of course, an affirmative response of $\mathcal{S}$ to α cannot be taken as a *conclusive* evidence for the fact that $\mathcal{S}$ believes that α, since belief is

> a theoretical construct, not definable in terms of overt behavior, be it linguistic or non-linguistic. (Carnap, 1954: 231)

---

[19] See Sect. 4.3.3 of Chap. 4.

[20] See Szabó (2020), Sect. 3. For another reason supporting (SP) see Salmon (1986), 80 f.; but see also Cartwright (1971), 119–133.

[21] "Representation" is used here in a technical sense explained in Chap. 1: «there is nothing 'represented' in the sense of representative theories of ideas, for example.» (Chomsky 2000: 173).

[22] The example is taken from Rizzi (2000: 279).

[23] Other obvious examples are polls, from which reliable data about people's various kinds of belief are extracted.

But this is true of any empirical theory, and does not constitute an objection against (DP); it is rather an invitation to reformulate (DP) in a way appropriate to empirical theories. Here is a possible reformulation:

(DP*) If $\mathcal{S}$ is a sincere, reflective, rational, attentive subject who understands "α", the best explanation of the fact that $\mathcal{S}$ assents to "α" is that $\mathcal{S}$ believes that α (i.e. believes the proposition expressed by "α"); and the best explanation of the fact that $\mathcal{S}$ dissents from "α" is that $\mathcal{S}$ does not believe that α.

Of course there are vague, border, and unclear cases, in which it is not easy to attain a reflective equilibrium between theoretical and descriptive requirements; for example, the grammar proposed by a linguist may predict that a sentence with three layers of nested relative clauses is grammatical, whereas a native speaker may be uncertain about accepting it, or even refuse it. In such cases it is not obvious that the theory is (descriptively) inadequate; on the contrary, its adequacy can be defended, provided that it is capable to give a systematic and non-ad hoc account of the discrepancies between its predictions and the data. The exact meaning of "systematic" and "non-ad hoc" may vary with the example considered, but, given a specific case, it is in general not difficult to recognize an account as having those features. For instance, the linguist may stress the presence, in the grammar (s)he is propounding, of strong idealizations about the user's memory, attention, etc., and explain the discrepancies in terms of limitations of memory, attention, etc. actually present in empirical language-users; this would be an example of systematic account, which is not ad hoc in the sense that the necessity of idealizations can be justified on independent grounds. Concluding, the existence of cases in which it is difficult to find a point of reflective equilibrium is not a reason for abandoning (DP*); on the contrary, assent and dissent seem to be reliable guides to what is and to what is not believed.

There is another aspect of the question that should be considered. When one tries to develop a scientific theory of belief, one is making reference to a kind of belief that either a subject is *justified* in having, or an observer is *justified* in attributing to a subject: a belief lacking either kind of justification would simply not admit a *theory*.[24] Justified belief (in either sense) can be analyzed in different ways; one of them consistes in conceptually articulating it into two components: the relation of belief between a subject and a proposition, and the justifications the subject has for that proposition.[25] Now, *if* we analyse justified belief in this way, problems similar to the ones illustrated above concerning belief arise in relation to justifications to believe. For example, we can wonder whether the following arguments, analogous to (16) and (19), are valid:

(16′)
(a) $\mathcal{S}$ has justifications to believe that 9 is greater than 7

[24] Throughout this chapter I shall use "belief", "believe" etc. to refer to justified belief, to justifiedly believing, and so on.

[25] Below I shall put forward a proposal about how justifications to believe should be conceived. An argument for analysing justified belief in terms of justifications can be found in Usberti (2004).

(b) 9 is the number of planets
∴ (c) $\mathcal{S}$ has justifications to believe that the number of planets is greater than 7;
(19′)
(a) $\mathcal{S}$ has justifications to believe that John is unmarried
(b) For every x, x is unmarried iff x is a bachelor
∴ (c) $\mathcal{S}$ has justifications to believe that John is a bachelor.

We cannot answer these questions unless we have defined a notion of justification *for* an arbitrary sentence. This notion certainly has intuitive aspects, but at the same time has behind it a long tradition of conceptual analyses and of attempts to tranform it into a theoretical notion. For example, if α is a mathematical sentence, a justification to believe that α is a proof of α, and there are in the literature several attempts to define a notion of proof capable to be taken as a basis for different philosophical and mathematical enterprises (proof theory, intuitionism, formal epistemology, etc.). In general, therefore, it will be necessary to define a *theoretical* notion of justification for α, in terms of which a version of (DP*) appropriate to justified belief becomes possible:

(DP_j) Assuming that $\mathcal{S}$ is a sincere, reflective, rational, attentive subject who understands "α", if $\mathcal{S}$ assents to "α" then a reporter $\mathcal{R}$ is justified in asserting that $\mathcal{S}$ believes that α (i.e. believes the proposition expressed by "α"); and if $\mathcal{S}$ dissents from "α" then $\mathcal{R}$ is justified in asserting that $\mathcal{S}$ does not believe that α.

Scott Soames quotes an example of Mark Richard's to support the claim that «dissent is not reliable in this way», i.e. as a guide to what is not believed.[26] I assume, for the sake of the argument, that Richard's example does present a case in which a subject dissents from a sentence expressing a proposition (s)he in fact believes. The point is that such an example does not support the claim. That dissent is not a *reliable* guide to what is not believed means that it is false that *most cases* in which a subject $\mathcal{S}$ dissents from a sentence "α" are cases in which $\mathcal{S}$ does not believe that α: to exhibit *one* case in which $\mathcal{S}$ dissents from "α" but believes that α is not sufficient. It would be sufficient if that case were representative of the majority of cases; but Richard's example is not representative at all: it is a case in which the reporter $\mathcal{R}$ of a belief report coincides with the subject $\mathcal{S}$ of belief, but $\mathcal{R}$ does not know himself to be $\mathcal{S}$—not a frequent, nor a typical, case; certainly not a case capable to discredit scientific practices like the ones mentioned above, based on the reliability of assent and dissent as guides to what is and to what is not believed.

Soames generalizes Richard's example in order to make it representative of a vast class of cases. I shall discuss his generalization in a moment. Before that I want to come back to a question raised above: the question whether the judgement that, in appropriate circumstances, the (c)-sentences in (16)–(20) are false can be defended on principled grounds. I have mentioned (DP) as a principle that can be invoked to support that judgement; but there is another important principle. Consider the

[26] Soames (1987b: 217–218). However, according to Soames (2002), 10, the central objection to (DP_j) is represented by Kripke's puzzle. I shall try to answer this puzzle in Sect. 10.

following case.[27] Claude is an anthropologist studying the population of a Borneo village; he sees that $\mathcal{A}$ always manifests a great respect to $\mathcal{B}$, the village shaman, when he meets him dressed up for some ceremony, his face hidden by the ritual mask; on the contrary he sees $\mathcal{A}$ showing contempt of $\mathcal{C}$, a very poor man living in a hut at the border of the village; as a matter of fact $\mathcal{C}$ is the same person as $\mathcal{B}$, as Claude discovers shortly after. Claude wonders about the reasons of such a contrasting behaviour towards the same person, and he formulates several alternative explanations, among which the one expressed by the following sentences:

(a) $\mathcal{A}$ believes that $\mathcal{B}$ is respectable
(b) $\mathcal{A}$ believes that $\mathcal{C}$ is not respectable
(c) $\mathcal{A}$ does not know that $\mathcal{B} = \mathcal{C}$.

Claude interrogates various inhabitants of the village, and verifies that (c) is true; at this point he chooses the hypothesis expressed by the conjunction of (a) and (b) as the best explanation of the data, and infers that (a) and (b) are true. Finally, from the truth of (b) and from considerations about $\mathcal{A}$'s rationality based on observations of his behaviour, he infers that the sentence

(d) $\mathcal{A}$ believes that $\mathcal{C}$ is respectable

is false. Claude has effected an abductive inference, perfectly legitimate in empirical sciences; in this case the best explanation of the given data entails that (a) is true and (d) is false, where (d) is obtained from (a) by replacing the name "$\mathcal{B}$" with the coreferential "$\mathcal{C}$".

If we examine more closely Claude's reasoning, we realize that, before applying abduction, he has implicitly used the following principle

(*) (i) It is possible that $\tau = \tau'$ and $\mathcal{S}$ does not believe that $\tau = \tau'$;

(ii) If $\tau = \tau'$ and $\mathcal{S}$ does not believe that $\tau = \tau'$, then "$\mathcal{S}$ believes that $\alpha[\tau]$" does not entail "$\mathcal{S}$ believes that $\alpha[\tau'//\tau]$";

for, Claude has first verified that $\mathcal{A}$ does not know that $\mathcal{B} = \mathcal{C}$; from this he has inferred, by (*), that it is possible that $\mathcal{A}$ believes that $\mathcal{B}$ is respectable and $\mathcal{A}$ does not believe that $\mathcal{C}$ is respectable; and only at this point has he effected his inference to the best explanation.

(*) seems unquestionable, and is normally used in psychological explanations. In the case of (16), for example, it is quite natural to explain the fact that the conclusion (c) is false by the fact that, although 9 is the number of planets, $\mathcal{S}$ may not *know* that 9 is the number of planets; analogously, in the case of (17), it is quite natural to explain the fact that the conclusion (c) is false by the fact that, although Sophia Loren is Sophia Scicolone, $\mathcal{S}$ may not *know* that Sophia Loren is Sophia Scicolone. The problem is that (*) (i) is incompatible with the validity of (SP) in doxastic contexts, at least for theories of two kinds: theories according to which, if $\tau = \tau'$

[27] Of course the following case can be seen also as a counterexample to (SP). Its interest lies in the fact that Soames seems to claim that all putative counterexamples to (SP) are cases of conflict either with (DP) or with (TP) (Soames, 1987b: 217 and 236, footnote 22); at least, this claim is essential to his argument against (DP). The case introduced in the text is of neither kind.

(where τ and τ' are singular terms) then the proposition that $\alpha[\tau]$ is the same as the proposition that $\alpha[\tau'//\tau]$; and theories according to which, if $\pi \equiv \pi'$ (where π and π' are predicates) then the proposition that $\alpha[\pi]$ is the same as the proposition that $\alpha[\pi'//\pi]$; for according to such theories it is impossible for a subject to believe that $\alpha[\tau]$ (resp. $\alpha[\pi]$) and not to believe that $\alpha[\tau'//\tau]$ (resp. $\alpha[\pi'//\pi]$).[28]

We can now come back to Soames's generalization of Richard's example. Here is his example:

> Suppose [...] that ⌜A believes that Ruth Barcan is F⌝ is true relative to a context. ⌜A believes that I am F⌝ should then be true relative to a corresponding context in which Ruth Barcan (i.e. Ruth Marcus) is the agent (where F is free of first person pronouns). Suppose, in fact, that Ruth utters the sentence in a conversation with someone who knows her as "Ruth Marcus". It would seem that this person can truly report ⌜A believes that she (pointing at Ruth) is F⌝, or even ⌜A believes that Ruth Marcus is F⌝. Thus, substitution of one coreferential name or indexical for another preserves truth value. Since there seems to be nothing special about this example, we have a general argument for [(SPST)]. (Soames, 1987b: 218)

If the argument is general, we can apply it to Claude's case. If Claude's abductive inference is correct, (a) and (b) are true, and, since (d) is false, also the sentence

(e) $\mathcal{A}$ does not believe that $\mathcal{C}$ is respectable

is true.[29] By Soames's reasoning, both "$\mathcal{A}$ believes that I am respectable" and "$\mathcal{A}$ does not believe that I am respectable" are true relative to a context in which $\mathcal{B}$ (i.e. $\mathcal{C}$) is the agent. Now suppose that $\mathcal{B}$ utters the sentence in a conversation with someone who knows him as $\mathcal{C}$; then this person can truly report "$\mathcal{A}$ believes that $\mathcal{C}$ is respectable" and "$\mathcal{A}$ does not believe that $\mathcal{C}$ is respectable"; but this is a contradiction. In conclusion, the argument is *not* general.

Summing up, there are serious reasons supporting the validity of (SP), but equally serious reasons for the validity of (DP*)/(DP_j) and of psychological explanations based on (*); on the other hand, the validity of (SP) in doxastic contexts conflicts both with (DP*)/(DP_j) and with the possibility of accounting for the validity of psychological explanations based on (*). In all cases, the conflict is not between the predictions of a theory and intuition, but between those predictions and scientific practices.

7.2.3 Optimal Solutions

If we agree that the reasons for (SP) and (DP*)/(DP_j) are equally important, an optimal solution of the conflict we have described should find a way of not giving either principle up. This is what almost all the former approaches to Frege's Puzzle have tried to do. If we keep present the initial assumption that belief is a relation between subjects and propositions, and we observe that (DP*)/(DP_j) makes reference

[28] Direct reference theories are of the first kind; almost all theories are of the second kind.

[29] Observe that no use of (DP) has been made by Claude.

to the relation of (semantical) expression between a sentence and a proposition, it is not difficult to see where a space for an optimal solution has been looked for. Consider for example arguments (16), (18) and (19): *if* we can argue that (a)-sentences of (27)–(29) express propositions different from the ones expressed by (b)-sentences,

(27) (a) 9 is greater than 7

(b) the number of planets is greater than 7

(28) (a) John is a man

(b) John is a featherless biped

(29) (a) John is an unmarried man

(b) John is a bachelor,

then we have neutralized counterexamples to either (SP) or (DP_j) (according to our philosophical taste) as being only apparent, and we can deem both principles valid. The simultaneous validity of both (SP) and (DP_j) in doxastic contexts is therefore explained by introducing a *theoretical* notion of proposition such that it turns out that sentences (a) and (b) of (27)–(29) express different propositions. Essentially, the idea consists in introducing more and more demanding criteria for the identity of propositions (hence for substitution), i.e. more and more fine-grained theoretical notions of proposition, in order to neutralize intuitive counterexamples. For example, necessary equivalence permits to neutralize counterexamples to the validity of arguments (16) and (18),[30] intensional isomorphism neutralizes counterexamples to the validity of (19) (Carnap, 1947), expressing the same Russellian proposition neutralizes counterexamples to the validity of (21).[31]

However, a solution of this kind is optimal only if, besides neutralizing counterexamples to (SP), it can be argued that the theoretical notion of proposition introduced

[30] Carnap's solution to Frege's puzzle in *modal* contexts (in Carnap, 1947) is not essentially different. It is usually described (by Carnap himself) by saying that the truth-value of a modal sentence depends (not on the extensions/denotations, but) on the intensions of the expressions occurring within the modal context; under this description, (SPST) is not valid, while the following Intensional Substitutivity Principle for Singular Terms does hold:

(ISPST) if $\vDash \alpha[\tau]$
and τ has the same intension as τ'
then $\vDash \alpha[\tau'//\tau]$.

But Carnap's solution can be described equally well by saying that the truth-value of the modal sentence "It is necessary that α" depends on the proposition *denoted* by the clause "that α"; under this description, (ISPST) is valid but irrelevant, since the proposition in question is individuated by the intensions of the expressions occurring after the "that".

[31] Cp. for instance Soames (1987b: 228), where the solution is applied to the similar argument

(a) $\mathcal{S}$ believes that "Hesperus" refers to Hesperus and "Phosphorus" refers to Hesperus.
(b) Hesperus = Phosphorus.
∴ (c) $\mathcal{S}$ believes that "Hesperus" refers to Hesperus and "Phosphorus" refers to Hesperus and, for some x, "Hesperus" refers to x and "Phosphorus" refers to x.

is such that there are no counterexamples to (DP_j) nor is it impossible to account for scientific practices based on (*). On the other hand, there are structural limits to the fine-grainedness of theoretical notions of propositions. For instance, Carnap's solution has a structural limit in the fact that, in the case of syntactically simple expressions, equi-intensionality and intensional isomorphism collapse, so that the validity of (SP) clashes with the validity of (DP_j); analogously, given that Sophia Loren *is* Sofia Scicolone, the Russellian proposition expressed by "Sophia Loren is an actress" is the same as the Russellian proposition expressed by "Sofia Scicolone is an actress", so that any solution based on the adoption of Russellian propositions is not optimal owing to the same conflict between (SP) and (DP_j).

7.2.4 Mates' Argument

This situation is further complicated by Mates' puzzle. I have already introduced the puzzle, but Mates' original argument contains in fact much more than what is traditionally called "Mates' puzzle". Here is the argument:

> Carnap has proposed the concept of intensional isomorphism as an appropriate explicatum for synonymity. It seems to me that this is the best proposal that has been made by anyone to date. However, it has, along with its merits, some rather odd consequences. For instance, let "D" and "D′" be abbreviations for two intensionally isomorphic sentences. The following sentences are also intensionally isomorphic:
>
> [(30)] Whoever believes that D, believes that D.
>
> [(31)] Whoever believes that D, believes that D′.
>
> But nobody doubts that whoever believes that D, believes that D. Therefore, nobody doubts that whoever believes that D, believes that D′. This seems to suggest that, for any pair of intensionally isomorphic sentences - let them be abbreviated by "D" and "D′" – if anybody even doubts that whoever believes that D believes that D′, then Carnap's explication is incorrect. What is more, any adequate explication of synonymity will have this result, for the validity of the argument is not affected if we replace the words "intensionally isomorphic" by the word "synonymous" throughout. (Mates, 1950: 125)

The underlying logical structure of the argument may not clearly result from Mates' formulation; it becomes more evident if we split it into two theses.[32] The former is that, if there are two intensionally isomorphic sentences δ and δ' and a subject $\mathcal{S}$ such that $\mathcal{S}$ believes that δ but does not believe that δ', then δ and δ' are not interchangeable in each sentence in $\mathcal{L}$ *salva veritat*e. The latter is that such pairs of synonymous sentences and such a subject do exist.

Mates does not give an explicit proof of the former thesis; however, Carnap gave in *Meaning and Necessity* an argument for the analogous thesis that belief-sentences are not intensional or, equivalently, that L-equivalence is not a sufficient condition of interchangeability *salva veritate*; I submit that Mates would have argued in a similar way. Here is how the Carnap-style argument would run. Let δ and δ' be two

[32] I owe the following formulation of Mates' puzzle to Casalegno (1997), 155–156.

intensionally isomorphic sentences of $\mathcal{L}$, and suppose we ask $\mathcal{S}$ whether he believes what δ and δ′ say: $\mathcal{S}$ answers «Yes» to δ, «No» to δ′: «Since we know him to be truthful, we take his affirmative or negative answer as evidence for his belief or disbelief.» (Carnap, 1947: 54) In other terms, we apply (DP). As a consequence, (32) is true and (33) is false:

(32) $\mathcal{S}$ believes that δ
(33) $\mathcal{S}$ believes that δ′,

which means that δ and δ′ are not interchangeable in $\mathcal{L}$.

As Mates remarks at the end, «the validity of the argument is not affected if we replace the words "intensionally isomorphic" by the word "synonymous" throughout». It is worthwhile to reflect on the meaning of this remark. First, Mates observes that what is shown is that «any adequate explication of synonymity will have this result»; but "adequate explication" means here a theoretical notion having the essential features of the intuitive notion; hence what is shown concerns the very *intuitive* notion of synonymy, independently of how it is defined. Second, what, exactly, is shown? Let us consider the new argument. Let δ and δ′ be two synonymous sentences, and $\mathcal{S}$ a subject who believes that δ but does not believe that δ′; then δ and δ′ are not interchangeable (because (32) is true and (33) is false), hence they are not synonymous. The last step (from non-interchangeability to non-synonymy) is justified by the following condition of adequacy for definitions of synonymy:

(34) ***Adequacy Condition***

Two expressions are synonymous in a language $\mathcal{L}$ if and only if they may be interchanged in each sentence in $\mathcal{L}$ without altering the truth-value of that sentence. (Mates, 1950: 119)

So, what Mates' former thesis asserts, concerning the intuitive notion of synonymy, is that, if there is a subject that assents to one of two synonymous sentences and not to the other, then the two sentences are not synonymous.

The latter thesis is more difficult to prove than it might appear. It might seem sufficient to exhibit two synonymous sentences like (35)

(35) (a) Paul is a doctor

(b) Paul is a physician

and to argue, by exploiting (DP), that, when δ and δ′ are, respectively, (35) (a) and (35) (b), there is a subject $\mathcal{S}$ such that (32) is true and (33) is false. But this example would be unconvincing for anyone who for some independent reason holds it impossible that (32) is true and (33) is false, and that therefore (DP) must be given up. We need a couple of sentences "δ" and "δ[ε′//ε]" such that it is beyond doubt that every subject believes that δ, and that there is a subject who does not believes that δ[ε′//ε]. Now, no one can doubt that, for every subject $\mathcal{S}$, if $\mathcal{S}$ believes that α, then $\mathcal{S}$ believes that α; so take as "δ" the following sentence:

(36) Whoever believes that α, believes that α;

on the other hand, take as $\delta[\varepsilon'//\varepsilon]$ the following:

(37) Whoever believes that α, believes that $\alpha[\varepsilon'//\varepsilon]$,

where ε and ε′ are two synonymous expressions; a subject who does not believe (or doubts) that $\delta[\varepsilon'//\varepsilon]$, thereby making (33) false, does exist: it is Mates himself, since this is exactly the content of his latter thesis. If we now observe that (36) and (37) are synonymous, since (37) is the result of replacing, within (36), "α" with the synonymous "$\alpha[\varepsilon'//\varepsilon]$", we realize that Mates, (36) and (37) have the required properties.[33]

Mates' two theses, taken together and construed as referring to the intuitive notion of synonymy, entail that if ε and ε′ are two syntactically different designators, then they are not synonymous.[34] And this conflicts with the following intuition about the existence of different synonymous expressions:

(EDS) In every language there are syntactically different synonymous expressions.

Summing up, Mates' problem reveals the presence of two conflicting intuitions concerning synonymy: (EDS) and (34); they are conflicting because (34) entails that every expression is synonymous only with itself, against (EDS). This entailment subsists if and only if either the principle (DP*) or the principle (*) is accepted. If we remark that, if a language $\mathcal{L}$ is rich enough to contain belief contexts, the adequacy condition (34) amounts to the validity of (SP) in such contexts, we can state the search for an optimal solution in the following way:

(38) **The problem**

Give an adequate definition of the meaning of "$\mathcal{S}$ believes that α" and of a theoretical notion of synonymy, i.e. a definition satisfying (SP), (DP*)/(DP_j), (*) and (EDS).

7.2.5 Answers to Mates' Argument: A Dead End

In this section I argue that, if we want to comply with the intuition expressed by (EDS), an optimal solution does not exist within the framework of externalist semantics. I argue for this claim by surveying and criticizing some of the main strategies of

[33] For these reasons I shall occasionally call Mates' argument for his second thesis "the autoreferential argument".

[34] Of course, if ε and ε′ are two tokens of the same syntactical type, "α[ε]" and "$\alpha[\varepsilon'//\varepsilon]$" are the same sentence, hence Mates' second thesis cannot be true: for every subject $\mathcal{S}$, if $\mathcal{S}$ believes that δ, then $\mathcal{S}$ believes that δ.

solution proposed in the literature to Mates' puzzle, and by extracting from the analysis a general remark.

7.2.5.1 Is Synonymy Epistemically Transparent?

I shall start from a strategy that seems to be implicit in remarks of several authors but that, as far as I know, has been actually applied only by Kripke, in footnote 46 to Kripke (1979). Consider the following question:

(39) Is it possible that a subject $\mathcal{S}$ is competent about two synonymous predicates π and π' (in the sense that $\mathcal{S}$ knows the meaning both of π and of π'), withouth knowing that they are synonymous?

Assuming that the answer is negative, it is easy to neutralize counterexamples to (SP) based on (DP_j) as cases in which the subject does not know the meaning of either π or π', hence as cases in which not all the assumptions contained in (DP_j) are satisfied, hence (DP_j) doesn't apply.[35]

Kripke elaborates this strategy. First he assumes that, if someone assents to "Jones is a doctor" but not to "Jones is a physician", then he «either does not understand one of the sentences normally, or he should be able to correct himself "on reflection"» (Kripke, 1979: 276, fn. 23).[36] Then he uses this assumption to analyze the appropriate instance of Mates' doubt (37), i.e. the proposition expressed by "Whoever believes that Jones is a doctor, believes that Jones is a physician" (Kripke, 1979: 281–282, footnote 46). What is the source of Mates' dissent from this sentence? Either Mates doesn't realize that "doctor" and "physician" are synonymous, in which case (DP_j) doesn't apply; or Mates does realize that they are synonymous, but he applies (DP_j) to a man who assents to "Jones is a doctor" but not to "Jones is a physician": a misapplication of (DP_j), since it ignores the caution implicit in the previous assumption.

The problem with this line of thought is that it is not easy to *argue* for a negative answer to (39), i.e. for the thesis that synonymy of predicates is epistemically transparent. The only argument seems to be that a negative response to (39) is a conceptual truth: if we want that a certain semantical property or relation R (for instance entailment between two sentences α and β, or synonymy of two expressions ε and ε', or the name ν's property of denoting the object o) plays a role in the explanation of

---

[35] A negative answer to (39) seems to be implicit in Paul Horwich's remarks about Mates' problem (Horwich, 1998: 100–102), according to which a subject who says "Not all πs are π's" should understand one of the two predicates imperfectly, or be «somewhat confused». However, the way out of the problem Horwich proposes goes in another direction. According to it, if π and π' are synonymous and a subject $\mathcal{S}$ says and believes that not all πs are π's, then $\mathcal{S}$ says and believes that not all πs are πs. This amounts to giving up (DP*), because $\mathcal{S}$ would of course assent to the sentence "All πs are πs", and from this it follows, by (DP*), that $\mathcal{S}$ (says and) believes that all πs are πs. The proposed way out is therefore not optimal, for the reasons explained above.

[36] Kripke does not argue for this assumption; plausibly, a rationale for it is a negative answer to (39).

linguistic or logical competence of a subject $\mathcal{S}$, it is necessary that R is *known* by $\mathcal{S}$, or at least that $\mathcal{S}$ is in a position to know R.[37] But against this pretended conceptual truth there seem to be counterexamples. One is proposed by S. Rieber:

> Assume that 'bet' and 'wager' are synonymous. […]
>
> Joan, let us say, is a normally competent English-speaker who uses the words 'bet' and 'wager' as almost everyone else does. […]
>
> However, Joan is sceptical about apparent synonymies. In the past, she often thought that certain pairs of words were synonymous, then discovered (or was shown) slight differences in meaning. Perhaps she is an analytic philosopher humbled by many counterexamples to claims about co-extensiveness. In any case, she now doubts that 'bet' and 'wager' are synonymous. It is not that she can think of a counterexample to the claim that all and only bets are wagers: she just strongly suspects that there is one.
>
> Here we have someone who understands a pair of synonyms but does not believe that they are synonymous. (Rieber, 1992: 225–226)

Another example can be fashioned through a simple modification of Kripke's puzzle. Peter has lived in Greece in two different periods of his life; during the first he knew the land where he lived under the name "Greece", during the second under the name "Hellas", and he never realized that it was the same land; he knows that "Greek" means inhabitant of Greece and that "Hellene" means inhabitant of Hellas, so he is perfectly competent about the two synonymous predicates, but he does not know that they are synonymous.[38]

A related reason to doubt that a negative response to (39) is a conceptual truth is the fact that different authors give opposite answers to (39), mostly without argument; for instance Frege, Dummett and Kripke answer No,[39] Soames, Salmon and Rieber Yes (Rieber is the sole who argues for his answer). It seems to me that one reason for these divergent opinions is that, as we have seen in Chap. 4, our intuitions about meaning, and consequently about synonymy, seem to be so different in the case of predicates and names that I have suggested different *explicantia* for synonymy (equi-intensionality and equi-extensionality, respectively). As a consequence, when the meaning of a pair of predicates is defined in terms of names—as in the case of "Greek" and "Hellene"—and we are confronted with a subject $\mathcal{S}$ who ignores that they are synonymous, we oscillate between judging $\mathcal{S}$ incompetent about (at least one of) them, because they are predicates, and judging $\mathcal{S}$ competent about (both of) them, because ignoring whether two names denote the same object is not being incompetent about their meaning. This is why Kripke's strategy of footnote 46 seems unsuccessful.

---

[37] On this point see Boghossian (1994: 42–45).

[38] See the following footnote.

[39] Frege (1918: 359), Dummett (1981: 95), Kripke (1979: 276, ffn. 23). Notice that Kripke adds that no parallel linguistic or conceptual confusion can be ascribed to someone who says "Cicero was bald but Tully was not".

7.2.5.2 Putnam

Putnam (1954) proposed a solution of Mates' puzzle based on the idea that (DP) [and presumably (DP_j)] is valid and (SP) invalid. (SPST) should be replaced by the following restricted principle

(RSPST) if $\vDash \alpha[\tau]$
and τ has the same denotation as τ'
then $\vDash \alpha[\tau'//\tau]$, provided that $\alpha[\tau]$ and $\alpha[\tau'//\tau]$ have the same logical structure,

where

> Two sentences are said to have the same *logical structure*, when two occurrences of the same sign in one correspond to occurrences of the same sign in the other. (Putnam, 1954: 122, fn. 8)

Analogously, (SPP) should be replaced by a similarly restricted principle (RSPP). The adoption of (RSPST) and (RSPP), not explicit in Putnam (1954), may be seen as a consequence of the following principle, explicitly adopted by him:

> The sense of a sentence is a function of the senses of its parts *and of its logical structure*. (Putnam, 1954: 118)

Mates' puzzle is solved because, according to this principle, (30) and (31) are not synonymous, since they have not the same logical structure.

Putnam's solution has received a lot of criticism from Soames (1987a), which I find convincing; let me mention here one more, due to Chomsky.[40] Let δ, δ' and δ'' be three different and Putnam-synonymous sentences; then the sentences (31) and

(40) Whoever believes that δ, believes that δ''

are Putnam-synonymous; but it cannot be excluded that everybody believes (31) and somebody does not believe (40); therefore (31) and (40), although Putnam-synonymous, are not interchangeable within the context "Whoever believes that…".

The most common strategy is to refuse Mates' latter thesis, and to deny that there are a pair of synonymous sentences δ and δ' and a subject $\mathcal{S}$ such that $\mathcal{S}$ believes that δ but does not believe that δ'. The solutions of Carnap, Church, Sellars and Soames are all of this kind.[41]

7.2.5.3 Carnap

Here is Carnap's argument for the conclusion that, if δ and δ' are synonymous, there is not a subject $\mathcal{S}$ such that $\mathcal{S}$ believes that δ but does not believe δ'. Take δ and

[40] Cf. Scheffler (1955: 42, fn. 7), where the suggestion is attributed to Chomsky.

[41] Carnap (1954: 230–231), Church (1954: 68–71), Sellars (1955: 120), Soames (1987a: Sects. 4, 7, 8).

δ′ as abbreviations of (35) (a) and (35) (b), which are synonymous if "doctor" and "physician" are. Let us ask first how one could establish that $\mathcal{S}$ does not believe that Paul is a physician. The natural answer relies on (DP): by asking $\mathcal{S}$; but—Carnap remarks—from an affirmative or negative response of a subject to the sentence α we cannot *deduce* that he believes, or does not believe, that α: at best, we can infer it as probable.[42] Therefore, although it is possible that $\mathcal{S}$ answers «No» to the question «Is Paul a physician?», we can continue to hold that $\mathcal{S}$ does believe that Paul is a physician.

Carnap's remark is certainly correct; but, if $\mathcal{S}$'s negative response allows us to infer that it is *probable* that he does not believe that Paul is a physician, why does Carnap infer, on the contrary, that he does? This is a further assumption, essentially ad hoc, in the sense that it is justified by the very necessity to get out of Mates' trap.[43] Moreover, it conflicts with (DP_j), which is exactly the natural reformulation of (DP) induced by the considerations above, strictly similar to Carnap's own remark.[44]

7.2.5.4 Church

The solutions we are going to examine now are based on a common strategy, which might be called *reinterpretation* or *reconstruction*[45]: Mates' doubt should not be taken at face value, but should be reinterpreted. The strategy encompasses two steps: (a) an argument for the impossibility of taking Mates' doubt at face value, i.e. as being about the proposition that δ′; (b) a proposed reinterpretation.

---

[42] As observed by Carnap (1954: 230).

[43] As a matter of fact Carnap remarks that the subject's negative answer is «perhaps due to his momentary confusion» (Carnap 1954: 231), so that his response is non-indicative. This suggestion seems to me misleading, since it is not difficult to envisage a perfectly attentive and clear-headed subject accepting for instance (20) (a) and refusing (20) (c).

[44] This is a point which whoever is concerned with a scientific study of semantic competence is sensible to. Cp. for example Partee:

> But the linguist, although he may agree wholeheartedly that 'believes' is a term for whose correct application no single kind of observational evidence is criterial, is not thereby free to discount a priori whatever observational evidence happens to conflict with his favorite hypothesis. (Partee, 1973: 316)

Still another questionable aspect of Carnap's argument is that it could be used to rehabilitate simple L-equivalence (or N (for necessary)-equivalence) as an *explicans* of synonymy. For, suppose that δ and δ′ are L(N)-equivalent: the procedure by which it is ascertained that $\mathcal{S}$ believes that δ and not that δ′ is presumably the same I have just described, so $\mathcal{S}$'s negative answer permits us to infer only as probable that (7) is false; therefore we could hold (7) *true* on the basis of the fact that (6) is true and that δ and δ′ are L(N)-equivalent. What is lacking is some reason for applying the argument to intensionally isomorphic sentences and not to L(N)-equivalent ones.

[45] Cf. Burge (1978: 123). Another important solution implementing this strategy is Stalnaker's. For reasons of space I will not be able to deal with it.

Church's solution (in Church, 1954) is the first example of this sort. It consists in introducing a distinction and making four claims. The distinction is between the pair (36)–(37) and the 'metalinguistic' pair

(41) Whoever satisfies (in English) the sentential matrix "x believes that α", satisfies (in English) the sentential matrix "x believes that α"
(42) Whoever satisfies (in English) the sentential matrix "x believes that α", satisfies (in English) the sentential matrix "x believes that α[ε′//ε]";

the four claims are that (i) it is impossible for any subject to believe that δ and not believe that δ[ε′//ε]; (ii) it is *possible* to believe the proposition expressed by (41) and not believe the proposition expressed by (42); (iii) Mates' doubt (i.e., non-belief, or abstention from belief) cannot be the doubt that δ[ε′//ε]; (iv) it may be reinterpreted as the doubt concerning the proposition expressed by (42) (Church, 1954: 69). Clearly claims (iii) and (iv) are implementations of steps (a) and (b).

Church does not give explicit reasons for (i), but it is not difficult to reconstruct them: since belief is a relation with propositions, and by hypothesis "δ" and "δ′" express the same proposition, necessarily, if $\mathcal{S}$ believes that δ, then $\mathcal{S}$ believes that δ′. On the other hand, the proposition expressed by (41) is different from the proposition expressed by (42), because the two mentioned sentential matrices are different; hence it is possible to believe the former and not the latter, as (ii) says.

As for (iii), let me start from a preliminary remark. An easy way to prove it would consist in using (SP). We might observe that, since (36) and (37) are synonymous (because of the fact that α and α[ε′//ε] are), also the two sentences

(43) Mates believes that whoever believes that α believes that α
(44) Mates believes that whoever believes that α believes that α[ε′//ε]

are synonymous, so it is impossible, because of (SP), that (43) is true and (44) is false. But this would be unfair to Mates, or, better, incorrect. Mates might have observed that some subjects assent to "α" and not to "α[ε′//ε]", and from this he might have inferred (applying (DP)) that (SP) does not hold in doxastic contexts; so he would not be worried about his own violating (SP), and might conclude that, on second thoughts, "δ" and "δ[ε′//ε]" are *not* synonymous. Therefore, in order to prove that it is impossible for Mates to believe that α and not to believe that α[ε′//ε] we must *not* use (SP). The only one who has clearly seen this point is Church himself, when he writes that

> the very possibility of entertaining the doubt that [(37)], without simultaneously doubting that [(36)], shows [(37)] and [(36)] to be non-interchangeable in belief contexts. (Church, 1954: 69)

Church argues for (iii) by exploiting Langford's celebrated translation test. He takes as α and α[ε′//ε] in (36) and (37) the sentences

(45) The seventh consulate of Marius lasted less than a fortnight;
(46) The seventh consulate of Marius lasted less than a period of fourteen days,

thereby obtaining the two sentences

(36′) Whoever believes that the seventh consulate of Marius lasted less than a fortnight, believes that the seventh consulate of Marius lasted less than a fortnight;

(37′) Whoever believes that the seventh consulate of Marius lasted less than a fortnight, believes that the seventh consulate of Marius lasted less than a period of fourteen days.

Then he informs us that

> the German language has no single word which translates the word 'fortnight,' and that the literal translation of the word 'fortnight' from English into German is 'Zeitraum von vierzehn Tagen. (Church, 1954: 70)

At this point he makes what I shall call "the critical inference"; he writes:

> In consequence, the German translations of [(36′)] and [(37′)] are identical, as follows:
>
> [(36″) (37″)] Wer glaubt dass das siebente Konsulat des Marius weniger als einen Zeitraum von vierzehn Tagen gedauert habe, glaubt dass das siebente Konsulat des Marius weniger als einen Zeitraum von vierzehn Tagen gedauert habe. (Ibid.)

The conclusion is that the assumption that (43) is true and (44) is false is a contradiction in German, so it cannot be true in any other language either.

As for (iv), Church doesn't give an argument; it is formulated as a suggestion. More explicitly, the suggestion is that Mates' doubt—what Mates doesn't believe—is not that (37), but that

(47) Whoever satisfies (in English) the sentential matrix "x believes that α", satisfies (in English) the sentential matrix "x believes that α[ε′//ε]".[46]

Finally Church suggests that Mates «must have mistaken the doubt that [(47)] for the doubt that [(37)].» (Church, 1954: 71).

Church's argument for (iii) is, in my opinion, invalid because the critical inference has no justification. Which notion of translation justifies it? Surely not the intuitive notion. Consider for example Church's paper containing the argument we are considering, which is written in English, and suppose that a translator must translate it into German. When (s)he takes into consideration (45) and (46) (s)he cannot leave them in English, because, if (s)he did, many of Church's remarks would be unintelligible to his readers (who by assumption do not know English)[47]; but if (s)he

---

[46] Notice that (i) $\mathcal{S}$ satisfies the matrix "x believes that α" iff "$\mathcal{S}$ believes that α" is true; and (ii) "β iff γ" is true iff "β" is true iff "γ" is true. As a consequence, Mates' doubt that whoever satisfies the matrix "x believes that α" satisfies the matrix "x believes that α[ε′//ε]" is the doubt that (42) is true.

[47] For instance the following remark:

According to Mates, it is true that:

[(i)] Nobody doubts that whoever believes that the seventh consulate of Marius lasted less than a fortnight believes that the seventh consulate of Marius lasted less than a fortnight.

But is not true that:

[(ii)] Nobody doubts that whoever believes that the seventh consulate of Marius lasted less than a fortnight believes that the seventh consulate of Marius lasted less than a period of fourteen days. (Church, 1954, p. 69) translated (45) and (46) according to Church's prescription, her/his translations of the two sentences would be one and the same German sentence, with the consequence that Mates' position, together with the whole ensuing discussion, would become completely unintelligible. Of course, no actual translator would conform to Church's prescription; (s)he would rather change the example, so as to have at his disposal two distinct synonymous words in German, and only one in the target language; in this way (s)he would be, so to say, faithful to 'the spirit' of the original and not to the letter. More generally, «Obvious truths and falsehoods are normally used in arriving at a translation—not vice versa» (Burge, 1978: 122).

If the intuitive notion of translation does not justify Church's critical inference, maybe it can be justified by a theoretical notion. As a matter of fact, the critical inference would be justified only under the following assumptions: (i) that translation is compositional, and (ii) that distinct synonymous expressions are translated into distinct synonymous expressions, if it possible; if it is not possible, they are translated into the same expression. But the assumption that translation is a compositional operation is, in turn, utterly unjustified. For example, suppose that a book written in Italian is translated into English; the book contains the sentence

(48) Questo libro è scritto in italiano;

the compositional translation of (48) would be

but (48) is true, while (49) is false: a compositional translation does not even satisfy (TP).

(49) This book is written in Italian;

Moreover, the critical inference has in some cases unacceptable consequences. In Italian the expressions "faccia", "viso" and "volto" are synonymous; as a consequence the following sentences are synonymous in Italian:

(α) Heidegger aveva una faccia inquietante
(α′) Heidegger aveva un viso inquietante
(α″) Heidegger aveva un volto inquietante.

Now consider the sentences

(50) (a) Chiunque crede che α, crede che α

(b) Chiunque crede che α, crede che α′
(c) Chiunque crede che α, crede che α″,

and suppose that Gabriele, an Italian philosopher, doubts neither that (50) (a) nor that (50) (b), but does doubt that (50) (c). I don't know whether German has only one single word which translates the three Italian synonyms, but suppose this is the case. Now apply Church's critical inference: the translation into German of all three sentences (50) is the same. From this, however, a contradiction follows in the case of (c), *not* in the case of (b), since by hypothesis Gabriele does not doubt that (50) (b). As a consequence, the *content* of Gabriele's doubt *cannot* be the proposition expressed by (50) (c); it is rather the proposition expressed by "(50) (c) is true", according to Church's fourth claim; on the other hand, the *content* of Gabriele's belief *can* be the proposition expressed by (50) (b), and therefore it *is* that very proposition.[48] In conclusion: Gabriele's belief and Gabriele's disbelief have two different contents. This is obvious, since the in one case Gabriele believes and in the other he disbelieves. What seems absurd is that the content of his attitudes depends on his answer: before believing one thing and disbelieving the other, Gabriele has asked himself whether whoever believes that α, believes that α′, and whether whoever believes that α, believes that α″, and the *content* of each question must be definite *before* he answers the question.

Summing up, Church's implementation of the strategy of reinterpretation fails at step (a), since the argument for his claim (iii) is unconvincing; as a consequence, Church's solution conflicts with (DP_j), exactly like Carnap's one: by the standard of evaluation adopted above, it is not optimal.[49]

7.2.5.5 Sellars

Sellars (1955) gives an argument very similar to Church's,[50] but I shall stress a significant difference. Sellars considers the belief reports

(51) Jones believes that all Greeks are Greeks
(52) Jones believes that all Greeks are Hellenes,

and concerning (52) he remarks that «even as a sentence in *our* language may well have more than one employment»: we may employ the words "Greeks" and "Hellenes" either *as we normally use them*, i.e. as synonyms; or *as used by Jones*, i.e. as non-synonyms. To this distinction he adds a qualification: the former employment is a «purely *using* use», while the latter is a «covert *mentioning* use», so that (52) is equivalent, in fact, to

(53) The sentence "all Greeks are Hellenes" as used by Jones expresses something he believes.

However, Sellars gives no argument to support the qualification, and we have seen that Church's argument to the same effect is unconvincing. So what we remain with is the acknowledgment of two possible uses of expressions occurring within belief

[48] In principle it would be possible to stipulate that the content of Gabriele's belief is the proposition expressed by "(50) (b) is true"; but in that case we should apply the same stipulation also to (35) (a) and, more generally, to all belief reports; and this would be incorrect, since believing that α and believing that "α" is true are logically independent from each other.

[49] Notice that also Church's implementation of step (b)—proposing a convincing reinterpretation—is questionable; Burge has convincingly argued that Church's claim (iv) does not really provide an alternative to taking Mates' doubt at face value (Burge, 1978: 123).

[50] Cf. Sellars (1955: 117, fn. 1), Church (1954: 73, fn. 21).

reports. This is an important remark, and I shall draw on it below.[51] Sellars hastens to add that

> Clearly it is only on the former supposition that the question "Does the synonymity of [(51)] and [(52)] as sentences in our language follow from the synonymity, in our language, of 'Greek' and 'Hellene'?" is a relevant question to ask. (Sellars, 1955: 119)

But the question we are concerned with here is *not* the one mentioned by Sellars: it is—as Church points out—«the *possibility* of doubting that [(46)] without doubting that [(45)]» (Church, 1954: 69); and to answer this question both uses are relevant.

7.2.5.6 Soames

Soames's solution is similar to Church's one, in the sense that Church's first three claims are endorsed; the difference lies in the reinterpretation step: Mates is the victim of an error, but according to Soames his confusion is not between use and mention; it is rather between the proposition semantically expressed by a sentence and the proposition that sentence may be used to assert in certain contexts.

In order to justify this reinterpretation Soames introduces some semantical and pragmatical principles concerning belief ascriptions.[52]

(P_1) (Soames, 2002: 208)

If $\mathcal{R}$ is a competent speaker who assertively utters "$\mathcal{S}$ believes that α[ν]" in a context C, using it with its literal meaning (nonmetaforically, nonironically, nonsarcastically, and without a defeating conversational implicature), then $\mathcal{R}$ asserts that $[\![\mathcal{S}]\!]_C$ believes $[\![\alpha[x//\nu]]\!]C,a{:}[\![\nu]\!]_C$, where $[\![\alpha[x//\nu]]\!]C,a{:}[\![\nu]\!]_C$ is the Russellian proposition semantically expressed in C by the clause "α[x//ν]" with respect to the assignement of the denotation of ν in C to "x".

(P_2) (Soames, 2002: 221)

If $\mathcal{R}$ assertively utters "$\mathcal{S}$ believes that α[ν]" in a context C and common background information i shared by speakers and hearers in C is such that

(i) the name ν is associated by them to the description "thex(x is D and x = ν)", and as a result of this an assertive utterance of "α[ν]" in C would result in an assertion of the proposition $[\![\alpha[\text{thex}(x \text{ is D and } x = \nu)]]\!]_C$;
(ii) speakers and hearers in C will readily assume that
   (a) if $\mathcal{R}$'s assertive utterance is true in C, then $[\![\mathcal{S}]\!]$ believes $[\![\alpha[\text{thex}(x \text{ is D and } x = \nu)]]\!]_C$;

[51] To my knowledge, Sellars' is the first published acknowledgement of the opaque/transparent ambiguity.

[52] I will use the following symbols: "$[\![\alpha]\!]_C$" for the proposition semantically expressed by the sentence α in the context C; "$[\![\varepsilon]\!]_C$" for the denotation of the expression ε in C; "p", "q" for arbitrary propositions.

(b) speakers and hearers in C know that (a);

then $\mathcal{R}$ asserts that $[\![\mathcal{S}]\!]_C$ believes $[\![\alpha[\text{thex}(x \text{ is D and } x = \nu)]]\!]_C$.

Let us apply these principles to the solution of Mates' problem. What we must do is (i) to show that sentence (54) is true, whenever ε and ε′ are synonymous

(54) Mates believes that whoever believes that $\alpha[\varepsilon]$, believes that $\alpha[\varepsilon'//\varepsilon]$,

and (ii) to explain why (54) *seems* to be false. Let us consider separately the cases when ε and ε′ are proper names and when they are predicates, starting from proper names.

Let C be a context in which the common background information i shared by speakers and hearers is that the names "Carl Hempel" and "Peter Hempel" are associated with descriptive information thex(x is a philosopher & x = Carl Hempel) and thex(x is a white-haired gentleman & x = Peter Hempel), respectively, and let us suppose that the reporter Scott Soames assertively utters in C the sentence (55), using it with its literal meaning (nonmetaforically, nonironically, nonsarcastically, and without a defeating conversational implicature):

(55) Mates believes that whoever believes that Carl Hempel died last week, believes that Peter Hempel died last week;

by (P_1), Soames has asserted that Mates believes the Russellian proposition p semantically expressed in C by the clause "whoever believes that x died last week, believes that υ died last week" with respect to the assignement of $[\![\text{Carl Hempel}]\!]_C$ to "x" and of $[\![\text{Peter Hempel}]\!]_C$ to "υ"; since $[\![\text{Carl Hempel}]\!]_C = [\![\text{Peter Hempel}]\!]_C$, p is the same as the Russellian proposition q semantically expressed in C by the clause "whoever believes that x died last week, believes that x died last week" with respect to the assignement of $[\![\text{Carl Hempel}]\!]_C$ to "x", which is believed by Mates by hypothesis; therefore (55) is true.

Why does it seem to be false? Because, in virtue of (P_2), Soames, with his assertive utterance in C of the sentence (55), has asserted *also* that

(56) Mates believes that whoever believes that the philosopher Carl Hempel died last week, believes that the white-haired gentleman Peter Hempel died last week,

and (56) *is* false: Mates would never had taken the subordinate clause of (56) to be true. Mates' error has been to mistake the *semantic content* of (i.e. the proposition *semantically expressed* by).

(57) Whoever believes that Carl Hempel died last week, believes that Peter Hempel died last week

for the proposition *asserted* by means of (57) in the context C, namely the proposition semantically expressed by

(58) Whoever believes that the philosopher Carl Hempel died last week, believes that the white-haired gentleman Peter Hempel died last week.

Let us consider now the case when ε and ε' are predicates. Let C be a context in which the reporter Scott Soames assertively utters the sentence (59), using it with its literal meaning (nonmetaforically, nonironically, nonsarcastically, and without a defeating conversational implicature):

(59) Mates believes that whoever believes that Paul is a doctor, believes that Paul is a physician.

Here (P_1) cannot be applied; however, invoking an analogous principle (P_1')[53] and by means of an argument analogous to the preceding one concerning (55), it is easy to show that (59) is true. But how to explain that it seems to be false, in absence of a principle analogous to (P_2)? In Soames (1987a) a pragmatic principle is stated, which the reporter of a belief attribution should follow:

(P_3) «Remain faithful to the words of the agent unless there is reason to deviate» (Soames, 1987a: 119)[54];

at this point we can presumably reason as follows[55]: (59) violates (P_3), because there is no reason to deviate from the words Mates would have used to express his own belief:

(60) Whoever believes that Paul is a doctor, believes that Paul is a doctor.

However, (P_3) is a pragmatic principle, not a semantic one; as a consequence (59) is pragmatically inappropriate, not semantically false.

It seems to me that Soames's solution is neither optimal nor convincing, for several reasons.

First, Soames's explanation of the reason why (55) seems to be false doesn't work without a further assumption. Notice that, owing to (P_1), Mates' doubt concerns *also* the Russellian proposition semantically expressed by δ'; therefore Soames's reasoning based on (P_2) does not exclude that it concerns the proposition semantically expressed by δ'; and the problem of explaining how this is possible without postulating that Mates is utterly irrational remains open. If, on the contrary, we want

53 Something like:

(P_1') If $\mathcal{R}$ is a competent speaker who assertively utters "$\mathcal{S}$ believes that $\alpha[\pi]$" in a context C, using it with its literal meaning (…), then $\mathcal{R}$ asserts that $[\![\mathcal{S}]\!]_C$ believes $[\![\alpha[x/\pi]]\!]_{C,a:[\![\pi]\!]_C}$, where $[\![\alpha[x/\pi]]\!]_{C,a:[\![\pi]\!]_C}$ is the Russellian proposition semantically expressed in C by the clause $\alpha[x/\pi]$ with respect to the assignement of the denotation of "π" in C to x.

54 Soames adds that

> Someone who assertively utters a sentence standardly asserts the proposition semantically expressed by the sentence in the context. However, the speaker often asserts other propositions as well. (Soames, 1987a: 120).

But he doesn't suggest a criterion to determine which other propositions can be asserted. This gap is filled in Soames (2002) for the case of substitution of names, but not of predicates.

55 The reason of the "presumably" is that I am applying to the present case the explanation proposed by Soames (1987a) of the apparent falsity of other belief reports, and I am not sure that Soames would agree that the two sorts of cases are analogous; as for me, I cannot see any relevant difference.

that Mates' doubt *is not* the doubt that δ', we must prove that it is impossible that it is, and the only argument to this end seems to be Church's argument, which I have already argued to be unconvincing.

Second, there are several cases not covered by the solution.

(a) Consider a context C in which Mates knows that "Max" and "Ben" are names of one and the same person, but also knows that Ann associates absolutely no information to the names "Max" and "Ben", save the 'syntactical' information that Max is a male human being, and that Ben is a male human being; Mates observes Ann asking John which ones of the invitees have arrived for the party, and John answering «Only Max»; then Mates observes Scott asking Ann: «Is Ben arrived?» and Ann answering «No». At this point Mates considers Ann as a good counterexample to the truth of "Whoever believes that Max has arrived, believes that Ben has arrived"; consider an assertive utterance of

(61) Mates believes that whoever believes that Max has arrived, believes that Ben has arrived:

since in C no common background information satisfying (P_2) (i) and (ii) is associated to "Ben" and "Max", (P_1) cannot be used, and Soames's explanation is not available: we have a case of pure conflict of (SP) with (DP_j), and consequently Soames's solution is not optimal.[56]

Cases of this sort are not uncommon; here is another example. The first school day John, a teacher, asks Mary, the secretary, for the list of pupils of his class; Mary gives him two partial, and partially overlapping, lists; he makes a single list out of them, putting on it every name occurring in at least one of the partial lists, and eliminating repetitions; the result is a list of 19 students; «Impossible: Mary says—the students are 18»; she looks at John's list and after a while says:

(62) Here is the explanation: you believe that Carl Smith is different from Peter Smith; in fact, Carl Smith is Peter Smith.

Here, again, no common background information satisfying (P_2) (i) and (ii) is associated to the names "Carl Smith" and "Peter Smith", so (P_2) cannot be invoked, but only (P_1); by (P_1), Mary is saying, in particular, that John believes the Russellian proposition constituted by the individual Carl Smith and the property Difference (whatever it is); and since that proposition is expressed also by the sentence "Peter Smith is different from Peter Smith", according to the theory Mary might have assertively uttered, instead of (62), the following sentence:

[56] A possible reply would be to insist that some background descriptive information associated to the two names does exist after all: "thex (x = Max)" and "thex (x = Ben)", respectively. But the property of being Max is the property of being Ben, so there is no way of differentiating the proposition semantically expressed by the subordinate clause of (61) from a proposition asserted by it in the context. Moreover, I wonder whether this move would not entail a rehabilitation of the descriptive theory of names.

(63) Here is the explanation: you believe that Peter Smith is different from Peter Smith; in fact, Peter Smith is Peter Smith;

but the idea that, with (63), Mary would have given an equally good explanation as with (62) is hard to swallow. In other words: in cases like this, the theory simply makes wrong predictions.

(b) Another sort of case not covered by Soames's solution is the one of Claude: here again (P_2) cannot be used, not because of lack of information associated to the name ν, but simply because no assertion of the form $\alpha[\nu]$ has been made; so we are left without any explanation of 'Claude's error'; on the contrary, we have excellent reasons for the claim that Claude has made no error and that his belief attribution must be taken at face value. The use of principle (*) is a deep-rooted practice of sciences investigating mind and behaviour, and in the use of that principle the assumption that the subject $\mathcal{S}$ does not believe$_j$ *that* $\tau = \tau'$, for instance, or even believes$_j$ *that* $\tau \neq \tau'$, although $\tau = \tau'$, plays a crucial role; it seems therefore necessary that a semantics conceived as a scientific enterprise is capable to account for *what* is believed$_j$ by a subject who believes$_j$ that $\tau \neq \tau'$, when $\tau = \tau'$: to say that $\mathcal{S}$ is irrational because (s)he believes$_j$ a contradiction seems inappropriate, and to say that the whole practice of attributing to $\mathcal{S}$ the non-belief that $\tau = \tau'$ is based on a mistake seems in turn erroneous: a *systematic* error is not an error. If these are the only alternatives (as it seems to be the case), then a constitutive limit of the semantics adopted clearly emerges: its uncapability to provide an adequate description of what is believed by someone who believes a proposition that, according to *that* semantics, is an impossibile proposition; and the reason for such impossibility seems to be the externalist limitation imposed onto such a description, according to which it must make reference only to relations between linguistic expressions and entities of the external world. I shall try to articulate this point in Sect. 7.2.6.

Third, we have seen that the case of substitution of predicates is presumably covered by invoking (P_3); however, the explanation of why (59) seems to be false is not convincing, at least in the version I have suggested (the only one I can see), for at least two reasons.

(a) The explanation assumes that there is no reason to deviate from the words Mates would have used to express his own belief, i.e. (60); but this is false, a reason does exist for Soames, the reporter: Soames knows that "doctor" and "physician" are synonymous, he has observed that Mates assents to "doctors are doctors", and he has inferred (59), by (SP); therefore Soames's assertion of (59) is perfectly compliant with (P_3), and the explanation fails.

(b) Suppose, for the sake of the argument, that there is no reason to deviate from the words Mates would have used to express his own belief, so that (P_3) is in fact violated: it would still be necessary to explain why Soames (both as a reporter and as a propounder of his theory) has violated the principle. Notice that the violation would be systematic, since the theory claims (violating the principle) that "$\mathcal{S}$ believes that $\alpha[\varepsilon'//\varepsilon]$" is truly asserted whenever "$\mathcal{S}$ believes that $\alpha[\varepsilon]$" is truly asserted and ε' is synonymous with ε. If a presumed rule R is systematically violated by the subjects who should follow it, we had better infer that R doesn't exist, or that subjects follow *another* rule; postulating R and adding that it is systematically violated seems an entirely ad hoc move to save the theory.

Fourth, according to Soames the process of interpreting an utterance in a context is in fact more complex than suggested so far. It is important to realize that, according to him, principle (P_2) and a similar one

> should not be considered hard-and-fast criteria for determining what proposition a speaker is using the complement clause of an attitude ascription to pick out. Rather, they are heuristic principles that are often highly useful in interpreting a speaker's remarks, but sometimes play no significant role. [...] the task of determining what someone who has uttered an attitude ascription has attempted to assert and convey is often a matter of interpretation that is both open-ended and highly sensitive to salient information that is part of the set of shared common background assumptions present in the context at the time of the utterance. (Soames, 2002: 227)

If I understand correctly Soames's proposal, in a context in which $\mathcal{R}$ asserts

(64) $\mathcal{S}$ believes that $\pi(\nu)$,

and the name ν is associated to no information, principle (P_1) is applied; principle (P_2) is applied when some background information, shared by speakers and listeners, is associated to ν. The process of interpretation/understanding of (64) in the former cases (let us call them "A-cases: no information is associated to ν) is therefore different from the process of interpretation in the latter (call them "B-cases": some information is associated to ν). In A-cases we associate to $\pi(\nu)$ the Russellian proposition p constituted by the individual denoted by ν and the property denoted by π; let t be the time necessary to associate p to $\pi(\nu)$. In B-cases the process is articulated in two phases: first, our mind generates the propositions $p_1, \ldots, p_m$ that *could* be associated to $\pi(\nu)$ according (P'_2) (m is the number of descriptive pieces of information associated to ν); second, one proposition among $p_1, \ldots, p_m$ is selected, on the basis of some, unspecified, selection criterion.[57] If $t_1, \ldots, t_m$ are the times necessary to associate $p_1, \ldots, p_m$, respectively, to $\pi(\nu)$, and k is the time necessary to effect the selection, the natural empirical prediction we can extract from Soames's theory is that the time necessary, in B-cases, to associate to $\pi(\nu)$ a specific proposition is $t_1 + \cdots + t_m + k$, i.e. (under the simplifying assumption that the time necessary to associate to $\pi(\nu)$ any proposition is the same and equal to t) $(t \times m) + k$: a time much greater than t, if m is sufficiently large. Is this prediction confirmed by empirical data? To my knowledge the question is open.[58]

[57] This is my tentative interpretation of the passage quoted above.

[58] Another difficulty is that, as there is no limit to the amount of descriptive information a name can be associated with, there is no upper bound to the number of believed propositions a belief report attributes to a subject, hence there is no suggestion about when the generation of possibly asserted propositions ends and the process of selection starts.

7.2.5.7 Conclusion

Concluding, no attempt to solve Mates' puzzle based on the reinterpretation strategy answers the crucial question: why can't Mates' doubt be interpreted as concerning the proposition *semantically expressed* by the subordinate clause? The only answer I know is Church's argument, but it is not convincing; the reinterpretation strategy seems therefore to fail. Hence Mates is right when he says that our *intuitive* notion of synonymy forces us to conclude that there are not two distinct synonymous sentences, against (EDS). Of course, in order to avoid this conclusion it is possible to give up (DP) (as Soames, for instance, suggests), but this—as I have argued above—would not be an optimal solution.

The moral to draw from the analysis of the answers proposed to Mates' argument seems to be that the intuitive notion of synonymy has to be abandoned as it does not allow a scientific semantic analysis of belief reports. But this is not the last word: the intuitive notion of synonymy we have made reference till now might not be the only one.

7.2.6 Relativizing Semantic Notions to Subjects

Generally, when we want to develop a scientific investigation of a domain about which we have conflicting intuitions, a good strategy is to revise in some way the intuitive notions, either by modifying them or by observing that, on closer inspection, what seemed to be one notion is in fact a cluster of simpler and clearer notions. The latter seems to be the case with the notion of synonymy. The intuitive notion can be understood in an absolute sense, i.e. in the way normally privileged by philosophers (including Mates), but also in a relative sense, according to which ε and ε' are or are not synonymous *for a subject*.

7.2.6.1 Scheffler

The relevance of the relative notion has been stressed by I. Scheffler in some remarks of fundamental importance about Carnap's notion of intensional isomorphism. In *Meaning and Necessity* Carnap remarks that since

(65) John believes that δ

does not follow from

(66) John is disposed to an affirmative response to some sentence in some language which expresses the proposition that δ,

we must interpret (65) as saying as much as (66) but something more; «and—he remarks—this additional content seems difficult to formulate» (Carnap, 1947: 55).

This seem to me the exact point at which Carnap went astray in the research of an optimal solution. An optimal intuitive solution would have been at hand: in analogy with the fact that modal contexts are (not extensional but) intensional, i.e. satisfy principles of substitutivity of equi-intensional expressions, for instance

(SPEIST) If $\vDash \Box\alpha[\tau]$
and τ has the same intension as τ'
then $\vDash \Box\alpha\left[\tau'//\tau\right]$,

one might have considered a principle of substitutivity of "doxastically equivalent" expressions, for instance

(67) If $\vDash \mathcal{S}$ believes that $\alpha[\tau]$
and $\vDash \mathcal{S}$ believes that $\tau = \tau'$
then $\vDash \mathcal{S}$ believes that $\alpha[\tau'//\tau]$,

and tried to develop a semantics of belief that provably satisfied it.[59] Carnap follows instead a different path: he requires that «The two sentences [$\alpha[\tau]$ and $\alpha[\tau'//\tau]$] must, so to speak, be understood in the same way» (*Ibid.*), and makes precise this intuitive idea by defining the relation of intensional isomorphism. This is precisely the choice Scheffler criticizes:

> "understood" is here crucially ambiguous. Carnap's treatment of "intensional isomorphism" refers throughout to purely linguistic entities and not at all to pragmatics, i.e. to the psychological reactions of persons to sentences. Hence, "being understood in the same way" refers to some semantic characteristic, and no inference may be drawn from the intensional isomorphism of two sentences to the nature of the psychological reactions to them. To exclude the possibility that John may truly be said to believe one and not the other is to express a *psychological* theory as well as a *semantical* one, and a highly improbable one at that. For the same limitations which prevent John from seeing that a sentence is L-equivalent to another may prevent his seeing that one sentence is intensionally isomorphic to another. (Scheffler, 1955: 41)

In other terms, a subject's reaction (of assent or dissent) to a sentence does not depend only on the form of the sentence, but on other factors too, in particular on who the subject is. What Scheffler is stressing here is the *perspectival* or *relative* nature of synonymy, in the sense that two expressions may be synonymous *for* one subject and not *for* another. Later on he remarks:

> Hence, on general grounds, it seems implausible to hold that only a relation as narrow as intensional isomorphism will do for all contexts. Suppose, e.g., we question not John but Russell, using L-true sentences from *Principia*. (Scheffler, 1955: 42)

More explicitly, if we take two L-true (hence L-equivalent) sentences δ and δ' from *Principia*, and question Russell about them, we have a subject who answers «Yes» to δ if and only if he answers «Yes» to δ', even if δ and δ' are not intensionally isomorphic, hence a subject for whom intensional isomorphism is not a necessary condition of synonymy. Again, Scheffler is pointing at an intuitive notion of synonymy that is relative to subjects; equivalently, he is pointing out that synonymy can be conceived

[59] This is the strategy I shall elaborate.

not as a relation between two linguistic expressions, but as a ternary relation between a knowing subject and two linguistic expressions: the relation existing between $\mathcal{S}$, ε and ε′ iff ε is synonymous with ε′ *for* $\mathcal{S}$.

However, Scheffler exclusively imputes the perspectival/relative nature of synonymy to the *limitations* a knowing subject may have, and in this way he relegates it to the pertinence of individual psychology. On the contrary, we have seen at the beginning that the problems raised by belief reports concern in fact justified belief, and that justifications to believe raise problems strictly analogous to the ones raised by justified belief. So it is natural to ascribe the relative character of (this notion of) synonymy not to the psychological peculiarities of a knowing subject, but to objective factors such as the information available to her/him/it and the structure of her/his/its computational apparatus.

7.2.6.2 Relative Synonymy and Internalism

If we make reference to the notion of synonymy relativized to subjects, the condition of adequacy (34) must of course be modified. The natural rephrasal, which I suggest as an adequacy condition for definitions of synonymy-for-$\mathcal{S}$, is the following. Let $\mathcal{NL}$ be a natural language; for every subject name $\mathcal{S}$ of $\mathcal{NL}$, the set $P_{\mathcal{S}}$ of the $\mathcal{S}$-*sentences* is defined as follows[60]:

Definition 1 $P_{\mathcal{S}}$ is the smallest set of sentences of $\mathcal{NL}$ containing

(i) All sentences containing no predicate expressing belief;
(ii) "$\mathcal{S}$ believes that α", if $\alpha \in P_{\mathcal{S}}$;
(iii) the negation, the universal and the existential quantifications of α, if $\alpha \in P_{\mathcal{S}}$ [61];
(iv) the truth-functional combinations of α and β, if $\alpha \in P_{\mathcal{S}}$ and $\beta \in P_{\mathcal{S}}$.

(68) ***Adequacy Condition***

Given a subject $\mathcal{S}$, two expressions are synonymous-for-$\mathcal{S}$ (in $\mathcal{NL}$) if, and only if, they may be interchanged in each $\mathcal{S}$-sentence without altering the truth-value of that sentence.

Let me argue for the plausibility of this condition. If two expressions are synonymous-for-$\mathcal{S}$, then they surely cannot be interchanged in each sentence in $\mathcal{NL}$: think of all the sentences of the form "$\mathcal{S}$' believes that α" and interpreted *de dicto*, where $\mathcal{S}$' is any subject for which the two expressions are not synonymous; on the other hand, they *may* be interchanged in all the sentences in which the 'point of view'

[60] $\mathcal{NL}$ is intended to be the I-language of $\mathcal{S}$.

[61] For example, "Everyone dances" and "Someone dances" are the universal and the existential quantifications of "John dances".

of $\mathcal{S}$ is the only relevant one; and these are precisely the $\mathcal{S}$-sentences. Conversely, nothing short of synonymy-for-$\mathcal{S}$ will guarantee interchangeability of two expressions in all $\mathcal{S}$-sentences: the truth-value of "$\mathcal{S}$ believes that α" depends upon the meaning α has *for* $\mathcal{S}$; if α is replaced by a sentence not having the same meaning for $\mathcal{S}$, the truth-value of "$\mathcal{S}$ believes that α" *may* be changed.

If we make reference to the relative notion of synonymy Mates' puzzle is virtually solved in an optimal way; in particular, it is Mates' second thesis that becomes false, since it becomes possible to argue that, for every subject $\mathcal{S}$, if $\mathcal{S}$ assents to δ, and δ' is synonymous with δ *for* $\mathcal{S}$, then $\mathcal{S}$ assents to δ'. The argument is essentially the following: the validity of (DP_j) is a consequence of a negative answer to (39), i.e. of a principle of epistemic transparency of synonymy; and the counterexamples to epistemic transparency quoted above can be neutralized, i.e. shown to be only apparent counterexamples, if we make reference to the relative notion of synonymy. More precisely, the dissent from δ' in the cases of (35) (b) and of my modification of Pierre's story are explained away as cases in which δ and δ' are not synonymous-*for*-$\mathcal{S}$, while the legitimacy of the substitution by which (35) (b) is inferred from (35) (a) is explained by the fact that δ and δ' are synonymous-*for-the reporter*. Also Rieber's counterexample can now be neutralized: it can be analyzed as a case of divergence between the standpoint of the believer (Joan), who doubts, i.e. does not believe, that the two words are synonymous,[62] and the standpoint of the reporter (ourselves, in this case), who observes that Joan uses the words "bet" and "wager" 'correctly' (i.e. like (s)he would use them herself/himself) and from this infers that Joan is competent about their meaning.

More generally, the distinction between synonymy-for-the-believer and synonymy-for-the-reporter allows to accept both the principle of epistemic transparency of synonymy and (apparent) counterexamples to it: it is perfectly possible that a subject $\mathcal{S}$ is competent for the reporter about ε and ε', that ε and ε' are synonymous for us, and that ε and ε' are not synonymous for $\mathcal{S}$; but if ε and ε' are synonymous for $\mathcal{S}$, then $\mathcal{S}$ is in a position to know that they are.

Owing to the essential identity of Mates' and Frege's puzzles, if we make reference to the relative notion of denotation, i.e. denotation *for a subject*, also Frege's puzzle is virtually solved in an optimal way, since it will be natural to require that, for ν to be interchangeable with ν' in the sentence "$\mathcal{S}$ believes that $\alpha[\nu]$", ν and ν' denote the same object *for* $\mathcal{S}$. In this case, however, the true nature of the semantics I am going to sketch pops up more clearly: it will be an *internalist* semantics, according to which meanings, and denotations, are mental entities, not entities of the external world.

---

[62] Notice that saying that synonymy-for-$\mathcal{S}$ is transparent amounts to saying that if ε andε' are synonymous-for-$\mathcal{S}$ (i.e. in S's own I-language), then $\mathcal{S}$ is in a position to know that they are; the possibility is not excluded that there are pairs $\prec\varepsilon, \varepsilon'\succ$ such that $\mathcal{S}$ cannot answer either "Yes" or "No" the question whether ε andε' are synonymous; this seems to be Joan's epistemic situation in Rieber's example.

7.2.7 An Internalist Semantics for Belief Reports

After having selected the intuitive *relative* notions of synonymy and denotation as the ones permitting an optimal solution of Mates' and Frege's puzzles, hence a scientific (i.e. computational-representatational) approach to the semantics of belief reports, we must now define formal counterparts of them.

My proposal is to extend to $\mathcal{L}_B$ (see the Preface) the internalist semantics for $\mathcal{L}$ developed in Chaps. 4 and 5. To this purpose it is sufficient to specify the denotations of the new kinds of terms and what a justification is for the new kind of formulas.

Belief is a relation between subjects and propositions. The notion of (knowing) subject has been defined in Chap. 4. Propositions, the intended denotations of p-terms, will be equated to $\mathcal{C}$-intensions of sentences.

Definition 2 Let τ be a singular term, π an n-place predicate, α a sentence; then

1. The *$\mathcal{C}$-intension of* τ *for* $\mathcal{S}$ (the *$\mathcal{S}$-intension of* τ, for short), $[\![\tau]\!]_{\mathcal{S}}$, is a function associating to every cognitive state $\sigma \in \mathcal{S}$ the denotation $[\![\tau]\!]_{\mathcal{S},\sigma}$ of τ in σ for $\mathcal{S}$.
2. The *$\mathcal{S}$-intension of* π, $[\![\pi]\!]_{\mathcal{S}}$, is a function associating to every cognitive state σ $\in \mathcal{S}$ the denotation $[\![\pi]\!]_{\mathcal{S},\sigma}$ of π in σ for $\mathcal{S}$.
3. The *$\mathcal{S}$-intension of* α, $[\![\alpha]\!]_{\mathcal{S}}$, is a function associating to every cognitive state $\sigma \in \mathcal{S}$ the value 1 if σ is an $\mathcal{S}$-justification for α, the value 0 if σ is not an $\mathcal{S}$-justification for α.[63]

Definition 3 For every subject $\mathcal{S}$, for every sentence α, *the proposition (semantically) expressed by* α *for* $\mathcal{S}$ is $[\![\alpha]\!]_{\mathcal{S}}$.

On this view, a sentence of $\mathcal{L}$ semantically expresses as many propositions as there are subjects speaking $\mathcal{L}$; in other terms, the relation of semantic expression is relativized to subjects.[64]

Let us now consider the relation of belief. First of all it should be remembered (see Sect. 7.2.2) that we are concerned here not with belief *tout-court* but with *justified* belief, and that, according to my proposal, justified belief should be 'factorized' into a notion of (*ex-ante*) justification and a psychological primitive relation of believing. The intuitive idea therefore emerges that the relation '$\mathcal{S}$ (justifiedly) believes that α' amounts to satisfying three conditions:

(i) $\mathcal{S}$ has a justification for α;

---

63 As $[\![\alpha]\!]_{\mathcal{S}}$ (σ) is either 1 or 0, the $\mathcal{S}$-intension of α may alternatively be identified with $J_{\mathcal{S}}(\alpha) = \{\sigma | \sigma$ is a $\mathcal{S}$-justification for $\alpha\}$.

64 As I said, the background of the internalist semantics sketched above is the Chomskian framework, according to which the only notion of language of whose competence a computational-representatational theory can be given is the notion of I-language, where the "I" points, among other things, to its being the language of an individual subject. As a consequence, different subjects speak, in principle, different languages, possibly with large overlapping areas. However, this does not prevent the possibility for a subject to represent the language of another subject, in particular the meanings assigned by the other subject to the words of his own language.

(ii) $\mathcal{S}$ stands in the (primitive psychological) relation B of belief with the proposition p expressed by α;
(iii) $\mathcal{S}$ stands in the relation B with p because of the justification $\mathcal{S}$ has for α.

Is it possible that these conditions are satisfied by performing a computation? My tentative answer is that it is, provided that two central points are made explicit.

First, we have seen that, in the semantics for $\mathcal{L}$, the $\mathcal{C}$-justifications for an arbitrary α are cognitive states of an arbitrary subject; but when α is a belief report, *two* cognitive states come into play: the one that is a justification for the report itself, and the one that constitutes the cognitive state observed or at least asserted as subsisting; I shall call *reporter* the subject occupying the former state, *believer* the one occupying the latter.[65] A justification for "$\mathcal{S}$ believes that α"[66] will be a cognitive state *of the reporter* in which a certain amount of information is available about the cognitive state *of the believer*, in particular about her/his (justifiedly) believing the proposition expressed by α. A question naturally arises at this point: believing the proposition expressed by α *for which subject*? In principle two possibilities are open: the proposition expressed by α *for the believer*, or the proposition expressed by α *for the reporter*. Several linguistic data suggest that *both* alternatives are available, hence that belief reports of natural languages are ambiguous. I shall come back to this point in a moment.

Second, how to give a computational account of condition (iii), i.e. of the fact that a subject believes a proposition *because* (s)he has a justification for it? It might be tempting to introduce a sort of doxastic maxim:

(DM) Believe all the propositions for which you have justifications,

and to account for condition (iii) by saying that $\mathcal{S}$ follows it. However, this would amount to treating justified belief as a normative relation, whereas I am suggesting that it can be described as a computational relation. I will therefore conceive (iii) in terms of a basing relation between justifications (i.e. cognitive states) and psychological states:

Definition 4 Let B($\mathcal{S}$,p) be the relation of belief between a subject $\mathcal{S}$ and a proposition p, and let C be a condition; then

B($\mathcal{S}$,p) is *based on* C $=_{\text{def}}$ B($\mathcal{S}$,p) only if C.

In order to define justifications for belief reports we must extend the notion of cognitive state; to this end let us introduce the notion of a *mind-reading cognitive state*:

65 "Believer" and "reporter" refer here to two roles, not to two persons; the same person may play both roles, the same role may be played by different persons.

66 "Thatα" is intended to cover the two cases "that$_{\mathcal{Bel}}$α" and "that$_{\mathcal{Rep}}$α", whose meaning will be explained in a moment.

Definition 5 A *mind-reading cognitive state for* $\mathcal{L}_B$ is a cognitive state ρ satisfying the following conditions:

1. ρ is inhabited (also) by subjects;
2. in ρ the relation of belief between subjects and propositions is recognizable and, for every for every proposition p, two roles are assigned to subjects: the role $\mathcal{Bel}_p$of the believer of p, and the role $\mathcal{Rep}_{<\mathcal{S},p>}$ of the reporter of the relation of belief between the subject S and p;
3. in ρ some cognitive states are represented and ascribed to subjects;
4. in ρ the concept f_B is manageable, i.e. a feature-checking algorithm p_B is available, verifying the presence of the following feature configuration:
   there are a proposition p and a cognitive state σ such that

   (i) $\mathcal{Bel}_p$ occupies σ;
   (ii) $j_\sigma(p) = 1$;
   (iii) B($\mathcal{Bel}_p$,p);
   (iv) B($\mathcal{Bel}_p$,p) is based on conditions (i) and (iii).

Some remarks about this definition. In Chap. 4 I have observed that, when an atomic cognitive state has been specified, the answer to the Application Question of a $\mathcal{C}$-concept f_C to an n-tuple of $\mathcal{C}$-objects $o_1 \ldots o_n$ is determined by a computation; here I am assuming that the same holds when the $\mathcal{C}$-concept f_B of belief is applied to a pair of a subject $\mathcal{S}$ and a proposition p. More precisely, I am assuming that, if a mind-reading cognitive state ρ has been specified, $\mathcal{S}$ is a subject and p is a proposition, then the answer (if any) to the following

(69) *Application Question for* f_B(s,p) *in* ρ:

Which is the value of f_B(s,p) in ρ?

is determined by a computation, consisting in checking whether conditions 4.(i)–(v) are satisfied. It is a strong assumption, and a discussion of it is beyond the scope of this book.[67] I simply remark that the computational nature of checking (i) and (ii) is obvious, since it amounts to verifying that certain pieces of information are available to the believer and the reporter; and that verifying (iii) consists in a computation has been argued at length in Chaps. 4 and 5. As for (iv), an essential role is played of course by the principle (DP*), whose reformulation (DP_j), proposed in Sect. 7.2.2, points out that the computation by which (iv) is checked involves inference to the best explanation, which, again, has been described in Chap. 4 as a computational process.[68] Finally, (v) is to be understood essentially as a methodological principle,

[67] For extensive discussions of the assumption see Carruthers and Smith (1996).

[68] In Sect. 7.2.2 I have stressed the role of the principle (*) as well.

in the sense that, once (iv) has been ascertained, abstracting from reasons for it different from (i) and (iii) is a legitimate idealization.

It is time to come back to the first point made above, about the ambiguity of belief reports. The idea is to exploit the possibility, within the internalist semantics developed so far, of defining semantic notions relativized to subjects in order to account for the transparent/opaque ambiguity of belief reports of natural languages.[69] From this point of view the transparent/opaque ambiguity concerns the two propositions ambiguously denoted by the "that" that introduces the subordinate clause α of the belief report: the proposition expressed by α *for the believer*, and the one expressed by α *for the reporter*. In the formal language $\mathcal{L}_B$ the ambiguous denotation of "that α" is disambiguated by means of the two operators $\text{that}_{\mathcal{Bel}}$ and $\text{that}_{\mathcal{Rep}}$; consequently, in the case of a natural language belief report two distinct Justification Questions arise:

(70) 1. *The Justification Question for* $B(\mathcal{S},\text{that}_{\mathcal{Bel}}\alpha)$ *in* ρ:
When is $B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\alpha)$ justified in ρ? In other terms: when is the justificational value $j_\rho(B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\alpha))$ equal to 1?

2. *The Justification Question for* $B(\mathcal{S}, \text{that}_{\mathcal{Rep}}\alpha)$ *in* ρ:
When is $B(\mathcal{S},\text{that}_{\mathcal{Rep}}\alpha)$ justified in ρ? In other terms: when is the justificational value $j_\rho(B(\mathcal{S},\text{that}_{\mathcal{Rep}}\alpha))$ equal to 1?

In order to characterize a computational answer to the Justification Questions we follow the same strategy outlined in Chap. 4 for the other sorts of atomic sentences. First we distinguish between the 'direct' and the 'indirect' cases in which one can be intuitively justified in ρ to believe one's belief report, according as one has or does not have sufficient information to give one answer to the Application Question for $f_B(\mathcal{S},p)$ in ρ; in direct cases the evidential factors of the answer available to the reporter are the same as the evidential factors of the answer to the corresponding Application Question (essentially, the features 6.1–6.3 of Definition 5), in indirect cases inference to the best explanation comes into play. Second, we give the following definitions, analogous to Definitions 6–8 and 11 of Chap. 4:

Definition 6 Let ρ be an arbitrary mind-reading cognitive state for $\mathcal{L}_B$ of a subject $\mathcal{R}$ such that $[\![B]\!]_\rho = f_B$, $\mathcal{S}$ is the believer $\mathcal{Bel}$, $\mathcal{R}$ is the reporter $\mathcal{Rep}$, $[\![\mathcal{S}]\!]_\rho = \mathcal{S}$, $[\![\text{that}_{\mathcal{Bel}}\alpha]\!]_\rho = [\![\alpha]\!]_{\mathcal{S}}$, $[\![\text{that}_{\mathcal{Rep}}\alpha]\!]_\rho = [\![\alpha]\!]_{\mathcal{R}}$; then

- the set $EF_{B(\mathcal{S},\text{that}_{\mathcal{Bel}}\alpha),\rho}$ of the *evidential factors* of $B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\alpha)$ in ρ is $f_B(\mathcal{S}, [\![\alpha]\!]_{\mathcal{S}}) \cup \mathcal{A} = \{<Q, B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\alpha> \mid Q \text{ is a why-question arising in } \rho \text{ and } f_B(\mathcal{S},[\![\alpha]\!]_{\mathcal{S}}) = {}_\rho 1 \text{ is a potential answer to } Q\}$;

[69] By "accounting for the ambiguity" of a sentence α of a natural language $\mathcal{NL}$ I mean defining two possible translations of α into $\mathcal{L}_B$, and defining the notion of justification for each one. Convention: from here on I will omit the subscripts of "$\mathcal{Bel}$" and "$\mathcal{Rep}$" when it is not strictly necessary.

- the set $EF_{B(\mathcal{S},\text{that}_{\mathcal{Rep}}\alpha),\rho}$ of the *evidential factors* of $B(\mathcal{S}, \text{that}_{\mathcal{Rep}}\alpha)$ in ρ is $f_B(\mathcal{S},[\![\alpha]\!]_{\mathcal{R}}) \cup \mathcal{A} = \{<Q,B(\mathcal{S},\text{that}_{\mathcal{Rep}}\alpha> \mid Q$ is a why-question arising in ρ and $f_B(\mathcal{S},[\![\alpha]\!]_{\mathcal{R}}) =_\rho 1$ is a potential answer to Q$\}$.

Definition 7

- if $a \in EF_{B(\mathcal{S},\text{that}_{\mathcal{Bel}}\alpha),\rho}$, then a makes evident $B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\alpha)$ in ρ (in symbols $a \vDash_\rho B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\alpha)$) iff
- either $a = f_B(\mathcal{S},[\![\alpha]\!]_{\mathcal{S}})$ and $f_B(\mathcal{S},[\![\alpha]\!]_{\mathcal{S}}) =_\rho 1$;
- or $a \in \mathcal{A}$ and $f_B(\mathcal{S},[\![\alpha]\!]_{\mathcal{S}}) =_\rho 1$ is the best answer to Q in ρ;
- if $a \in EF_{B(\mathcal{S},\text{that}_{\mathcal{Rep}}\alpha),\rho}$, then a makes evident $B(\mathcal{S}, \text{that}_{\mathcal{Rep}}\alpha)$ in ρ (in symbols $a \vDash_\rho B(\mathcal{S}, \text{that}_{\mathcal{Rep}}\alpha)$) iff
- either $a = f_B(\mathcal{S},[\![\alpha]\!]_{\mathcal{R}})$ and $f_B(\mathcal{S},[\![\alpha]\!]_{\mathcal{R}}) =_\rho 1$;
- or $a \in \mathcal{A}$ and $f_B(\mathcal{S},[\![\alpha]\!]_{\mathcal{R}}) =_\rho 1$ is the best answer to Q in ρ.

Definition 8

- $j_\rho(B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\alpha)) = 1$ iff there is an $a \in EF_{B(\mathcal{S},\text{that}_{\mathcal{Bel}}\alpha),\rho}$ such that $a \vDash_\rho B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\alpha)$;
- $j_\rho(B(\mathcal{S}, \text{that}_{\mathcal{Rep}}\alpha)) = 1$ iff there is an $a \in EF_{B(\mathcal{S},\text{that}_{\mathcal{Rep}}\alpha),\rho}$ such that $a \vDash_\rho B(\mathcal{S}, \text{that}_{\mathcal{Rep}}\alpha)$.

Definition 9

- An *$\mathcal{R}$-justification for the sentence* $B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\alpha)$ is a mind-reading cognitive state ρ of the subject $\mathcal{R}$ such that $j_\rho(B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\alpha)) = 1$;
- An *$\mathcal{R}$-justification for the sentence* $B(\mathcal{S}, \text{that}_{\mathcal{Rep}}\alpha)$ is a mind-reading cognitive state ρ of the subject $\mathcal{R}$ such that $j_\rho(B(\mathcal{S}, \text{that}_{\mathcal{Rep}}\alpha)) = 1$.

If we now consider an English sentence A whose translation into $\mathcal{L}_B$ is α, the English sentence "$\mathcal{S}$ believes that A", can be translated into $\mathcal{L}_B$, for the reasons explained above, as either of the two sentences

(71) (i) $B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\alpha)$,

(b) $B(\mathcal{S}, \text{that}_{\mathcal{Rep}}\alpha)$.

I will call (71)(i) and (ii) the ($\mathcal{Bel}$)-reading and the ($\mathcal{Rep}$)-reading of "S believes that A", respectively. I shall not consider here the question of which factors could determine the choice of the translation, i.e. the disambiguation of the report.

7.2.8 *The Reporter/Believer Ambiguity as an* **Explicans** *of the Transparent/Opaque Ambiguity*

In this section I shall compare the ambiguity just defined (which I shall call "the reporter/believer ambiguity") with Quine's transparent/opaque ambiguity of belief reports. I shall argue that they are two alternative accounts of the same intuitive phenomenon (which, for simplicity, I shall designate with the same name "TO ambiguity").

The intuitive phenomenon is clearly individuated by Quine by making reference to the principle of substitutivity. According to him,

> An opaque construction is one in which you cannot in general supplant a singular term by a *codesignative* term (one referring to the same object) without disturbing the truth value of the containing sentence. In an opaque construction you also cannot in general supplant a general term by a *coextensive* term (one true of the same objects), nor a component sentence by a sentence of the same truth value, without disturbing the truth value of the containing sentence. All three failures are called failures of *extensionality*. (Quine, 1960: 136)

A construction that is not opaque is transparent. Some constructions may be either opaque or transparent, in the sense that they may be interpreted in two alternative ways. One of them is the belief construction:

> Thus suppose that though
>
> [(i)] Tom believes that Cicero denounced Catiline,
>
> he is ill-informed enough to think that the Cicero of the orations and the Tully of *De Senectute* were two. Faced with his unequivocal denial of 'Tully denounced Catiline', we are perhaps prepared both to affirm [(i)] and to deny that Tom believes that Tully denounced Catiline. If so, the position of 'Cicero' in [(i)] is not purely referential. But the position of 'Cicero' in the part 'Cicero denounced Catiline', considered apart, is purely referential. So 'believes that' (so conceived) is opaque.
>
> At the same time there is an alternative way of construing belief that is referentially transparent. [...] The difference is as follows. In the opaque sense of belief considered above, Tom's earnest 'Tully never denounced Catiline' counts as showing that he does not believe that Tully denounced Catiline, even while he believes that Cicero did. In the transparent sense of belief, on the other hand, Tom's earnest 'Cicero denounced Catiline' counts as showing that he does believe that Tully denounced Catiline, despite his own misguided verbal disclaimer. (Quine, 1960: 131)

Here it is clear that the validity of (SP) is the criterion of transparent reading, since it is in virtue of (SP) that we can infer that Tom believes that Tully denounced Catiline from the fact that Tom believes that Cicero denounced Catiline, although this conclusion overtly conflicts with (DP).

I shall localize the origin of the transparent/opaque ambiguity not in the construction "believes that", but in the word "that". In this case it may happen that the that-construction is in turn embedded into a transparent construction; for instance, "It is true that…" is a transparent construction. As Barbara Partee remarks, the existence of such cases «need not contradict the generalization that [that-clauses are] opaque, since the transparency is deducible from the meaning of the word *true*» (Partee, 1974: 838).

If we adopt the validity of (SP) as criterial for the TO distinction, in order to establish that the reporter/believer distinction is an *explicans* of the TO distinction we must verify that (RSP)[70]—the principle playing within internalist semantics the same role played by (SP) in externalist semantics—does not hold for "$B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\,\alpha)$" and does hold for "$B(\mathcal{S}, \text{that}_{\mathcal{Rep}}\alpha)$".

Consider the argument (17), and suppose that ρ is the mind-reading cognitive state of the reporter $\mathcal{R}$, in which the believer is $\mathcal{S}$; then the translations of (17) (a) and (c) in the $\mathcal{Bel}$-readingare "$B(\mathcal{S}, \text{that}_{\mathcal{S}}\ A(l))$" and "$B(\mathcal{S}, \text{that}_{\mathcal{S}}\ A(s))$", respectively.[71] Assume that $\rho \in \mathcal{J}_{\mathcal{R}}\ (B(\mathcal{S}, \text{that}_{\mathcal{S}}A(l)))$[72] and $\rho \in \mathbf{J}_{\mathcal{R}}\ (l = s)$. If $[\![l]\!]_{\mathcal{S}} \neq [\![s]\!]_{\mathcal{S}}$, then $[\![A(l)]\!]_{\mathcal{S}} \neq [\![A(s)]\!]_{\mathcal{S}}$; it is therefore possible that $[\![B]\!]_{\rho}(<[\![\mathcal{S}]\!]_{\rho}, [\![\text{that}_{\mathcal{S}}A(l)]\!]_{\rho}>) = 1$ and $[\![B]\!]_{\rho}(<[\![\mathcal{S}]\!]_{\rho}, [\![\text{that}_{\mathcal{S}}A(s)]\!]_{\rho}>) = 0$; in this case, by Definition 9, $\rho \notin \mathbf{J}_{\mathcal{R}}(B(\mathcal{S}, \text{that}_{\mathcal{S}}\ A(s)))$.

The translations of (17) (a) and (c) in the $\mathcal{Rep}$-reading , are "$B(\mathcal{S}, \text{that}_{\mathcal{R}}A(l))$" and "$B(\mathcal{S}, \text{that}_{\mathcal{R}}A(s))$", respectively. Assume that $\rho \in \mathbf{J}_{\mathcal{R}}(B(\mathcal{S}, \text{that}_{\mathcal{R}}A(l)))$ and $\rho \in \mathbf{J}_{\mathcal{R}}(l = s)$. Since $[\![l]\!]_{\mathcal{R}} = [\![s]\!]_{\mathcal{R}}$, $[\![A(l)]\!]_{\mathcal{R}} = [\![A(s)]\!]_{\mathcal{R}}$; hence, by Definition 9, $\rho \in \mathbf{J}_{\mathcal{R}}(B(\mathcal{S}, \text{that}_{\mathcal{R}}A(s)))$.

The other cases are analogous. In conclusion, the the reporter/believer ambiguity is an *explicans* of the intuitive transparent/opaque ambiguity. That's why, when I will occasionally need to mark explicitly the intended interpretation of the natural language report, I shall use those subscripts: "$\mathcal{S}$ believes $\text{that}_{O}\alpha$" and "$\mathcal{S}$ believes $\text{that}_{T}\alpha$".

On this view, the transparent/opaque ambiguity is not, as Quine (1956) suggests, an ambiguity between two *kinds* of belief (*de dicto* belief and *de re* belief), but an ambiguity concerning the believed proposition, hence, ultimately, between the *standpoints* from which the report is effected: the subject's point of view, in the case of the opaque reading, the reporter's point of view, in the case of the trasparent reading. In other words, there is one kind of belief, i.e. a relation between a subject and a proposition, but there are two points of view from which it may be described, hence two different propositions the subject is described as being in relation to.

It seems to me that Sellars' distinction—mentioned above—between two *uses* of a word in a belief context, our use and the subject's, on the one hand coincides with the TO distinction, on the other hand is close to my proposal to explain it as a distinction between two points of view from which the meaning of "that α" can be seen. The only real obstacle to explaining it that way would be the difficulty, for an externalist semantics, to account for the notion of point of view; whereas we have seen that, within the internalist semantics developed here, there is no such difficulty: points of view simply are subjects.[73]

---

70 See Chap. 5, Sect. 5.4.2.

71 Translations: "A" for "is an actress", "l" for "Sofia Loren", "s" for "Sofia Scicolone".

72 Remember that $\mathbf{J}_{\mathcal{S}}(\alpha) = \{\sigma | \sigma \in \sum \text{ and } [\![\alpha]\!]_{\sigma} = 1\}$ is the set of $\mathcal{S}$-justifications for α (cp. Chap. 5, Remark 1).

73 Sellars' qualification concerning a purely using use, as opposed to a covert mentioning use, is not the description of the phenomenon, but the sketch of an analysis, analogous to Church's analysis;

7.2.9 *Answering Question (6)*

Let us come back to question (6). In consequence of the fact that "$\mathcal{S}$ believes of t that it is P" is understood as equivalent to "$\mathcal{S}$ believes that P(t)" in the transparent reading, the inference involved is no longer an exportation, but an inference from the O-reading to the T-reading:

$$(\text{O/T})\ \frac{\text{s believes that P(t)[O-reading]}\quad \text{E(t)}}{\text{s believes that P(t)[T-reading]}};$$

when is it correct? Keeping present the translations into $\mathcal{L}_{\text{Bel}}$ of its premises and conclusion the answer is obvious:

(72) (O/T) is correct when the proposition expressed by P(t) for the believer is the same as the proposition expressed by P(t) for the reporter.

Some remarks concerning (72). First, notice that the proposition expressed (for either subject) by the clause "P(t)" depends not only on the denotation (for either subject) of the singular term, but also on the denotation of the predicate; on the other hand, (tExp) concerns only the denotation of the singular term; therefore condition (72) is a priori more demanding than any requirement that might be imposed to assure the correctness of (tExp). This seems to comply with Quine's idea that the TO distinction applies to predicates as well.[74]

Second, it is important to observe that, when it is represented by the reporter/believer distinction, the TO distinction is not a binary distinction, since the assignment of roles to subjects depends on the believed proposition, as suggested by condition 2. of Definition 5. (Remember the convention of Footnote 69.) I have not developed a systematic account, but the consideration of an example may provide some hints on this matter. Consider the sentence

(73) Quine believes that Ralph believes that the man in the brown hat is a spy;

it seems intuitively plausible that its possible translations into $\mathcal{L}_{\text{Bel}}$ are the following:

(74) 1. $\text{B}(\mathit{Quine}, \text{that}_{\mathit{Quine}}\text{B}(\mathcal{R}\mathit{alph}, \text{that}_{\mathcal{R}\mathit{alph}}\text{SPY(m)}))$

2. $\text{B}(\mathit{Quine}, \text{that}_{\mathit{Quine}}\text{B}(\mathcal{R}\mathit{alph}, \text{that}_{\mathit{Quine}}\text{SPY(m)}))$
3. $\text{B}(\mathit{Quine}, \text{that}_{\mathcal{R}}\text{B}(\mathcal{R}\mathit{alph}, \text{that}_{\mathcal{R}\mathit{alph}}\text{SPY(m)}))$
4. $\text{B}(\mathit{Quine}, \text{that}_{\mathcal{R}}\text{B}(\mathcal{R}\mathit{alph}, \text{that}_{\mathit{Quine}}\text{SPY(m)}))$

---

and we have seen that Church's argument for such an analysis is unconvincing. Nor is it convincing Sellars' remark that the subject's use is 'not relevant', for reasons explained above.

[74] Quine speaks of general terms; cp. Quine (1960): Sects. 31 and 20.

Here is a tentative way of accounting for these readings. Consider a mind reading cognitive state σ and the following sequence of sentences:

α_1: $\mathcal{S}_1$ believes that α
α_2: $\mathcal{S}_2$ believes that $\mathcal{S}_1$ believes that α
α_3: $\mathcal{S}_3$ believes that $\mathcal{S}_2$ believes that $\mathcal{S}_1$ believes that α
⋮
α_n: $\mathcal{S}_n$ believes that … that $\mathcal{S}_1$ believes that α

(where α is a sentence not containing belief ascriptions).

$\mathcal{Bel}_{\alpha_i}$ ($1 \leq i \leq n$) is the grammatical subject of α_i, $\mathcal{Rep}_{<\mathcal{S},\alpha_i>} = \mathcal{Bel}_{\alpha_{i+1}}$, if (s)he exists, otherwise $\mathcal{Rep}_{<\mathcal{S},\alpha_i>}$ is the subject whom σ belongs to. For each α_i we distinguish a $\mathcal{Bel}$- and a $\mathcal{Rep}$-reading, defined in terms of the possible readings of α_{i-1}; as the possible readings of α_1 are 2, the ones of α_2 are 2^2, the ones of α_3 are 2^3, and so on.

In conclusion I discuss a general objection that might be raised against my proposal to represent the TO distinction by means of the reporter/believer distinction. Consider again sentence (73), and suppose that, unbeknownst to Quine, Ortcutt is also the man met by Ralph in the morning at the local bank. In this situation, the truth of (73) entails the truth of

(75) Quine believes that Ralph believes that the man met at the local bank is a spy,

when (73) is understood in the T-reading, not when it is understood in the O-reading; what does it happen with the reporter/believer distinction? When (73) is understood in the Bel-reading, its truth does not entail the truth of (75), because Quine might not assent to "Ralph believes that the man met at the local bank is a spy"; when (73) is understood in the Rep-reading, its truth does entail the truth of (75), but only under the assumption that there is a reporter of (75) who knows that Ortcutt is the man met by Ralph at the local bank (for example myself). The objection is that the existence of a reporter whom the relevant pieces of information are available to cannot be assumed in general: there can be cases in which a sentence is true (or false) even if there is no possible subject who is in a position to know that it is true, hence to have a justification to believe it. In such cases—the objector might observe—the same sentence is true under the T-reading, but not under the Rep-reading; hence the the reporter/believer ambiguity cannot be the explicans of the transparent/opaque ambiguity, since they are not even extensionally equivalent.

My answer is that the objection is based on the realist assumption that, if no possible subject were in a position to assert (75), i.e. if (75) were not assertible, it would however be true. The anti-realist's rejoinder might be that "true" just means assertible, hence (75) would *not* be true if there were not a (possible) subject in a position to assert it (i.e., a cognitive state in which it is known). If the realist objected that "true" does not mean assertible, because there are true but non-assertible sentences, the antirealist would ask to exhibit one, and the realist would not be capable to (on pain of contradiction: "A is non-assertible" is intuitionistically equivalent to "¬A is assertible", against the hypothesis that A is true); (s)he could at most show that it is impossible that there is no true non-assertible sentence, but this would not be enough for the anti-realist. The moral to draw seems to be that the objection is not conclusive, in the sense that it is not convincing for someone who does not share realist convictions with the objector. As a matter of fact, the semantics adopted here is based on a theory of meaning of an intuitionistic insipiration, not because such a theory is considered in some sense more 'correct', but because it seems necessary to illuminate interesting aspects of the semantics of natural languages such as the transparent/opaque ambiguity and, as we shall see in a moment, the ESN ambiguity as well.

7.2.10 Substitutivity, Disquotation and Translation Revisited

The fact that an optimal solution of problem (38) does not exist does not mean that problem (38) cannot be reformulated in such a way that its modified version admits an optimal solution. The internalist semantics illustrated above clearly suggests the following reformulation of (38):

(38′) **The problem**

Give an adequate definition of the meaning of "$\mathcal{S}$ believes that α" and of a theoretical notion of synonymy-for-$\mathcal{S}$, i.e. a definition satisfying (RSP), (DP_j), (*) and (EDS).

In this section we will see that an optimal solution does exist.

We have seen in Chap. 5, Sect. 5.4.2, that (RSP) is valid for all sentences of $\mathcal{L}$, and in Sect. 7.2.8 of this chapter that it is valid also for the sentences of $\mathcal{L}_B$ in which only "$that_{\mathcal{Rep}}$" occurs, while it is not valid for the sentences of $\mathcal{L}_B$ in which "$that_{\mathcal{Bel}}$" occurs. A good explanation of why it is not valid is now available: in those sentences the relevant standpoint is the point of view of the believer, and for the believer the subject substituendum and substituent have *not* the same denotation (in the case of Frege's puzzle), or are *not* synonymous (in the case of Mates' puzzle).

More importantly, we can prove that, for those sentences, a duly restricted principle of substitutivity does hold. Let us define, for every subject $\mathcal{S}$, the set $P_{\mathcal{S}}$ of *S-sentences of* $\mathcal{L}_B$[75]:

Definition 10 $P_{\mathcal{S}}$ is the smallest set of sentences of $\mathcal{L}_B$ such that

(i) if $\alpha \in \mathcal{L}$, then $\alpha \in P_{\mathcal{S}}$;
(ii) if $\alpha \in P_{\mathcal{S}}$, then $B(\mathcal{S}, that_{\mathcal{Bel}}\alpha) \in P_{\mathcal{S}}$;
(iii) if $\alpha \in P_{\mathcal{S}}$, then $\neg\alpha, \forall x\alpha, \exists x\alpha \in P_{\mathcal{S}}$;
(iv) if $\alpha \in P_{\mathcal{S}}$ and $\beta \in P_{\mathcal{S}}$, then $\alpha\wedge\beta, \alpha\vee\beta, \alpha \rightarrow \beta \in P_{\mathcal{S}}$.

[75] For evident reasons I give to this set of sentences of $\mathcal{L}_{Bel}$ the same name given to the set of sentences of English defined by Definition 1.

If we now consider the notion I have proposed as the *explicans* of the intuitive relation of synonymy between sentences, i.e. identity of $\mathcal{S}$-intension, we can verify, by induction on α, the validity of the following Restricted Relativized Substitutivity Principle:

(RRSP) For every ρ, for every $\mathcal{S}$, for every $\alpha \in P_{\mathcal{S}}$, if $\rho \in \mathbf{J}_{\mathcal{R}}(\alpha[\varepsilon])$ and $\rho \in \mathbf{J}_{\mathcal{R}}(B(\mathcal{S}, \text{that}_{\mathcal{S}}\varepsilon = \varepsilon'))$, then $\rho \in \mathbf{J}_{\mathcal{R}}(\alpha[\varepsilon'//\varepsilon])$.

Consider the base clause: α is of the form $B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\beta)$, where $\beta \in L$. Suppose that $\rho \in \mathbf{J}_{\mathcal{R}}(B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\beta[\varepsilon]))$ and that $\rho \in \mathbf{J}_{\mathcal{R}}(B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\varepsilon = \varepsilon'))$, i.e. that $\llbracket B\rrbracket_{\rho}(<\llbracket\mathcal{S}\rrbracket_{\rho}, \llbracket\text{that}_{\mathcal{Bel}}\beta[\varepsilon]\rrbracket\rho>) = 1$ and that $\llbracket B\rrbracket_{\rho}(<\llbracket\mathcal{S}\rrbracket_{\rho}, \llbracket\text{that}_{\mathcal{S}}\varepsilon = \varepsilon'\rrbracket_{\rho}>) = 1$; then from the fact that $\llbracket\text{that}_{\mathcal{S}}\beta[\varepsilon]\rrbracket_{\rho} = \llbracket\beta[\varepsilon]\rrbracket_{\mathcal{S}}$, that $\llbracket\text{that}_{\mathcal{Bel}}\varepsilon = \varepsilon'\rrbracket_{\rho} = \llbracket\varepsilon = \varepsilon'\rrbracket_{\mathcal{S}}$, and that $\llbracket\varepsilon\rrbracket_{\mathcal{S}} = \llbracket\varepsilon'\rrbracket_{\mathcal{S}}$, it follows that $\llbracket B\rrbracket_{\rho}(<\llbracket\mathcal{S}\rrbracket_{\rho}, \llbracket\text{that}_{\mathcal{Bel}}\beta[\varepsilon'//\varepsilon]\rrbracket_{\rho}>) = 1$, hence that $\rho \in \mathbf{J}_{\mathcal{R}}(B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\beta[\varepsilon'//\varepsilon]))$. The other clauses are proved analogously. The validity of (RRSP) shows the material adequacy of Definition 16 of Chap. 4.

The principles of disquotation and (*) need to be reformulated as well, because of the ambiguity we now recognize in "that α". (DP_j) makes reference to the proposition expressed by "α", but this is no longer meaningful, since the relation of semantical expression is now relative to subjects; since the subject expressing assent or dissent is the believer, the obvious reformulation is the following:

(DP'_j) Assuming that $\mathcal{S}$ is a sincere, reflective, rational, attentive subject, who understands "α", if $\mathcal{S}$ assents to "α" then a reporter $\mathcal{R}$ is justified in asserting that $\mathcal{S}$ believes $\text{that}_O\ \alpha$; and if $\mathcal{S}$ dissents from "α", then $\mathcal{R}$ is justified in asserting that $\mathcal{S}$ does not believe $\text{that}_O\ \alpha$.[76]

Analogously, the obvious reformulation of (*) is the following:

(*′) If $\tau = \tau'$ and $\mathcal{S}$ does not believe $\text{that}_O\ \tau = \tau'$, then "$\mathcal{S}$ believes $\text{that}_O\ \alpha[\tau]$" does not entail "$\mathcal{S}$ believes $\text{that}_O\ \alpha[\tau'//\tau]$".[77]

Let us see how the problems introduced at the beginning are solved within the internalist semantics I have adopted.

In the case of arguments (1)–(5), which are instances of either **Frege's puzzle** or **Mates' puzzle**, the solution is uniform. The (a)- and the (c)-sentences of arguments (1)–(5) are ambiguous with respect to the transparent/opaque ambiguity, in the sense that their translations into $\mathcal{L}_B$ are of the forms "$B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\ \alpha)$" (corresponding to the O-reading) and "$B(\mathcal{S}, \text{that}_{\mathcal{Rep}}\alpha)$" (corresponding to the T-reading), respectively. The single substitutivity principle (SP) of externalist semantics has been replaced, within the internalist semantics, with the two (schematic) principles (RRSP) and (RSP), which are valid for sentences occurring in the contexts $B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\ldots\ldots)$ and $B(\mathcal{S}, \text{that}_{\mathcal{Rep}}\ldots\ldots)$, respectively. Also ($DP'_j$) is valid: when the belief report is interpreted according to the Bel-reading, no counterexample to it can arise because

[76] Cp. Kripke's remark that the principle (DP) is satisfied only when the belief report is interpreted opaquely (Kripke, 1979: 276, fn. 22).

[77] If $\tau = \tau'$, then "$\mathcal{S}$ believes $\text{that}_T\alpha[\tau]$" does entail "$\mathcal{S}$ believes $\text{that}_T\alpha[\tau'//\tau]$", since substitutivity holds for the T reading.

of an application of (RRSP), because substituendum and substituent are synonymous [or equi-extensional, in the case of (18)] for the believer. The validity of (*′) is obvious.

As for (EDS), the possibility of syntactically different synonymous expressions is no longer in question, once synonymy is relativized to subjects: since $\mathcal{S}$-synonymous expressions can be be substituted to each other in every sentence belonging to $P_{\mathcal{S}}$, Mates' argument is blocked.

Summing up, the solution is optimal, according to the definition of optimality adopted above. The conflict between substitutivity and disquotation is, from the standpoint of the present approach, an appearence due to the lack of a distinction between synonymy-for-the-believer and synonymy-for-the-reporter.

To conclude, let us consider **Kripke's puzzle**. We have seen that this puzzle is taken as the central objection to the principle (DP) by Soames, who considers it as «an apparent reductio ad absurdum» of the principle. But Kripke's reductio assumes (TP), so it proves not that (DP) is false, but that *either (DP) or (TP)* is. So let us consider (TP) more carefully.

A first reason of perplexity about it is strictly related to Burge's remark quoted in Sect. 7.2.5.4—«Obvious truths and falsehoods are normally used in arriving at a translation—not vice versa»—: the principle presupposes that the relation 'β is a translation of α' is well-defined, whereas we have seen in Sect. 7.2.5.4. Some reasons to the contrary. But let us skip this point and assume, for the sake of the argument, that translation is a well-defined relation.

A second reason has to do with ambiguity. Consider the ambiguous Italian sentence α: "Ogni marinaio ama una ragazza carina" and its English translation α′: "Every sailor loves a nice girl": does the translation observe (TP)? Yes and No: Yes if both α and α′ receive the same interpretation (either ∀∃ or ∃∀), No if α receives the ∀∃-interpretation, α′ the ∃∀-interpretation and each sailor loves a different nice girl. If we want to take into account the phenomenon of ambiguity, as it seems necessary, and to avoid such contrasting answers, the following modification of (TP) is necessary:

(TP′) If a sentence of one language expresses a truth in that language under the interpretation I, then any translation of it into any other language also expresses a truth (in that other language) under the same interpretation.

Finally, although translation is not in itself a semantical relation, but a relation between languages, it is strictly determined by semantical relations, as the very principle (TP) shows, with its reference to the notion of truth. Since I have argued that, in order to give an adequate treatment of belief reports, reference should be made to semantical relations relativized to subjects, the same arguments apply to translation, and (TP′) has to be reformulated in terms of an intuitive notion of translation *for a subject*; as a concequence, also reference to truth must be replaced by reference to knowledge. The outcome is the following:

(TP″) For every two natural languages $\mathcal{N}$ and $\mathcal{N}'$, for every subject $\mathcal{S}$, for every sentence $\alpha \in \mathcal{N}$ and $\alpha' \in \mathcal{N}'$, if $\mathcal{S}$ knows that α, under the interpretation I, and α' is a translation for $\mathcal{S}$ of α into $\mathcal{N}'$, then $\mathcal{S}$ knows that α', under the same interpretation.

Let me remark, by the way, that this principle seems to fare better than (TP) from the heuristical point of view. Consider for example (44): if we use (TP), its translations must preserve its truth-value; but the question is: which is its truth-value? According to Mates and Putnam, among others, it is false, according to Church and others it is true: there is no unique answer, so (TP) is of no use. If we use (TP”), its translations must preserve its ‘$\mathcal{S}$-knowledge-value’, for every subject $\mathcal{S}$; in this case, given a subject $\mathcal{S}$, the question is: which is knowledge-value for $\mathcal{S}$? And the answer to *this* question is definite.

Consider now Kripke’s reductio: the truth of (24) is inferred from the truth of (21) by means of (TP). But now (TP) is replaced by (TP″), and the relevant interpretation under which (21) is known by the reporter $\mathcal{R}$ is the one semantically expressed by

(21′) Pierre croit que$_{\mathrm{O}}$ Londres est jolie,

because (21) is inferred by (DP), which, as we have seen, is valid only when the belief report is understood opaquely. As a consequence, in order to apply legitimately (TP″) it should hold that the sentence

(24′) Pierre believes that$_{\mathrm{O}}$ London is pretty

is a translation of (21′) into English for the reporter $\mathcal{R}$; but (24′) is evidently *not* such a translation for $\mathcal{R}$: if someone asked Pierre: “Is London pretty?”, he would obviously answer “No”, and it is perfectly possible that the reporter knows that fact, and consequently does not believe that (24′) translates (21′). The reductio is therefore blocked.

What the reporter does know is that

(24″) Pierre believes that$_{\mathrm{T}}$ London is pretty;

but this is not relevant: the translation of (24″) into $\mathcal{L}_{\mathrm{Bel}}$ is

(24*) B($\mathcal{Pierre}$, that$_{\mathcal{Rep}}$ London is pretty),

and this is not the contradictory of

(23*) $\neg$B($\mathcal{Pierre}$, that$_{\mathcal{Bel}}$London is pretty),

which is the translation of (23) under the O-interpretation. From this point of view Kripke’s puzzle is based on an equivocation fallacy.

7.2.11 Conclusion

The main concern of this section has been the possibility of developing a scientific semantics for propositional attitudes, namely a semantics compatible with a view of knowledge of meaning as a system of computational structures and processes. From this point of view the importance of the disquotational principle and of other principles normally adopted in scientific practice has been stressed. The main obstacle to such a possibility has turned out to be the presupposition, characteristic of externalist semantics, that synonymy is an absolute relation between linguistic expressions; on the contrary, focusing on the relative notion of synonymy-for-a-subject turned out to be the keystone to an optimal solution of the foundational puzzles of Mates and Kripke. The attempt to develop a theory of the relative notion and of the notion of knowing subject has entailed the necessity of articulating an internalist semantics.

The relative notion of synonymy satisfies a principle of epistemic transparency, i.e. a negative answer to (39); more generally, internalist meaning is epistemically transparent, as we have seen in Chap. 4.

7.3 The Paradox of Analysis

The Paradox of Analysis has been stated by C. H. Langford in the following way:

> Let us call what is to be analyzed the analysandum, and let us call that which does the analyzing the analysans. The analysis then states an appropriate relation of equivalence between the analysandum and the analysans. And the paradox of analysis is to the effect that, if the verbal expression representing the analysandum has the same meaning as the verbal expression representing the analysans, the analysis states a bare identity and is trivial; but if the two expressions do not have the same meaning, the analysis is incorrect. (Langford, 1942: 323)

Consider for instance the two sentences

(76) The concept BROTHER is identical with the concept BROTHER
(77) The concept BROTHER is identical with the concept MALE SIBLING;

Both are true identities, but only the latter is informative. How to account for this?

The paradox is strictly related to the ones I have introduced at the beginning, because of its connection with the relation of synonymy (analysandum and analysans must be in some sense synonymous), but it is different as far as it concerns the nature of linguistic competence, even of a subject's own I-language: the intuitions at the basis of our judgements of adequacy of an analysis stem from our linguistic competence, in particular from our linguistic competence about predicates.[78]

Carnap's answer (Carnap, 1947: 64) consists in distinguishing, within the notion of synonymy, two notions: synonymy proper, whose explicans is intensional isomorphism, and cognitive equivalence, whose explicans is identity of intension; at this

[78] Carnap, for instance, remarks, about a sentence similar to (77) in relevant respects, that it

> is a sentence conveying fruitful information, although of a logical, not a factual, nature; it states the result of an analysis of the analysandum, the concept Brother. (Carnap, 1947: 63)

It seems to me that "logical" is best understood here as meaning information stemming from linguistic competence.

point he stipulates that an analysis is correct if analysandum and analysans have the same intension—which happens both in (76) and in (77); and that an analysis is informative if analysandum and analysans are not intensionally isomorphic—which happens only in (77). The difficulty with this answer is that it does not account for the difference in informative value between the analyses

(78) The concept PHYSICIAN is the same as the concept PHYSICIAN
(79) The concept PHYSICIAN is the same as the concept DOCTOR,

since analysandum and analysans are intensionally isomorphic also in (79), which is clearly informative.[79]

Carnap would presumably suggest the following answer: (79) is *not* informative, since it speaks of *one and the same* concept, and not of two concepts, although it refers to it twice with the help of two different signs; (79) says of this one concept that it is identical with itself, and it is therefore "rather trivial". He would also suggest why (79) *seems* informative: he would distinguish between (79) and

(80) "The concept PHYSICIAN" has the same meaning as "the concept DOCTOR",

and would impute to his objector a misinterpretation of (79)—which cannot be false if (78) is true—as meaning the same as (80)—which *can* be false even if (78) is true.[80]

If my reconstruction of Carnap's answer is correct, he anticipates here the strategy of reconstruction illustrated above (Sect. 7.2.5.4) in connection with Mates' puzzle, as it is implemented in particular by Church. And my first objection would be similar: Carnap is implicitly assuming that, if the objector deems (79) as informative, it is impossible for him to take it at face value; a convincing and non *ad-hoc* argument to this effect would be necessary, but none is given.[81] Second, the reason Carnap adduces for the non-informativeness of (79) applies to (77) as well: it speaks of *one and the same* concept, and the fact that it refers to it twice with the help of two different signs is irrelevant; so why is (77) deemed as informative and (79) as trivial? The answer that "Brother" "Male Sibling" are not intensionally isomorphic, while "Physician" and "Doctor" are, would be unacceptable: we need a reason independent of the predictions of the theory.

Church has proposed a solution to the paradox based on the observation that

> The paradox of analysis has an obvious analogy with Frege's puzzle [...], as to how an equation, say '*a*=*b*', can ever be informative – because, it seems, if the equation is true then '*b*' is replaceable by '*a*', and hence '*a*=*b*' is the same in meaning as '*a*=*a*'. (Church, 1946, 133.)

[79] The difficulty has been stated in Linsky (1949), where Linsky credits it to Benson Mates. However, Linsky's counterexample contained not names of concepts, but names of objects.

[80] I am extrapolating from Carnap's remarks about "the number 5" and "the number V" in his reply to Linsky (Carnap, 1949: 347–348).

[81] I have explained above (Sect. 7.2.5.4) why I find unconvincing Church's argument for a similar conclusion.

The solution takes advantage of Frege's distinction between sense and denotation, and postulates that the context "The concept…" is to be conceived as an oblique context; as a consequence "DOCTOR" and "PHYSICIAN", in (79), denote the ordinary Fregean senses of the two predicates (Church calls them "concepts"). As (79) is true, the ordinary sense of the two predicates is the same; but nothing prevents that their *indirect* senses be different, and if we postulate that they are we get an explanation of the informativeness of (79).

It is interesting to observe that this postulate is inconsistent with the answer Church proposes to Mates' Problem. As we have seen when we have analysed Church's argument for his claim (iii) (in Sect. 7.2.5.4), the premises of Church's implicit reasoning seem to be: (i) that translation should be compositional; (ii) that distinct synonymous expressions must be translated into distinct synonymous expressions, if possible. Notice that two synonymous expressions must be conceived as having the same direct *and n-indirect* (for every n) Fregean sense, otherwise there would be some n such that they are not interchangeable in some n-iteration of the context "$\mathcal{S}$ believes that". It is therefore at this point that Church is making the implicit assumption that "fortnight" and "period of fourteen days" have the same indirect sense, in conflict with the postulate that "Doctor" and "Physician" have different 1-indirect senses. Of course this inconsistency is not an argument against the postulate, but only against *either* the postulate *or* Church's solution to Mates' Problem; the discussion above suggests that the best choice is to give up Church's solution to Mates' Puzzle.[82]

Another interesting solution has been proposed by M. Richard (Richard, 2001). According to him the paradox consists in the fact that the following principles are incompatible with the intuitively true claim that there are correct not trivial analyses:

(i) *Identity*:

An analysis is what is said by a sentence of the form "The concept C is the concept C′", where "is" names the relation of identity.

(b) *Synonymy*:

When "The concept C is the concept C′" expresses a correct analysis, "C" and "C′" are synonymous.

(iii) *Compositionality*:

If "C" and "C′" are synonymous, "The concept C is the concept C′" and "The concept C is the concept C" say the same thing.

(d) *Triviality*:

What is said by "The concept C is the concept C" is trivial.

According to Richard in (i)–(iv), and consequently in both (78)–(79) and (76)–(77), two senses of "synonymous" are confused, which are both intuitively legitimate but should be kept clearly distinct. Let us say that the *proper contribution* of

[82] For a different opinion see Salmon (1993).

an expression (to what is said) is its semantic value (a property for a common noun, its bearer for a proper name, etc.); an expression is *responsible* for the proper contributions of all of its subexpressions (for example, "that all men are created equal" is responsible for the property of equality, which is the proper contribution of its subexpression "equal"); two expressions are *phrasally synonymous* if their proper contribution is the same; they are *structurally synonymous* if they are responsible for the same proper contributions. If "$\text{synonymous}_{\text{phr}}$" means phrasally synonymous and "$\text{synonymous}_{\text{str}}$" means structurally synonymous, the correct reformulation of (ii)–(iii) is the following:

(ii′) Synonymy:

When "The concept C is the concept C′" expresses a correct analysis, "C" and "C′" are $\text{synonymous}_{\text{phr}}$.

(iii′) Compositionality:

If "C" and "C′" are $\text{synonymous}_{\text{str}}$, "The concept C is the concept C′" and "The concept C is the concept C" say the same thing.

In the case of (76)–(77) it is clear that "The concept BROTHER" and "The concept MALE SIBLING" are $\text{synonymous}_{\text{phr}}$ but *not* $\text{synonymous}_{\text{str}}$; so the analysis is both correct and not trivial.

Richard's solution seems to me unconvincing for three reasons. First, the distinction between the two kinds of synonymy is introduced as an intuitive distinction; on the other hand, it is meaningful only for someone who adopts a 'Russellian' view of propositions as linguistically structured entities—a highly theoretical view; whoever does not adopt it can legitimately blame the distinction for being ad hoc. Second, if we assume that Peter Hempel is Carl Hempel and that names having the same denotation are synonymous, we have that

(81) The doctrine of Peter Hempel
(82) The doctrine of Carl Hempel

are $\text{synonymous}_{\text{str}}$; but this clashes with intuitive data: the analysis

(83) The doctrine of Peter Hempel is the doctrine of Carl Hempel

is intuitively both correct and not trivial—in particular it is useful for whoever knows the philosopher Carl Hempel but does not know that Peter Hempel is the same person. Third, in the case of (78)–(79), if we want that "The concept PHYSICIAN" is *not* $\text{synonymous}_{\text{str}}$ of "The concept DOCTOR" (as Richard's solution seems to require), we must postulate that the property of being a doctor is different from the property of being a physician; but then how can the analysis be correct?

The solution I suggest to the Paradox is the following. Intuitively, an identity "a = b" is true if the denotation of "a" is the same as the denotation of "b"; it is trivial if the identity of denotation of "a" and "b" is known a priori, i.e. before having empirically established that a is the same as b. Within the internalist semantics proposed here, "a priori" can be reinterpreted as meaning *before having computed the denotation* of "a" and “b”. So, given a subject $\mathcal{S}$ and a cognitive state σ of $\mathcal{S}$, let us define an identity π = π’ as $\mathcal{S}$-*correct* iff $[\![\pi]\!]_{\mathcal{S}} = [\![\pi']\!]_{\mathcal{S}}$,[83] as $\mathcal{S}$-*trivial* iff, in order to establish whether it is $\mathcal{S}$-correct, it is not necessary to compute neither $[\![\pi]\!]_{\mathcal{S}}$ nor $[\![\pi']\!]_{\mathcal{S}}$. As a matter of fact there are at least three kinds of trivial identities between predicates:

(a) when π and π′ are tokens of the same syntactical type;
(b) when the piece of information that $\pi = \pi'$ is contained in the lexical content either of π or of π′;
(c) when $lec_{\pi} =_{\text{subst}} lec_{\pi'}$.[84]

An example of the kind (a) is the identity “BACHELOR ≡ BACHELOR”: in order to establish whether it is $\mathcal{S}$-correct it is sufficient to observe that “BACHELOR” and “BACHELOR” are tokens of the same syntactical type. An example of the kind (b) is the identity “BACHELOR ≡ MALE UNMARRIED MAN”, if lec_{BACHELOR} contains the piece of information that a bachelor is a male unmarried man: in order to establish whether it is $\mathcal{S}$-correct it is sufficient to observe that the piece of information that a bachelor is a male unmarried man is contained in lec_{BACHELOR}. An example of the kind (c) is the identity “DOCTOR ≡ PHYSICIAN”, if lec_{DOCTOR} consists in the piece of information that a doctor is a person skilled in the art of healing and $lec_{\text{PHYSICIAN}}$ consists in the piece of information that a physician is a person skilled in the art of healing: in order to establish whether it is $\mathcal{S}$-correct it is sufficient to observe that $lec_{\text{DOCTOR}} = lec_{\text{PHYSICIAN}}$. An example of non-trivial identity between predicates is the identity “DOCTOR ≡ PHYSICIAN”, if lec_{DOCTOR} consists in the piece of information that a doctor is a person engaged in the practice of medicine and $lec_{\text{PHYSICIAN}}$ consists in the piece of information that a physician is a person skilled in the art of healing: in order to establish whether it is $\mathcal{S}$-correct it is necessary to compute the denotations of the two predicates, i.e. to establish that “person skilled in the art of healing” and “person engaged in the practice of medicine” define the same discriminating function, the same $\mathcal{C}$-concept.

The case of “GREEK” and “HELLENE” (and of all the predicates whose meaning is defined by means of a name) is different: the subject is not in a position to compute whether they are synonymous because (s)he lacks an essential piece of information (whether Greece = Hellas), i.e. something distinguishing a cognitive state from another; in the case of “DOCTOR” and “PHYSICIAN” no datum is lacking, but the execution of a computation the subject is in a position to carry out.

[83] Remember that $[\![\pi]\!]_{\mathcal{S}} = [\![\pi']\!]_{\mathcal{S}}$ iff $[\![\pi]\!]_{\mathcal{S},\sigma} = [\![\pi']\!]_{\mathcal{S},\sigma}$, for some σ of $\mathcal{S}$.

[84] As we have seen in footnote 44 and 45 of Chap. 4, the atomic sentences expressing verbal information associated to a designator δ must contain occurrences of δ; as a consequence it is impossible that lec_{δ} and $lec_{\delta'}$ are identical if δ and δ′ are different expressions. However, it is possible that they are equal modulo the susbstitution, within each piece of information of $lec\delta$, of δ with δ′, or viceversa; this is just the meaning of “$=_{\text{subst}}$”.

7.4 The Epistemic Specific/Non-specific Distinction

Let us now consider question (8). As a preliminary, in Sect. 7.4.1 I shall argue that the ESN distinction cannot be represented in terms of scope; in Sect. 7.4.2 I introduce a distinction between to kinds of cognitive states which can serve as justifications for sentences of the form $\exists x\alpha$; in Sect. 7.4.3 the distinction is connected to the one between the assertibility conditions, within intuitionistic logic, of $\exists x\alpha$ and $\neg\forall x\neg\alpha$; in Sect. 7.4.4 and in Sect. 7.4.5 the distinction is used to account for the epistemic specific/non-specific ambiguity.

7.4.1 The ESN distinction Cannot Be Represented in Terms of Scope

Since the ESN distinction is meaningful, as we have seen, also in relation to sentences containing no opaque verbs, it is evident that, at least in these cases, it cannot be represented in terms of scope of the existential quantifier. However, it seems to me arguable that even when we have to do with sentences containing opaque verbs, the distinction cannot be represented in terms of scope.

To start, consider sentence (9). It will be conceded that its natural translation into a first order language is of the form $\exists x\alpha$, and that the truth-conditions of a sentence of that form are such that it is not legitimate to say that it concerns a specific individual; for this respect an existential sentence is not different from a universal one: both express *general* thoughts, concerning the set of entities satisfying the predicate. Therefore, when we translate a sentence of a natural language like (9) into a logical sentence of the form $\exists x\alpha$, we are necessarily representing the *non-specific* reading of the indefinite.

Consider now a sentence structurally more complex, for example

(84) Every boy was kissed by a girl;

it is usually—and correctly—held to be ambiguous between a specific and a non-specific reading, and according to many such ambiguity can be conceived as a scope phenomenon, in the sense that the two readings can be represented by translating the sentence into the following two sentences of a first order formal language, differing from one another for the relative scope of the two quantifiers:

(84′) $\exists x\forall y((\text{GIRL}(x)\wedge\text{BOY}(y))\supset\text{KISSED}(x,y))$
(84″) $\forall y\exists x((\text{GIRL}(x)\wedge\text{BOY}(y))\supset\text{KISSED}(x,y))$.

But why should (84′) represent the *specific* reading of (84) (hence of the indefinite in (84)), if (84′) too is of the form $\exists x\alpha$, and consequently can represent only a non-specific reading? Maybe a reason for this confusion can be found in the following fact. Let us compare (84′) with (84″), usually held to represent the non-specific interpretation of (84). For (84″) to be true the value assigned to x may change according to the value assigned to y; on the contrary, for (84′) to be true the value assigned to x must keep *constant* as the values assigned to y change. It is just this constancy of the value of x that may be confused with (hence held to be capable of representing) its specificity.[85] However, constancy is *not* epistemic specificity. Suppose that John has observed lipstick traces of the same colour on the cheeks of each boy, and that he asserts (84) on the basis of this evidence: the content of his assertion is correctly rendered by (84′), but John has no referent in mind, nor could he specify any girl.[86]

Exactly for the same reasons the sentence (3) cannot represent the specific interpretation of (1): like (84′), (3) is of the form $\exists x\alpha$. What (2) and (3) say can be represented, respectively, in the following way:

(2′) Ralph believes that $\{x | x \text{ is a spy}\} \neq \emptyset$
(3′) $\{x | \text{Ralph believes of } x \text{ that (s)he is a spy}\} \neq \emptyset$

Although this is counterintuitive, (2′) is true if, and only if, (3′) is true (under the assumption that some spy exists). For, if (3′) holds, then Ralph will assent to "Someone is a spy", so (2′) is true; conversely, supposing that (2′) is true is equivalent to supposing that (2) is true, and we saw above that (2) and the fact expressed by (11) entail that (12) is true; on the other hand, the shortest spy exists and is unique; hence it is (s)he who makes Ralph's belief expressed by (12) true; hence Ralph believes of her/him that (s)he is a spy; therefore (3′) is true.

The equivalence of (2′) and (3′) highlights an important fact: elementary logic does not contain the resources for representing the ESN distinction,[87] if it is impossible to devise a way to represent it better than Quine's under this respect.

A natural reaction to this situation would be to resort to pragmatics to settle the problem of representing the ESN distinction. However, there is some reason to hold that in doxastic contexts the SND gives rise to a *semantical* distinction. Consider the following situation. The cashier of a bar notices that she has cashed a false $100 banknote. She thinks at once of Mr. Lapin, whose evil repute she knows very well and who paid with a €100 banknote just that morning, and she persuades herself

---

[85] Fodor & Sag, speaking about (9), argue for the logical equivalence of the wide scope (existential) quantifier reading and a 'referential' reading according to which "a student" is understood as «a referring expression such as a proper name or demonstrative phrase»:

> If some particular student has cheated, then the set of students who have cheated is not empty; and if the set of students who have cheated is not empty, then some particular student must have cheated. (Fodor & Sag, 1982: 356)

This might give the impression that, whenever a sentence of the form $\exists x\alpha$ is a theorem of a theory based on classical logic, there is a term t such that $\alpha[t/x]$ is a theorem. This is not the case; for example, take Peano Arithmetic PA, and let $\alpha[x]$ be the formula $(G \rightarrow (x = 1))\&\wedge(\neg G \rightarrow x = 0)$, where G is Gödel's indecidable sentence: then $\vdash_{PA} \exists x\alpha$, but there is no closed term t such that $\vdash_{PA} \alpha[t/x]$, since PA does not settle G. See also the remarks made in the main text about (7).

[86] Unless we admit such specifications as "the girl who kissed all boys"; I shall discuss this possibility below.

[87] I agree with S. Crawford that Quine's question in Quine (1956) was: «Does elementary logic contain the resources for representing these two kinds of states mind [i.e. notional and relational]?» (Crawford, 2008: 77).

that he is a counterfeiter. As a matter of fact the counterfeiter is another customer, Mr. Lepen, whom the cashier judges a person above suspicion. In this situation it is true both that

(85) The cashier justifiedly believes that a customer of the bar is a counterfeiter

and that

(86) A customer of the bar is a counterfeiter;

so the cashier justifiedly believes a true proposition, but she does not *know* that a customer of the bar is a counterfeiter: a typical Gettier counterexample to the Platonic analysis of knowledge as justified true belief.

Consider now a variant of the preceding situation. The cashier, after having noticed that she has cashed a false banknote, reflects that only a customer of the bar can have given her it, and on this basis she persuades herself that a customer of the bar is a counterfeiter, which is true. The interesting thing to note is that in this situation it is perfectly correct to say not only that (85) and (86) are true, but also that

(87) The cashier *knows* that a customer of the bar is a counterfeiter.

Since in the former situation (85) is true under the specific interpretation, while in the latter it is true under the non-specific interpretation, it seems legitimate to generalize from this example by saying that the *meaning* of

(88) S knows that α,

where α is an existential sentence, can be equated to the meaning of

(89) S justifiedly and truly believes that α

when the indefinite occurring in α is interpreted non-specifically,[88] but not when it is interpreted specifically. Now, in general, the meaning of "S knows that α" is different form the meaning of "S knows that β" if, and only if, the meaning of α is different form the meaning of β. Hence, if the meaning of "S knows that α", when the indefinite occurring in α is interpreted non-specifically, is different from the meaning of "S knows that α", when the indefinite occurring in α is interpreted specifically, this must depend on the fact that the *meaning* of α, when the indefinite occurring in α is interpreted non-specifically, is different form the *meaning* of α, when the indefinite occurring in α is interpreted specifically. Since meaning falls within the competence of semantics, the ESN ambiguity should be treated as a semantical phenomenon.

It seems legitimate to conclude that, in order to give a semantical representation of the ESN distinction, it is necessary to give up some deeper assumption, i.e. some assumption that is constitutive of the semantic paradigm traditionally adopted. This is the direction I will follow.

[88] Or, at least, such equation is immune from Gettier's counterexamples.

7.4.2 *Two Kinds of Cognitive States (and Assertion Conditions)*

We have observed above that when we translate a sentence of a natural language like (9) into a sentence of $\mathcal{L}$ of the form ∃xα, its truth-conditions are such that we are necessarily representing the non-specific reading of the indefinite. However, it can be observed as well that (9) can be asserted by John in circumstances, or contexts, of two distinct types. John might have in mind a student, for instance Mary, because he has seen her cheating on the exam. Alternatively, John might not have in mind a specific student, but he might persuade himself of the truth of (9) because there are two almost identical school works; in a context of this second type John simply asserts by (9) that the set of the students who cheated is not empty.

Is it possible to give a *general* characterization of the two types of context in which (9) can be asserted, i.e. a characterization that can be applied to every sentence for which the ESN distinction is meaningful? It seems to me that the semantics adopted for the TO distinction can give us some valuable hints in this sense, if we remember that cognitive states, besides being justifications for sentences, can play as well the role of contexts in which those sentences can be asserted; for, from the internalist viewpoint that informs this semantics, a context is nothing but an amount of information available to the subject the believer and coming from various mental components (memory, imagination, the senses, etc.), hence a part of a cognitive state.

7.4.2.1 Direct Cognitive States

According to the explanation of the existential quantifier given in Chap. 5, a subject occupying a cognitive state that is a justification for ∃xα is in principle in a position to know, hence to exhibit, some element d of the domain and a justification of the fact that it satisfies $\alpha[\underline{d}/x]$. On the other hand, if we consider the first type of circumstances in which (9) is (justifiedly) assertible, we realize that we have found what we were looking for: John—we had said—asserts (9) having in mind Mary, because he has seen her cheating during the exam; the clause that, in the semantics given in Chap. 5, defines the justifications for ∃xα suggests to replace 'having in mind Mary' with John's being in a position to exhibit (i) Mary and (ii) a justification for "The student Mary in the syntax class cheated on the final exam" (which might be, for instance, a cognitive state of John's in which he associates to the name "Mary" some pieces of information: that she is a student in the syntax class, that he saw her cheating, etc.).[89]

Hence, a first approximation to a general characterization of the cognitive states of the former kind—which I will call *direct*—would be the following:

[89] Of course John may have very good (psychological) reasons for asserting (9) and not the singular sentence "Mary, who is a student in the syntax class, cheated on the final exam" (for example, he might want not to expose Mary to a punishment). But these reasons are not cognitive: from an epistemic point of view the reasons John has for asserting (9) and the singular sentence are exactly the same.

(90) A cognitive state σ that is a justification for an indefinite sentence[90] α of a natural language in which the indefinite i occurs is *direct* if there is a term t such that σ is a justification for α[t/i].

This characterization avoids any reference to such psychological notions as ‘having in mind’ and ‘intention to refer’: it is sufficient to say that the speaker has a justification to assert

(91) α[t/i],

for some term t. That this is one of its virtues can be appreciated through the following example. A mathematics teacher is giving a lecture; at a certain point he asserts:

(92) A prime number is even. It is smaller than 5.

At the time of utterance he is so absent-minded that he is thinking of, and intends to refer to, the number 4. The pupils, however, reflect upon the definition of prime numbers the teacher has given in a previous class and persuade themselves that the indefinite in (92) is true of the number 2. In this situation it is intuitively legitimate to say that when the class is over the pupils know that *the number 2* is a prime number, even if the teacher didn’t have it in mind, nor intended to refer to it. The moral to draw is that in certain circumstances it is not the speaker’s intention that determines the proposition (s)he is in a position to assert, but her/his *justification* to believe it. As a consequence, the fact that a subject is authorized to believe a proposition does not depend on her/his psychological idiosyncrasies, but only on the structure of the proposition and of the justification—more specifically, on the speaker’s cognitive structure, and on information available to her/him.

However, (90) can only be a first approximation to a definition, because not all terms can be admitted as warranting that the indefinite is used specifically. In the case of (9), for instance, the demonstrative “she” (together to a gesture pointing to Mary), the proper name “Mary” and also a description like “Paul’s girlfriend” (under the assumption that Mary is Paul’s girlfriend) would give admissible instances of (91); but take the description “the student in the syntax class who cheated on the final exam”: the sentence

(93) The student in the syntax class who cheated on the final exam is a student in the syntax class and (s-)he cheated on the final exam

would be an instance of (91) but, if we admitted it, (90) would no longer have any discriminating power, since the description in (93) can be used to refer to that student (if in fact there is one and only one student in the syntax class who cheated) in *every* context.[91] What we are looking for are terms (including descriptions) whose use by

[90] Terminological note: I shall call “indefinite sentence” a natural language sentence containing indefinite terms that can be used to express existential quantification; some examples in English: “A woman/some woman/someone entered the room”, “John killed a bird/two birds, several birds”.

[91] Because, when descriptions are defined according to Russell’s theory, the sentence translating (93) into first order predicate logic is a theorem (under the assumption that there is only one student in the syntax class who cheated on the final exam); when descriptions are treated as singular terms, the sentence translating (93) into ε-calculus is a theorem of that calculus (under the same assumption).

the speaker certifies her/his capacity to distinguish the object (s)he has in mind from all the others; but the problem of circumscribing the class of such terms is not trivial. I shall suggest here a property only as a sufficient condition for a description to be admissible. I call this property *freshness,* and I define it as follows:

Definition 11 Given a justification σ for a sentence α, a description *d* is *fresh* relative to σ iff it is not formed with any of the predicates or singular terms that occur in α or in a sentences occurring in a verbal expression of information available in σ, nor with predicates or singular terms synonymous with them.

By means of the freshness condition we can exclude from the instances of (91) not only (93) but also, for example, the sentence (94), as intuition suggests:

(94) One of the students who made these two school works.

These considerations suggest the following definitions:

Definition 12 A cognitive state σ that is a justification for an indefinite sentence α of a natural language in which the indefinite i occurrs is *direct* iff there is a singular term t that is fresh relative to σ and such that σ is a justification for α[t/i].

Definition 13 When an indefinite sentence α of a natural language is asserted by a subject $\mathcal{S}$ occupying a cognitive state that is a direct justification for α, $\mathcal{S}$ is authorized to use the indefinite occurring in α in its specific reading.

Let me conclude with an example that highlights a less visible aspect of this definition. Suppose that Bond, before leaving his room, has left a camera pointed towards the door; when he comes back he watches the video film and sees a woman passing through the door; later on he forgets all the details and remembers only that the person who entered was a woman. In such a situation Bond has no specific woman in mind but, according to Definition 13, he is authorized to use the indefinite "a woman" in its *specific* reading when he asserts

(95) A woman entered the room,

because he has a direct justification for (95); and he has such a justification because he knows a procedure (i.e., watching the film again) whose execution would make available to his mind, after a finite time, a specific woman d and a justification for "$\underline{\text{d}}$ entered the room".

We have now a general characterization of the contexts of the former kind, or direct contexts, in which the proper use of the indefinite is the specific one.

7.4.2.2 Indirect Cognitive States

Let us pass to the second type of context. In our example John comes to persuade himself of the truth of (9) by means of a reasoning that is, I suggest, essentially indirect. It can be reconstructed approximately in the following way: John has a justification for believing that if nobody in the syntax class had cheated on the final exam there would not be two almost identical school works; since two almost identical school works do exist, he draws the conclusion that it is impossible that nobody in the syntax class has cheated. His reasoning can be verbally reconstructed as follows:

(96)

$$
\begin{array}{l}
[\forall x(S(x) \rightarrow \neg C(x))]^1 \\
\qquad \vdots \\
\hline
B \rightarrow \bot \qquad\qquad B^2 \\
\hline
\qquad \bot \\
\hline
\neg\forall x(S(x) \rightarrow \neg C(x)) \quad {}^{,\,1} \\
\qquad \vdots \\
\hline
\neg\forall x \neg(S(x) \wedge C(x))
\end{array}
$$

where “S(x)” abbreviates “x was a student in the syntax class”, “C(x)” abbreviates “x cheated on the final exam” and B abbreviates “There are two almost identical school works”. The conclusion is logically equivalent to the translation of (9) into a standard first order language.

I hope it is clear now what I mean by saying that in the second case the subject’s reasoning is indirect: in (96) John derives an absurdity from the assumption that nobody cheated on the exam. Therefore the justification John has in this situation for believing that (9) is true is a method for deriving an absurdity from the assumption that (9) is false.

One may wonder whether John’s reasoning in the second case is *essentially* indirect. Let us consider another example. A guard in a gallery hears the signal given by the alarm system whenever a painting is taken from its place, and says:

(97) A painting is missing from the gallery.

In this case the most natural reconstruction of the reasoning that justifies the guard in believing the proposition seems to be the following: «If the alarm system gives a signal then a painting has been taken from its place; the alarm system does give a signal; therefore a painting has been taken from its place; therefore a painting is missing from the gallery.» Here the indirect character of (96) is avoided by appealing to *modus ponens*. Admittedly this reconstruction is legitimate and natural. The problem with it is that it is entirely useless for our purposes, since it makes, of the indefinite “a painting”, the same non-specific use as does the guard when (s)he asserts (97). What we need is a reconstruction of the guard’s justification that permits us to understand her/his non-specific use of the indefinite; of course such a reconstruction cannot contain in its turn indefinites used non-specifically, in order not to presuppose as explained what it is intended to explain: we need a 'non-question-begging' reconstruction; and a reconstruction meeting this restriction has essentially the same indirect structure as (96): «If no painting were missing from the gallery the alarm system would not give a signal; the alarm system does give a signal; therefore a painting is missing from the gallery.»

We can therefore give the following definitions:

Definition 14 A cognitive state σ that is a justification for an indefinite sentence α of a natural language is *indirect* if in σ a method is available for deriving a contradiction from the assumption that α is false.

Definition 15 When an indefinite sentence α of a natural language is asserted by a subject $\mathcal{S}$ occupying a cognitive state that is an indirect justification for α, $\mathcal{S}$ is authorized to use the indefinite occurring in α in its non-specific reading.

Applying these definitions to (9), for instance, we see that the assertibility conditions of the non-specific reading fit exactly von Heusinger's intuitive characterization quoted above: the subject occupies a cognitive state in which a method is available for deriving a contradiction from the assumption that no student in the syntax class cheated on the final exam, hence what (s)he is authorized to assert is only that it is false that the set of students in the syntax class who cheated on the final exam is empty, i.e. that the set of students in the syntax class who cheated on the final exam is not empty.

7.4.3 *Formalization*

We have given a general characterization of the two types of context in which an indefinite sentence can be asserted, exploiting what seems to be a systematic correlation between those two types and the assertibility conditions, in the semantics given in Chap. 5 as well as in Heyting's functional semantics, of first order sentences of two forms: $\exists x\alpha$ and $\neg\forall x\neg\alpha$. If we keep in mind that the truth-conditions of those two sentence-forms are the same, the observed correlation strongly suggests that the distinction between the assertibility conditions of $\exists x\alpha$ and $\neg\forall x\neg\alpha$ is blurred by the explanation of meaning in terms of truth-conditions.

This impression is confirmed by the following well-known fact about the relations between classical and intuitionistic logic. It is possible to define translations TR from classical logic **CL** *into* intuitionistic logic **IL**, usually called "negative translations", such that $\vdash_{\mathbf{CL}} A$ iff $\vdash_{\mathbf{IL}} TR(A)$; one of them is due to Gödel and Gentzen[92] and is defined by the following clauses:

[92] That's why I call it "GG". A detailed history of Gödel's 'negative translation' and of its extensions is contained in Troelstra's introduction to Gödel's article "Zur intuitionistischen Arithmetik und Zahlentheorie", in Gödel (1986: 282–7). For the other negative translations see Ferreira and Oliva (2012).

$$
(98)\quad
\begin{aligned}
\beta^{GG} &= \neg\neg\beta, \text{ for } \beta \text{ atomic}\\
(\beta \wedge \gamma)^{GG} &= \beta^{GG} \wedge \gamma^{GG}\\
(\beta \to \gamma)^{GG} &= \beta^{GG} \to \gamma^{GG}\\
(\neg\beta)^{GG} &= \neg\beta^{GG}\\
(\beta \vee \gamma)^{GG} &= \neg\left(\neg\beta^{GG} \wedge \neg\gamma^{GG}\right)\\
(\forall x\beta)^{GG} &= \forall x\beta^{GG}\\
(\exists x\beta)^{GG} &= \neg\forall x\neg\beta^{GG}
\end{aligned}
$$

In view of this translation intuitionistic logic, which syntactically is a subsystem of classical logic, can be conceived from the semantical point of view as an *extension* of classical logic, in the sense that intuitionistic logic can be seen as containing *two* existential quantifiers: its own $\exists$, and the classical one Σ, which is defined within it in terms of $\forall$ and $\neg$ ($\Sigma x\alpha =_{def} \neg\forall x\neg\alpha$); and two disjunctions: its own $\vee$, and the classical one $+$, which is defined within it in terms of $\wedge$ and $\neg$ ($\alpha + \beta =_{def} \neg(\neg\alpha\wedge\neg\beta)$). Therefore we could say that from the semantical point of view intuitionistic logic is the result of adding to classical logic two new logical constants: the constructive existential quantifier $\exists$ and the constructive disjunction $\vee$.[93]

Gödel-Gentzen's translation and the preceding remarks suggest that the intuitionistic existential quantifier can be used to represent the specific reading of the natural language indefinite sentences. This idea is supported by two facts. The first is the well-known fact that most mathematical theories based on intuitionistic logic have the existence property:

---

[93] This view is explicitely endorsed by Heyting:

> Constructivist mathematicians in the strict sense […] reserve the word "exist" for constructive existence. Then (Ex)A(x) can only be asserted after the construction of a b such that A(b). They object to the use of the word "exist" to indicate non-constructive existence, because this terminology suggests the idea of a mathematical reality independent of our knowledge. The assumption of such a mathematical reality is metaphysical in nature; the goal of constructivists is to free mathematics from such philosophical assumptions. […].
>
> To summarise, I say that (Ex)A(x) will mean that a mathematical object b such that A(b) has been constructed; ¬¬(Ex)A(x) will mean that the assumption that no object satisfies A(x) has been reduced to absurdity. (Heyting, 1960: 177–178)
>
> [Les mathématiciens constructivistes au sens plus strict […] réservent le mot "exister" pour l'existence constructive. Alors (Ex)A(x) ne peut être affirmé qu'après la construction d'un b tel que A(b). Ils s'opposent à l'emploi du mot "exister" pour indiquer l'existence non-constructive, parce que cette terminologie suggère l'idée d'une réalité mathématique indépendente de notre connaissance. L'hypothèse d'une telle réalité mathématique est de nature métaphysique; le but des constructivistes est de libérer les mathématiques de telles hypothèses philosophiques. […].
>
> Pour résumer, je dis que (Ex)A(x) signifiera qu'un objet mathematique b tel que A(b) a été construit; ¬¬(Ex)A(x) signifiera que l'hypothèse qu'aucun objet ne satisfait à A(x) a été réduite à l'absurde.].

Notice that $\neg\neg\exists x\alpha$ and $\neg\forall x\neg\alpha$ are intuitionistically equivalent.

(EP) If a sentence $\exists x\alpha$ is a theorem, then $\alpha[t/x]$ is a theorem, for some term t,

which expresses exactly the requirement a justification for $\exists x\alpha$ must meet in order to be direct. The second fact is that

(99) Gödel-Gentzen's translation of **CL** into the logic **L** corresponding to the semantics given in Chap. 5 is such that $\vdash_{\mathbf{CL}}\alpha$ iff $\vdash_{\mathbf{L}}\alpha^{GG}$.[94]

The following formalization strategy is therefore justified:

(100) (i) Whenever α is an indefinite sentence of a natural language, its formalization takes as input not simply α, but ordered pairs $\prec\alpha, \sigma\succ$, where σ is a justification for α;

(b) given a pair $\prec\alpha, \sigma\succ$, the standard translation of α into $\mathcal{L}$ will contain $\exists$ or Σ according to whether σ is direct or indirect, respectively.

I remind the reader that $\exists x\alpha$ and $\neg\forall x\neg\alpha$ are not logically equivalent in intuitionistic logic. It is precisely in virtue of this non-equivalence that intuitionistic logic, as opposed to classical logic, does contain the resources for representing the ESN ambiguity, not as a phenomenon of scope, but as a sort of lexical ambiguity. In this sense we can say that the adoption of an 'intuitionistic' semantics enables us to give a finer analysis than classical logic of the meanings of natural language indefinite and disjunctive sentences. Moreover, the commonsense point of view seems in this case to endorse an intuitionistic attitude in positing a difference in meaning between the specific and the non-specific reading of indefinite sentences that realist truth-conditional semantics does not recognize; for, while it is perfectly natural to say, for example, that (9) in its specific reading 'speaks of' Mary, it seems intuitively impossible to say that, in its non-specific reading, (9) speaks or 'is about' any object at all.

It is interesting to observe that it is now possible to represent the specific interpretation of (84) as well. I have said that (84′) cannot represent it, but the existential quantifier occurring in (84′) was the classical one, which is now translated into the intuitionistic language as Σ, while $\exists$ is the intuitionistic existential quantifier; hence the sentence that does *not* represent the specific reading of (84) is now, within an intuitionistic framework,

(84‴) $\Sigma x\forall y((\mathrm{GIRL}(x)\wedge\mathrm{BOY}(y)) \supset \mathrm{KISSED}(x,y))$,

while (84′) represents now just the specific reading.

---

[94] In order to prove this fact, it is sufficient to observe first of all that in the semantics given in Chap. 5 all the principles of minimal logic are valid, and that in minimal logic it holds that (*) If α is a negative formula (i.e. not containing $\vee$ or $\exists$ and such that in it atomic formulae only occur negated), then $\vDash\alpha\longleftrightarrow\neg\alpha$.

At this point (99) can be proved by induction on the complexity of α; the basis is guaranteed by (*); the inductive cases are proved by principles valid in minimal logic.

7.4.4 Answering Question (8)

In order to answer question (8) it is necessary to extend the strategy described above to belief reports whose subordinate clause is an indefinite sentence. Consider for instance (10): which are its assertibility conditions? We have seen that the TOD is meaningful for it; moreover, the ESND is meaningful for its subordinate clause; the theory therefore predicts that it is four-ways ambiguous[95]:

(101)
1. $B\big(\mathcal{Bond}, \text{that}\ _{\mathcal{Bel}}\exists x(W(x) \wedge E(x))\big)$ **S + O**
2. $B\big(\mathcal{Bond}, \text{that}\ _{\mathcal{Bel}}\Sigma x(W(x) \wedge E(x))\big)$ **N + O**
3. $B\big(\mathcal{Bond}, \text{that}\ _{\mathcal{Rep}}\exists x(W(x) \wedge E(x))\big)$ **S + T**
4. $B\big(\mathcal{Bond}, \text{that}\ _{\mathcal{Rep}}\Sigma x(W(x) \wedge E(x))\big)$ **N + T**

We can now answer question (8):

(102) Inference (Part) is correct if, and only if, the cognitive state observed by the reporter asserting "s believes that α", where α is an indefinite sentence, is direct (hence the formalization of the report is either "$B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\exists x\alpha(x))$" or "$B(\mathcal{S}, \text{that}_{\mathcal{Rep}}\exists x\alpha(x))$").

The sentence "$B(\mathcal{S}, \text{that}_{\mathcal{Bel}}\exists x\alpha(x))$" formalizes the case in which it is the believer who has a specific object in mind, while "$B(\mathcal{S}, \text{that}_{\mathcal{Rep}}\ \exists x\alpha(x))$" formalizes the case in which it is the reporter.

7.4.5 On the Vast Difference

What about the vast difference invoked by Quine between the notional and the relational reading of belief reports? We know that Quine's final reaction to the derivation of (3) from (2) was to give up the intuitive idea that the vast difference exists.[96] Many found this position difficult to swallow[97]; I agree with them, but not with the ways that have been suggested to save the intuitive vast difference. All these ways consist in restrictions imposed onto the exportation principle, in order to block the derivation; I propose a different analysis.

First, observe that the cognitive state warranting Ralph's belief that someone is a spy (in the sense of "someone is a spy" in which it is equivalent to "there are spies") is clearly indirect: Ralph would not be capable to exhibit a person and to assert that that person is a spy; the justification he has for his belief can be verbally reconstructed rather as follows: «If there were no spies, α would not be true (where α is for example: "A state's military secrets are discovered by another state"); but α

[95] The following are intended to be formal representations of the four readings of (10) distinguished in Sect. 7.1.2.

[96] «At first this seems intolerable, but it grows on one.» (Quine, 1979: 273).

[97] See for instance Kvart (1982: 298), Recanati (2000: 377), Kripke (2011: 326, fn. 17).

is true; hence it is false that there are no spies.» As a consequence, the formalization of the report of Ralph's belief is neither

(103) $B(\mathcal{Ralph}, \text{that}_{\mathcal{Bel}}\exists xSPY(x))$

nor

(104) $B(\mathcal{Ralph}, \text{that}_{\mathcal{Rep}}\exists xSPY(x))$,

but

(105) $B(\mathcal{Ralph}, \text{that}_{\mathcal{Bel}}\ \Sigma xSPY(x))$

or

(106) $B(\mathcal{Ralph}, \text{that}_{\mathcal{Rep}}\ \Sigma xSPY(x))$[98]:

in both cases the content of Ralph's belief is expressed by a sentence of the form $\Sigma x\alpha$. As we have seen, within the framework of the semantics given in Chap. 5 Ralph's mental states represented by (105) and (106) are deeply different from the ones represented by (103) and (104).

Second, consider more closely the derivation of (3) from (2) and (11); the reasons why (12) is inferred from (2) and (11) are (i) that "The shortest spy is a spy" follows from the two premises "someone is a spy" and "no two spies are of exactly the same height", and (ii) that Ralph is 'minimally reasonable', so that he realizes (i), hence believes the conclusion since he believes the premises. The exact way "The shortest spy is a spy" follows from the two premises depends on how the description "The shortest spy" is treated. If we adopt Russell's theory of descriptions it follows provided that the two premises are formalized by (107) and (108):

(107) $\exists x(SPY(x))$

(108) $\forall x\forall y((SPY(x)\wedge SPY(y)) \rightarrow ((SHORTER(x,y)\wedge SHORTER(y,x)) \rightarrow y = x))$;

for in this case we can infer

(109) $\exists x(SHORTEST\ SPY!(x)\wedge SPY(x))$,[99]

which formalizes "the shortest spy is a spy". If we treat descriptions as singular terms, we can extend the language of predicate calculus with the ε-symbol and the ι-symbol, associate to the new language the ε-axiom $\exists x\alpha \rightarrow \alpha(\varepsilon x\alpha)$, from which we obtain the theorem

(110) $\exists x\alpha! \rightarrow \alpha(\iota x\alpha)$[100];

an instance of this theorem entails

(111) $\exists x(SHORTEST\ SPY!(x) \rightarrow SPY(\iota xSHORTEST\ SPY(x)))$,

[98] The difference between (48) and (49) is inessential in the present context.

[99] "SHORTEST SPY!(x)" abbreviates "$SHORTEST\ SPY(x)\wedge\forall y(SHORTEST\ SPY(y) \rightarrow y = x)$".

[100] See Leisenring (1969: 101).

which, together with ∃x(SHORTEST SPY!(x)), entails

(112) SPY(ιx(SHORTEST SPY(x))).

What is relevant here is that in both cases the premise (107) must hold for the conclusion to hold; but we have seen that the first premise of (i), i.e. the content of Ralph's belief, cannot be formalized as (107); as it can be seen through (105) and (106), the formalization can only be

(105′) Σx(SPY(x));

and from (105′) and (108) it does not follow, intuitionistically, (109), but only

(107′) Σx(SHORTEST SPY!(x)∧SPY(x)).

The derivation of (112) from (110) is blocked in an analogous way.

The introduction of the description "the shortest spy" may give the impression that (12) asserts that Ralph has a person in mind and has a justification to believe that that person is a spy. But this is incorrect. The reason why "the shortest spy is a spy" follows from "There is exactly one shortest spy" (hence why Ralph justifiedly believes that it follows) is purely logical, in the sense that (111) is a logical consequence of the logical theorem (110). "The shortest spy" denotes an *ideal* object in Hilbert's sense, an indeterminate object of which we only know that it is the unique element of {x|SPY!(x)}. This entails, on the one hand, that we are not able to use a fresh term to refer to that individual which, as a matter of fact, exists and is unique; on the other hand, that when we say that Ralph believes that the shortest spy is a spy, we are saying that his belief is about a set, the set S of spies, and is the belief that S contains an element that is minimum with respect to height. This intuitive interpretation seems to be supported by the fact that ∃xα! is classically equivalent to ¬∀x¬α!, and it seems intuitively untenable that this formula speaks of an individual.

Summing up, the derivation of (12) from (2) and (11) is a (correct) derivation of the *general* belief that the set of spies contains a height-minimum, not of the singular belief that a certain individual is a spy; if we want to derive the singular belief, represented by (107), the derivation is blocked because what can be derived from (2) and (8) is not (107) but (105′), which is not equivalent to (107) in the semantics. The intuitive vast difference registered by Quine is thus vindicated, without any restriction imposed onto (tExp). This does not mean that exportation (or the passage from opaque to transparent reading of belief reports) is not a problem (we have seen that it is, and I have suggeste an answer); it only means that blocking the derivation of (3) from (2) by blocking unrestricted exportation has the price of blurring the difference between the TOD and the ESND, two distinctions that *are* different. Quine's vast difference between the notional and the relational sense of belief reports is the difference between general and specific belief, not the difference between opaque and transparent belief.

7.5 Conclusion

I have criticized the traditional formulation of the question «When does the notional formulation of a belief report imply the relational formulation?» because it presupposes Quine's idea that the *De Dicto/De Re* distinction is the *explicans* of the notional/relational distinction, where the *De Dicto/De Re* distinction conflates two different distinctions, the TO one and the ESN one, each of which should be analyzed and represented separately.

The analysis of the solutions proposed to Mates' problem suggests that no optimal solution can be found within externalist semantics. The essential reason can be stated by saying that, if $\nu = \nu'$ and $\mathcal{S}$ believes that $\nu \neq \nu'$, externalist semantics provides no way of representing *what* $\mathcal{S}$ believes without falling into the absurdity of ascribing to $\mathcal{S}$ an irrational belief. If this conclusion is accepted, we have the sketch of an argument against externalist semantics:

(i) If we want to give an account of scientific practices based on (DP_j) and (*), we must be able to distinguish between believing that $\alpha[\varepsilon]$ and believing that $\alpha[\varepsilon'//\varepsilon]$;
(ii) In an externalist semantics, if two arbitrary expressions ε and ε' are synonymous, $\alpha[\varepsilon]$ and $\alpha[\varepsilon'//\varepsilon]$ express the same proposition;
(iii) therefore, in an externalist semantics we cannot distinguish between believing that $\alpha[\varepsilon]$ and believing that $\alpha[\varepsilon'//\varepsilon]$;
(iv) Concluding, in an externalist semantics we we cannot give an account of scientific practices based on (DP_j) and (*).

I have proposed a change of semantic paradigm consisting in replacing the absolute semantic notions of externalist semantics with notions relativized to subjects, in defining these notions within an internalist framework, and in showing how optimal solutions to the foundational puzzles can be found. More specifically, I have proposed to conceive the transparent/opaque ambiguity as concerning not two kinds of belief, but the two propositions ambiguously denoted by the "that" that introduces the subordinate clause of the belief report: the proposition expressed by the subordinate clause for the believer, and the one expressed and for the reporter; the condition for legitimately inferring the transparent reading from the opaque one is that the two propositions are the same.

I have proposed to conceive the ESN ambiguity as a lexical ambiguity concerning the indefinite "a NP", translated into $\mathcal{L}_B$ by means of the two different quantifiers $\exists$ or Σ, according as the context of the assertion is direct or indirect; the inference of Σ translation from $\exists$ translation is legitimate, the converse not. In this way the intuitive vast difference between (epistemically) specific and non-specific reading of the indefinite NP is vindicated.

In the semantic representation of both ambiguities a crucial role is played by the justificationist semantics adopted. In this sense a justificationist semantics appears as emprically more adequate than a truth-conditional one to represent important aspects of the meaning of indefinites and of belief reports.[101] This is the core of an argument for a justificationist semantics different from the traditional intuitionistic and neo-verificationist ones, like Dummett's anti-realist argument and Prawitz's argument against the Tarskian analysis of logical conseguence.[102]

The change of paradigm I have proposed points in the same direction suggested by Chomsky: towards an internalist theory of meaning. We have seen in Chap. 1 that Chomsky's reasons—as they emerge in Chomsky (2000)—are essentially of a methodological nature: only if the entities denoted by names and predicates are of a mental nature is it possible to conceive knowledge of meaning as a system of computational structures and processes, thereby giving a scientific account of semantic competence, i.e. of cognitive preconditions of linguistic use. I want to put into evidence another aspect of the change of paradigm: the reasons that justify an internalist semantics converge in a significant way with the ones that justify Heyting's intuitionistic theory of the meaning of the logical constants—a theory traditionally classified as anti-realist, but which would more appropriately be called *mentalistic*. As we have seen, Heyting explains the meaning of each logical constant **C** by specifying what a proof of an arbitrary sentence having **C** as its principal operator amounts to, and he postulates that knowing the meaning of such a sentence amounts to being capable to recognise its proofs; since such proofs are (inductively) defined in terms of specific mental operations on entities having the nature of mental representations (in the non-relational sense of "representation"), Heyting's explanation presupposes in fact several hypotheses on the internal structure of our deductive faculty. Therefore, internalist semantics and intuitionistic theory of meaning can be seen as having a common object of theoretical interest: mind and its internal organization.

References

Boghossian, P. (1994). The transparency of mental content. *Philosophical Perspectives, 8*, 33–50.

Bonomi, A. (1983). *Eventi Mentali*. il Saggiatore.

Burge, T. (1978). Belief and synonymy. *The Journal of Philosophy*, *LXXV*(3), 119–138.

Carnap, R. (1947). *Meaning and necessity*. The University of Chicago Press. (Second Edition 1956).

Carnap, R. (1949). A reply to Leonard Linsky. *Philosophy of Science*, *XVI*, 347–350.

Carnap, R. (1954). On belief sentences. Reply to Alonzo Church. In M. Mcdonald (Ed.), *Philosophy and analysis.* Oxford. (Now in Carnap (1947), Second Edition (1956), 230–232)

Carruthers, P., & Smith, P. K. (Eds.). (1996). *Theories of theories of mind*. Cambridge University Press.

[101] Cp. Heyting's remark that «Intuitionistic logic makes finer distinctions possible, which classical two-valued logic is unable to express.» (Heyting, 1974: 88).

[102] See for instance Dummett (1976) and Prawitz (2005: 675).

Cartwright, R. (1971). Identity and substitutivity. In M. K. Munitz (Ed.), *Identity and individuation* (pp. 119–133). New York University Press.
Casalegno, P. (1997). *Filosofia del linguaggio*. La Nuova Italia Scientifica.
Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge University Press.
Church, A. (1946). Review of the black/white controversy. *The Journal of Symbolic Logic, XI*, 132–133.
Church, A. (1954). Intensional isomorphism and identity of belief. *Philosophical Studies, V*(5), 65–73.
Crawford, S. (2008). Quantifiers and propositional attitudes: Quine revisited. *Synthese, 160*, 75–96.
Dummett, M. (1976). What is a theory of meaning? (II). In G. Evans & J. Mcdowell (Eds.), *Truth and meaning: Essays in semantics* (pp. 67–137). Clarendon Press. (Now in Dummett (1993) (pp. 34–93)].
Dummett, M. (1981). *Frege. Philosophy of language* (2nd ed.). Duckworth.
Dutilh Novaes, C. (2004). A medieval reformulation of the *de dicto/de re* distinction. In L. Behounek (Ed.), *Logica yearbook 2003* (pp. 111–124). Filosofia.
Farkas, D. (1996). Specificity and scope. In L. Nash & G. Tsoulas (Eds.), *Langues et Grammaire 1* (pp. 119–137). Paris.
Ferreira, G., & Oliva, P. (2012). On the relation between various negative translations. In U. Berger & H. Schwichtenberg (Eds.), *Logic, construction, computation* (pp. 227–258). Ontos-Verlag.
Fodor, J. D. (1970). *The linguistic description of opaque contexts*. Ph.D. Thesis. Massachusetts Institute of Technology.
Fodor, J. D., & Sag, I. A. (1982). Referential and quantificational indefinites. *Linguistics and Philosophy, 5*, 355–398.
Frege, G. (1892). *On sense and meaning*. In Frege (1984) (pp. 157–177).
Frege, G. (1918). *Thoughts*. In Frege (1984) (pp. 351–372).
Gödel, K. (1986). *Collected works* (Vol. 1). Oxford University Press.
Heyting, A. (1960). Remarques sur le constructivisme. *Logique Et Analyse, 3*, 177–182.
Heyting, A. (1974). Intuitionistic views on the nature of mathematics. *Synthese, 27*, 79–91.
Hindley, J. R. (2015). *The root-2 proof as an example of non-constructivity*. http://www.users.waitrose.com/~hindley/Root2Proof2015.pdf
Horwich, P. (1998). *Meaning*. Oxford University Press.
Kamp, H., & Bende-Farkas, Á. (2019). Epistemic specificity from a communication-theoretic perspective. *Journal of Semantics, 36*, 1–51.
Kaplan, D. (1969). Quantifying in. In D. Davidson & G. Harman (Eds.), *Words and objections* (pp. 206–242). Reidel.
Kripke, S. (1979). A puzzle about belief. In A. Margalit (Ed.), *Meaning and use* (pp. 239–283). Reidel.
Kripke, S. A. (2011). Unrestricted exportation and some morals for the philosophy of language. *Philosophical troubles* (pp. 322–350). Oxford University Press.
Kvart, I. (1982). Quine and modalities De Re: A way out? *The Journal of Philosophy, 79*, 295–328.
Langford, C. H. (1942). The notion of analysis in Moore's philosophy. In P. Schilpp (Ed.), *The philosophy of G. E. Moore* (pp. 321–342).
Leisenring, A. C. (1969). *Mathematical logic and Hilbert's ε-Symbol*. MacDonald & Co.

Linsky, L. (1949). Some notes on Carnap's concept of intensional isomorphism and the paradox of analysis. *Philosophy of Science, XVI*, 323–347.

Mates, B. (1950). *Synonymity*. University of California Publications in Philosophy, XXV. (Reprinted in L. Linsky (Ed.), *Semantics and the philosophy of language* (pp. 111–138). University of Illinois Press, 1970).

Moffett, M. (2002). A note on the relationship between mates' puzzle and frege's puzzle. *Journal of Semantics, 19*, 159–166.

Neale, S. (2000). *On a milestone of empiricism*. In Orenstein & Kotatko (2000) (pp. 237–346).

Orenstein, A., & Kotatko, P. (Eds.). (2000). *Knowledge, language and logic: Questions for Quine*. Kluwer.

Partee, B., et al. (1973). The semantics of belief-sentences. In J. Hintikka (Ed.), *Approaches to natural language* (pp. 309–336). Reidel.

Partee, B. (1974). Opacity and Scope. In M. Munitz & P. Unger (Eds.), *Semantics and philosophy* (pp. 81–101). New York University Press. (Reprinted in P. Ludlow (Ed.), *Readings in the philosophy of language* (pp. 833–853). The MIT Press, 1997).

Prawitz, D. (2005). Logical consequence from a constructivist point of view. In S. Shapiro (Ed.), *The Oxford handbook of philosophy of mathematics and logic* (pp. 671–695). Oxford University Press.

Putnam, H. (1954). Synonymity, and the analysis of belief sentences. *Analysis, 14*(5), 114–122.

Quine, W. V. O. (1953). Reference and modality. In *From a logical point of view*, Second Revised Edition (pp. 139–157), 1961. Harper and Row.

Quine, W. V. O. (1956). Quantifiers and propositional attitudes. *The Journal of Philosophy, 53*(5), 177–187. (Reprinted in *The ways of paradox and other essays* (pp. 183–194), 1966. Random House).

Quine, W. V. O. (1960). *Word and object*. The MIT Press. (New edition 2013).

Quine, W. V. O. (1979). Intensions revisited. In P. A. French, T. E. Uehling, & H. C. Wettstein (Eds.), *Contemporary perspectives in the philosophy of language* (pp. 268–274). University of Minnesota Press.

Quine, W. V. O. (2000). *Response to Recanati*. In Orenstein & Kotatko (2000) (pp. 428–430).

Recanati, P. (2000). *Opacity and the attitudes*. In Orenstein & Kotatko (2000) (pp. 367–406).

Richard, M. (2001). Analysis, synonymy and sense. In A. Anderson & M. Zelëny (Eds.), *Logic, meaning and computation* (pp. 545–572). Kluwer.

Rieber, S. (1992). Understanding synonyms without knowing that they are synonymous. *Analysis, 52*(4), 224–228.

Rizzi, L. (2000). *Comparative syntax and language acquisition*. Routledge.

Salmon, N. (1986). *Frege's puzzle*. The MIT Press.

Salmon, N. (1993). A problem in the frege-church theory of sense and denotation. *Noûs, 27*(2), 158–166.

Scheffler, I. (1955). On synonymy and indirect discourse. *Philosophy of Science, 22*, 39–44.

Sellars, W. (1955). Putnam on synonymy and belief. *Analysis, 15*(2), 117–120.

Soames, S. (1987a). Substitutivity. In J. J. Thomson (Ed.), *On being and saying. Essays for Richard Cartwright* (pp. 99–132). The MIT Press.

Soames, S. (1987b). Direct reference, propositional attitudes, and semantic content. *Philosophical Topics 15*, 47–87. (Reprinted in N. Salmon & S. Soames (Eds.), *Propositions and attitudes* (pp. 197–239), 1988. Oxford University Press.)

Soames, S. (2002). *Beyond rigidity*. Oxford University Press.

Sosa, E. (1970). Propositional attitudes *de Dicto* and *De Re*. *The Journal of Philosophy, 67*, 883–896.

Thomas Aquinas (1976). De Propositionibus Modalibus. In *Opera Omnia*, Tomus XLIII. San Tommaso.

Usberti, G. (2004). On the notion of justification. *Croatian Journal of Philosophy, 4*(10), 99–122.
von Heusinger, K. (2011). Specificity. In K. von Heusinger, C. Maienborn, & P. Portner (Eds.), *Semantics* (pp. 1025–1058). de Gruyter.

Chapter 8
Knowledge and Gettier Problems

Abstract In the first part of this chapter (Sects. 8.1–8.3) a definition is proposed of the notion of $\mathcal{C}$-justification for epistemic reports, sentences of the form "$\mathcal{S}$ knows that α", and an analysis of Gettier problems. More specifically, in Sect. 8.1 the definition is given; in Sect. 8.2 a representative class of Gettier problems is introduced and analyzed, distinguishing between atomic and logically complex problems; in Sect. 8.3 a comparison is made between the present approach and J. Pollock's analysis. The second part is devoted to two strictly connected themes: assertibility conditions of empirical sentences (Sect. 8.4), defined in such a way as to result epistemically transparent; and Williamson's argument(s) against transparency (or luminosity, in his terminology) of knowledge (Sect. 8.5).

Keywords Theory of meaning · Epistemology · Knowledge · Gettier problems · Justification · Assertion · Luminosity

In the first part of this chapter (Sects. 8.1–8.3) I propose an extension of the definition of $\mathcal{C}$-justification to epistemic reports, sentences of the form "$\mathcal{S}$ knows that α", and an analysis of Gettier problems. More specifically, in Sect. 8.1 the definition is given; in Sect. 8.2 a representative class of Gettier problem is introduced and analyzed; in Sect. 8.3 a comparison is made between the present approach and J. Pollock's analysis. The second part is devoted to two strictly connected themes: assertibility conditions of empirical sentences (Sect. 8.4), defined in such a way as to result epistemically transparent, and Williamson's argument(s) against transparency (or luminosity, in his terminology) of knowledge (Sect. 8.5).

8.1 The Proposed Definition

In Chap. 2 I remarked that, in order to grant the assertibility of α, it is not strictly necessary that a justification for α is factive: what is really necessary is that it warrants *knowledge* that α, since it is knowledge that warrants assertibility; and I have called

G. Usberti, *Meaning and Justification. An Internalist Theory of Meaning*, Logic, Epistemology, and the Unity of Science 59,
https://doi.org/10.1007/978-3-031-24605-0_8

truth-grounds of α k-factive justifications for α; in other terms, I have used the intuitive notion of knowledge to isolate truth-grounds among justifications. In Chaps. 4 and 5 I have given a definition of the theoretical notion of truth-ground of α, for $\alpha \in \mathcal{L}$, according to which a truth-ground of α is nothing but a justification σ for α, provided that a certain relation exists between σ and the cognitive state of an 'observer' of σ. In Chap. 7 such an observer has been explicitly introduced in order to define justifications for doxastic reports[1]: I have called it *the reporter*. The last step is to define justifications for epistemic reports, i.e. sentences of the form "$\mathcal{S}$ knows that α".

Let me remind the reader the strategy that I sketched in Chap. 4:

(1)

- (i) The notion of truth-ground is *constitutive* of the notion of truth: a sentence we have a truth-ground of is true because we have a truth-ground of it, not viceversa. I have argued for this idea in Chap. 3.
- (ii) A subject $\mathcal{S}$ who occupies a cognitive state that is a truth-ground of α occupies an *ideal* cognitive state, i.e. a state in which $\mathcal{S}$ has, about α, all relevant information.
- (iii) The intuitive reason that justifies a belief in a cognitive state is the same as the reason that justifies that belief in an ideal cognitive state.

In Chap. 4 I focused on condition (ii), showing a tension between the notion of ideal cognitive state and the requirement of epistemic transparency of justifications, and I proposed to solve it by conceiving truth-grounds in terms of a relation between two cognitive states. In Chap. 7 the two cognitive states have been matched with the points of view of the two subjects involved in a belief report: the believer and the reporter. At this point condition (iii) requires that the intuitive reason that justifies the believer in believing that α is the same as the reason that justifies the reporter in believing the same proposition. I shall now focus on this requirement.

First of all, let us extend $\mathcal{L}_{\mathrm{B}}$ to $\mathcal{L}_{\mathrm{B+Kn}}$ (see the Preface). As a second step, let us extend the definition 5 of mind-reading cognitive state for $\mathcal{L}_{\mathrm{B}}$ given in Chap. 7 into a definition of mind-reading cognitive state *for* $\mathcal{L}_{\mathrm{B+Kn}}$ by adding the following clause:

(2) 5. in ρ the concept f_{Kn} is manageable, i.e. a feature-checking algorithm p_{Kn} is available, verifying the presence of the following feature configuration:

there are a proposition p and a cognitive state σ such that

- (i) $\mathcal{B}el_{p}$ occupies σ;
- (ii) σ is a truth-ground of p relative to ρ;
- (iii) $\mathrm{B}(\mathcal{B}el_{p}, \mathrm{p})$;
- (iv) $\mathrm{B}(\mathcal{B}el_{p}, \mathrm{p})$ is based on conditions (i) and (ii).

A remark concerning condition 5.(ii). Checking whether this condition is satisfied amounts in fact to checking the *reasons* of the believer's belief. In everyday language,

---

[1] Justifications, not truth-grounds; the rationale for this will be explained in Sect. 8.4.

when we want to explain a phenomenon, we use "reason" in typically different contexts. For example, a subject $\mathcal{S}_1$ may ask himself: "Why are there puddles in the street?", and find appropriate the answer "Because last night it has been raining"; in this case we say that last night's rain is, for $\mathcal{S}_1$, the reason of the puddles in the street. But it may also happen that, in the same situation, another subject $\mathcal{S}_2$ asks herself: "Why does $\mathcal{S}_1$ believe that last night it has been raining?", and finds appropriate the answer "Because he has seen puddles in the street"; in this case we say that the puddles in the street are, for $\mathcal{S}_2$, the reason of $\mathcal{S}_1$' belief that last night it has been raining. In the former case "reason" means the proposition a believer is authorized to believe by its being the best explanation of the data available to her/him; in the latter case it means the evidential factors of the proposition believed, on which the believer bases her/his belief. Since, in the present context, we are concerned with the cognitive state of a reporter who is judging the cognitive state of a subject who believes a proposition, the relevant sense is the latter; I will therefore use "reasons" to denote the evidential factors of the proposition believed by the believer.

Adopting this terminology we can say that when the reporter checks whether condition 5.(ii) is satisfied, what (s)he is checking is not only whether the proposition believed by the believer is evident also to him/herself, but also whether the reason why it is evident to the believer is the same as the reason why it is evident to him/herself.

The third step consists in following the same procedure adopted in Chap. 7 to define justifications for doxastic reports. (Convention: from here on I will omit the subscripts of "Bel" and "Rep" when it is not strictly necessary. See Chap. 7, fn. 69.)

Definition 1 Let ρ be an arbitrary mind-reading cognitive state for $\mathcal{L}_{\text{B+Kn}}$ of a subject $\mathcal{R}$ such that $[\![\text{Kn}]\!]_\rho = f_{\text{Kn}}$, $\mathcal{S}$ is the believer $\mathcal{Bel}$, $\mathcal{R}$ is the reporter $\mathcal{Rep}$, $[\![\mathcal{S}]\!]_\rho = \mathcal{S}$, $[\![\text{that}\,\alpha]\!]_\rho = [\![\text{that}\,_{\mathcal{Bel}}\alpha]\!]_\rho = [\![\alpha]\!]_\mathcal{S}$; then the set $\text{EF}_{\text{Kn}(\mathcal{S},\text{that}\,\alpha),\rho}$ of the *evidential factors* of Kn($\mathcal{S}$, that α) in ρ is $f_{\text{Kn}}\big(\mathcal{S}, [\![\alpha]\!]_\mathcal{S}\big) \cup \mathcal{A} = \{$<Q, B($\mathcal{S}$,that α>|Q is a why-question arising in ρ and $f_{\text{Kn}}\big(\mathcal{S}, [\![\alpha]\!]_\mathcal{S}\big) =_\rho 1$ is a potential answer to Q$\}$.

Definition 2 Let a $\in \text{EF}_{\text{Kn}(\mathcal{S},\,\text{that}\,\alpha),\rho}$; then a makes evident Kn($\mathcal{S}$, that α)) in ρ (in symbols a $\vDash_\rho$Kn($\mathcal{S}$, that α) iff

- either $a = f_{\text{Kn}}\big(\mathcal{S}, [\![\alpha]\!]_\mathcal{S}\big)$ and $f_{\text{Kn}}\big(\mathcal{S}, [\![\alpha]\!]_\mathcal{S}\big) = {}_\rho 1$;
- or a $\in \mathcal{A}$ and $f_{\text{Kn}}\big(\mathcal{S}, [\![\alpha]\!]_\mathcal{S}\big) = {}_\rho 1$ is the best answer to Q in ρ.

Definition 3 j_ρ(Kn($\mathcal{S}$, that α)) = 1 iff there is an a $\in \text{EF}_{\text{Kn}(\mathcal{S}\,\text{that}\,\alpha),\rho}$ such that a$\vDash_\rho$Kn($\mathcal{S}$, that α).

Definition 4 An *$\mathcal{R}$-justification for the sentence* Kn($\mathcal{S}$, that α) is a mind-reading cognitive state ρ of the subject such that j_ρ(Kn($\mathcal{S}$, that α)) = 1.

Remark 1 In consequence of Definitions 1–4, of Definition 5 of Chap. 7 and of (2), ρ is a justification for Kn($\mathcal{S}$, that α) iff ρ is a justification for B($\mathcal{S}$,that$_{\mathcal{Bel}}$α) and σ, the cognitive state occupied by $\mathcal{S}$, is a truth-ground of α relative to ρ.

My approach predicts therefore that epistemic reports are not subject to the TO ambiguity, as doxastic reports are. This seems to be confirmed by empirical data; when we ask whether $\mathcal{S}$ knows that α, one of the tests we apply is (DP), which means that we take "$\mathcal{S}$ knows that α" as entailing "$\mathcal{S}$ (justifiedly) believes that α" *in its O-reading*[2] which, according to my analysis, concerns the proposition expressed by α *for the believer*. If we knew that $\mathcal{S}$ does not believe that Carl Hempel is Peter Hempel, and we wanted to test whether $\mathcal{S}$ knows that Carl Hempel was a philosopher, it would be senseless that we ask her/him: "Do you believe that Peter Hempel was a philosopher?".

8.2 Gettier Problems

It is well known that Gettier stated only two counterexamples to the traditional definition of knowledge, but that it is possible to generate a potentially infinite family.[3] I will call "Gettier problems" all the cases in which we would intuitively say that a subject justifiedly believes the true proposition expressed by the statement α, but not that the subject knows that α; "atomic (Gettier) problems" the cases in which α is a statement that can be translated into an atomic sentence of $\mathcal{L}$, "conjunctive" the ones in which α is translated into a conjunction; and so on. In this terminology, Gettier's first original problem is an atomic problem, while the second is a disjunctive problem. Here is a small catalogue of problems, which will be discussed below and are representative of a vast class of cases.

8.2.1 Atomic Problems

Gettier 1. (Gettier, 1963)

Smith and Jones have applied for a certain job. Smith is justified in believing the following proposition:

(3) Jones is the man who will get the job, and Jones has ten coins in his pocket,

because the president of the company assured him that Jones would in the end be selected, and because he, Smith, had counted the coins in Jones's pocket ten minutes ago. From (3) Smith correctly infers

(4) The man who will get the job has ten coins in his pocket,

and accepts (4) on the grounds of (3). As a matter of fact, unknown to Smith, he himself, not Jones, will get the job. And, also, unknown to Smith, he himself has ten

[2] Cp. Chap. 7, fn. 76.

[3] Cp. Zagzebski (1994); a classical survey of Gettier problems is Shope (1983).

coins in his pocket. Smith is therefore justified in believing (4), and (4) is true, but Smith does not know that (4) is true.

The Fake Barns. (Goldman, 1976)

Tim believes that

(5) That is a barn

because he is seeing from the highway a building looking like a barn in good light and from not too great a distance; as a matter of fact, that building is a barn. But Tim is unaware that people there who want to appear affluent have erected many barn facades indistinguishable from real barns when seen from the highway, so that there are many more barn facades than real barns.

Tom Grabit. (Lehrer & Paxson, 1969)

Keith sees Tom Grabit steal a book from the library. But unsuspected by Keith, Tom's mother has said that Tom was miles away at the time of the theft and has a twin brother, John, almost indistinguishable from Tom, who was in the library at the time. Yet Tom's mother is a compulsive and pathological liar, and her statement is a neurotic lie. Intuitively, Keith does know that

(6) Tom Grabit stole a book.

The civil-rights leader assassination. (Harman, 1968)

Tom believes thats

(7) A famous civil-rights leader has been assassinated

because he has read it in a reliable newspaper. The news is true but, unsuspected by Tom, the assassination has been denied, even by eyewitnesses, the point of the denial being to avoid a racial explosion; the denials occurred too late to prevent the original and true story from appearing in the paper; but everyone else in the town has heard about the denials, and does not know what to believe. «Would we judge—Harman asks—Tom to be the only one who knows that the assassination has actually occurred? […] I do not think so.»

Norman & Mary 1. (Harman, 1980)

Mary calls Norman's office and is told that

(8) Norman is in Italy;

moreover, Norman is in Italy. However, when Mary calls Norman's office, there is, in the pile of unopened mail on Mary's desk before her, a letter written by Norman, in which he says (falsely) he is in San Francisco (and the letter has been been mailed to Mary by a friend of Norman's from San Francisco). Intuitively, Mary does not know that (8) is true.

Norman & Mary 2. (Pollock, 1986)

Like **Norman & Mary 1**, with the difference that Norman has not had his trick letter mailed from San Francisco, but has had a friend secrete it under Mary's doormat. Intuitively, Mary does know that (8) is true.

Norman & Mary 3

Like **Norman & Mary 1**, with the difference that Norman has not written a letter addressed to Mary, but has written (falsely) that he is in San Francisco in a private diary he keeps regularly, which Mary knows and often reads, believing (falsely) that Norman does not know she knows it. Intuitively, Mary does not know that (8) is true.

8.2.2 Logically Complex Problems

Gettier 2. (Gettier, 1963)

Smith has strong evidence for the following proposition:

(9) John owns a Ford;

obviously, (9) entails

(10) Either John owns a Ford, or Brown is in Barcelona;

Smith realizes the entailment and is therefore justified in believing (10) on the basis of (9). But imagine that two further conditions hold: first John does not own a Ford; and secondly, by the sheerest coincidence, and entirely unknown to Smith, Brown is in Barcelona. If these two conditions hold, then Smith does not know that (10) is true, even though (10) is true, Smith does believe that (10) is true, and Smith is justified in believing that (10) is true.

A variant of Gettier 1

Smith and Jones have applied for a certain job. Smith is justified in believing (4) because the president of the company, having examined both and having found them equally worthy under all respects, decreed that the job will be won by the man who has ten coins in his pocket (and the position will stay vacant if nobody has or both have ten coins in his/their pocket). As a matter of fact, unknown to Smith, he himself has ten coins in his pocket. In this case Smith does intuitively know that (4) is true.

This problem is to be classified as logically complex only if, differently from the rest of this book, we treat descriptions *à-la* Russell. I introduce it here because of the intrinsic interest of its analysis, as we will see in a moment.

8.2.3 Analysis of Atomic Problems

Let us consider atomic Gettier problems described above and see how Definitions of Sect. 8.1 account for the intuitive data.

Gettier 1

Let σ be Smith's cognitive state, ρ the mind-reading cognitive state of the reporter (i.e. Gettier himself), "$\mathcal{S}$" and "α" the translations into $\mathcal{L}_{\mathrm{B+Kn}}$ of "Smith" and (4), respectively. We must check whether ρ is a $\mathcal{C}$-justification for "Smith knows that α". In view of Remark 1 it is sufficient to check whether σ is a justification for $\mathrm{B}(\mathcal{S}, \mathrm{that}_{\mathcal{S}}\alpha)$ and σ is a truth-ground of α relative to ρ. Gettier's description of the case shows that Smith's believing the proposition expressed for him by α is based on Smith's occupying σ and on σ being a justification for α; hence, by Definitions 5–9 of Chap. 7, σ is a justification for $\mathrm{B}(\mathcal{S}, \mathrm{that}_{\mathcal{S}}\alpha)$. However, σ is not a truth-ground of α relative to ρ, because the evidential factor that makes evident α in σ is not the same as the evidential factor that makes it evident in ρ: in σ a why-question arises ("Why did the president say that Jones would in the end be selected in contrast to not saying it?") the best answer to which ("Because the president knew that Jones would in the end be selected") is different from the best answer in ρ (maybe "Because the president was insincere or ill-informed").

The Fake Barns

Let "$\mathcal{S}$" and "α" be the translations into $\mathcal{L}_{\mathrm{B+Kn}}$ of "Tim" and (5), respectively; furthermore, let σ be Tim's cognitive state, σ' the cognitive state (of an unspecified subject) resulting from adding to σ the piece of information that people there have erected many barn facades indistinguishable from real barns when seen from the highway; and ρ the mind-reading cognitive state of the reporter (Goldman), resulting from adding to σ' the piece of information that that specific building is a real barn. Goldman's description of the case shows that Tim's believing the proposition expressed for him by α is based on Tim's occupying σ and on σ being a justification for α; hence, by Definitions 5–9 of Chap. 7, $j_{\sigma}\big(\mathrm{B}\big(\mathcal{S},\mathrm{that}_{\mathcal{S}}\alpha\big)\big) = 1$, i.e. σ is a justification for $\mathrm{B}\big(\mathcal{S},\mathrm{that}_{\mathcal{S}}\alpha\big)$. However, σ is not a truth-ground of α relative to ρ, because the evidential factor that makes evident α in σ is not the same as the evidential factor that makes it evident in ρ: in σ a why-question Q arises ("Why does that look like a barn in contrast to looking like something other?") the best answer to which is R: "Because it is a barn"; in ρ question Q arises, but the best answer to it is not R (because of the fact that the piece of information characterizing σ' is available), but something like R′: "Because it's facade is a good imitation"; on the other hand, there is a why-question arising in ρ, and having R as best answer in ρ: it is not Q, but something like Q′: "Why does that look like a barn also when seen from near in contrast to looking like a barn only when seen from the highway?"). The pair <Q′, R> is different both from the pair <Q, R> and from the pair <Q, R′> .

Tom Grabit

Let "$\mathcal{S}$" and "α" be the translations into $\mathcal{L}_{\mathrm{B+Kn}}$ of "Keith" and (6), respectively; furthermore, let σ be Keith's cognitive state, σ' the cognitive state of a subject informed of Tom's mother's statement, and ρ the mind-reading cognitive state of the reporter, resulting from adding to σ' the piece of information that Tom's mother's statement is a neurotic lie. Keith's believing the proposition expressed for him by α

is based on Keith's occupying σ and on σ being a justification for α; hence, by Definitions 5–9 of Chap. 7, $j_\sigma(\mathrm{B}(\mathcal{S},\mathrm{that}_{\mathcal{S}}\alpha)) = 1$, i.e. σ is a justification for $\mathrm{B}(\mathcal{S},\mathrm{that}_{\mathcal{S}}\alpha)$. Moreover, in the three cognitive states the same why-question Q arises ("Why does that guy look like Tom Grabit in contrast to looking like someone other?"), and the best answer to it is the same in σ and in ρ ("Because that guy is Tom Grabit"), while in σ′ it is different ("Because that guy is Tom's twin brother John Grabit"). Therefore σ is not a truth-ground of α relatively to σ′, but it is relatively to ρ; in conclusion, $j_\rho(\mathrm{Kn}(\mathcal{S},\ \mathrm{that}\ \alpha)) = 1$.

The civil-rights leader assassination

Let "$\mathcal{S}$" and "α" be the translations into $\mathcal{L}_{\mathrm{B+Kn}}$ of "Tom" and (7), respectively; furthermore, let σ be Tom's cognitive state, σ′ the cognitive state of the people who heard that the assassination has been denied even by eyewitnesses, and ρ the cognitive state of the reporter (Harman), resulting from adding to σ′ the piece of information that the news in the newspaper is true and that the point of the denial was to avoid a racial explosion. Tom's believing the proposition expressed for him by α is based on Tom's occupying σ and on σ being a justification for α; hence σ is a justification for $\mathrm{B}(\mathcal{S},\ \mathrm{that}_{\mathcal{S}}\alpha)$. However, in σ the why-question Q arises: "Why did the newspaper say, in contrast to conjecturing, that α?", and the best answer is R: "Because the newspaper knew that α"; in ρ a different why-question Q′ arises[4]: "Why did the newspaper say that α, in contrast to what the eyewitnesses said?", and the best answer is R′: "Because the newspaper knew that α and the eyewitnesses wanted to avoid a racial explosion"; since $<\mathrm{Q}, \mathrm{R}> \neq <\mathrm{Q}', \mathrm{R}'>$, the evidential factor that makes α evident is not the same in σ and in ρ, hence σ is not a truth-ground of α relative to ρ.

Norman & Mary 1

Let "$\mathcal{S}$" and "α" be the translations into $\mathcal{L}_{\mathrm{B+Kn}}$ of "Mary" and (8), respectively; furthermore, let σ be Mary's cognitive state, σ′ the cognitive state resulting from adding to σ the piece of information that in the pile of unopened mail on Mary's desk there is a letter from San Francisco in which Norman says that he is in San Francisco; and ρ the mind-reading cognitive state of the reporter (Harman), resulting from adding to σ′ the piece of information that the content of Norman's letter is false. Mary's believing the proposition expressed for her by α is based on Mary's occupying σ and on σ being a justification for α; hence σ is a justification for $\mathrm{B}(\mathcal{S},\ \mathrm{that}_{\mathcal{S}}\alpha)$. However, in σ the why-question Q arises: "Why did Norman's office say, in contrast to conjecturing, that α?", and the best answer is R: "Because the office knows that α"; in ρ a different why-question Q′ arises: "Why did Norman's office say that α, in contrast to what Norman said in his letter?", and the best answer is R′: "Because Norman's office knows that α and Norman wanted to deceive Mary". Since $<\mathrm{Q}, \mathrm{R}> \neq <\mathrm{Q}', \mathrm{R}'>$, the evidential factor that makes α evident is not the same in σ and in ρ, hence σ is not a truth-ground of α relative to ρ. It is important to stress the role played by the cognitive state σ′ in this case: it modifies the contrast of the why-question Q,

[4] Notice that Q too arises in ρ, according to van Fraassen's definition of arising; but the best answer to it is not R, but something like "Because only the newspaper was sincere".

producing the different why-question Q′; and the reason of this modification is that the reporter knows that it is relevant to Mary's belief, in the sense that Mary's belief would have been modified, had she occupied σ′.

Norman & Mary 2

Let "$\mathcal{S}$", "α" and σ be as in the preceding case; furthermore, let σ′ the cognitive state resulting from adding to σ the piece of information that Norman (has not had his trick letter mailed from San Francisco, but) has had a friend secrete the letter under Mary's doormat; and ρ the mind-reading cognitive state of the reporter (Harman), resulting from adding to σ′ the piece of information that the content of Norman's letter is false. As before, Mary's believing the proposition expressed for her by α is based on Mary's occupying σ and on σ being a justification for α; hence σ is a justification for $B\left(\mathcal{S}, \text{that}_{\mathcal{S}}\alpha\right)$. Moreover, in ρ the why-question Q′ of the preceding case cannot arise, because Norman has not told Mary anything contrasting with what the office said (he has written a letter containing the proposition that he is in San Francisco, but he has not sent it to Mary: hence he has not said to Mary, through the letter, that he is in San Francisco). Therefore, the why-question arising in ρ is the very same Q arising in σ,[5] and the answer is the same as in σ, so the evidential factor that makes α evident in σ and in ρ is the same, hence σ is a truth-ground of α relative to ρ. In this case the cognitive state σ′ does not modify the contrast of the why-question Q, since the reporter knows that it is not relevant to Mary's belief, in the sense specified above.

Norman & Mary 3

Let "$\mathcal{S}$", "α" and σ be as in the preceding case; furthermore, let σ′ be the cognitive state resulting from adding to σ the piece of information that Norman has written that he is in San Francisco in his private diary, and ρ the cognitive state of the reporter, resulting from adding to σ′ the piece of information that the content of Norman's diary at the relevant date is false. In σ the why-question Q arises: "Why did Norman's office say, in contrast to conjecturing, that α?", and the best answer is R: "Because the office knows that α"; in ρ a different why-question Q′ arises: "Why did Norman's office say that α, in contrast to what Norman wrote in his diary?", and the best answer is R′: "Because Norman's office knows that α and Norman wanted to deceive Mary". Therefore the evidential factor that makes α evident is not the same in σ and in ρ, hence σ is not a truth-ground of α relative to ρ. In this case, like in **Norman & Mary 1**, the cognitive state σ′ does modify the contrast of the why-question Q, since the reporter knows that it is relevant to Mary's belief, in the sense specified above.

---

[5] Of course in ρ other questions arise, such as "Why has Norman written the letter in contrast with not writing it?" and "Why has Norman had a friend secrete his trick letter under Mary's doormath in contrast to having it mailed?"; but this is not relevant to our problem.

8.2.4 Analysis of Complex Problems

The definition of the notion of $\mathcal{C}$-truth-ground of α, for α logically complex, has been given in Chap. 5 (Definition 8); I shall use it to account for complex Gettier problems.

Gettier 2

Let "$\mathcal{S}$", "β" and "γ" be the translations into $\mathcal{L}_{B+Kn}$ of "Smith", (9) and "Brown is in Barcelona", respectively; furthermore, let σ be Smith's cognitive state, and ρ the mind-reading cognitive state of the reporter (Gettier), resulting from adding to σ the pieces of information that ¬β and that γ. Gettier's description of the case shows that Smith's believing the proposition expressed for him by β∨γ is based on Smith's occupying σ, on σ being a justification for β, and on Smith's inferential competence, in virtue of which Smith sees that β∨γ may be inferred from β. According to Remark 1, ρ is a justification for Kn($\mathcal{S}$, that β∨γ) iff ρ is a justification for B($\mathcal{S}$, that$_{\mathcal{Bel}}$ β ∨ γ) and σ, the cognitive state of the believer, is a truth-ground of β∨γ relative to ρ. According to Definition 8 of Chap. 5, σ is a truth-ground of β∨γ relative to ρ iff it is either a truth-ground of β relative to ρ or a truth-ground of γ relative to ρ; since both β and γ are atomic sentences, we must look at Definition 13 of chap. 4; according to it σ is not a truth-ground of γ relative to ρ, because $j_\sigma(\gamma)\neq 1$, and σ is not a truth-ground of β relative to ρ, because $j_\rho(\beta)\neq 1$. In conclusion, ρ is not a justification for Kn($\mathcal{S}$,that β∨γ).

A variant of Gettier 1

A clear intuition about (4) is that it is ambiguous between an epistemically specific and a non-specific interpretation; Indeed, (4) seems to be a perfect analogue of sentence (86) of Chap. 7:

(86) A customer of the bar is a counterfeiter,

with the only difference that in the latter an indefinite description occurs, in the former a definite one. In both cases, the specific and the non-specific interpretation are intuitively 'triggered' by the context, as it is chracterized by the description of the case; according to my approach, the triggering is provoked by the direct and indirect nature, respectively, of the corresponding cognitive state, provided that descriptions are treated *à-la* Russell: in **Gettier 1** Smith believes that (4) is true because he believes that *Jones* is the man who will get the job, while in the variant the rule set by the president can be verbally reconstructed as an essentially indirect reasoning, analogous (if more complicate) to the one posited in Chap. 7 for (96):

$$[\forall x\neg C(x)\vee\neg\forall x\forall y(((C(x)\wedge C(y))\rightarrow x=y)]^{1}$$

$$\vdots$$

$$(11)\ \dfrac{\dfrac{\overline{\quad V \quad} \qquad V\rightarrow\bot^{2}}{\bot}}{\neg(\forall x\neg C(x)\vee\exists x\exists y((C(x)\wedge C(y))\wedge\neg(x=y))).}{,1}$$

$$\vdots^{6}$$

$$\overline{\neg\forall x\neg C(x)\wedge\neg\exists x\exists y((C(x)\wedge C(y))\wedge\neg(x=y))).}$$

$$\vdots^{7}$$

$$\dfrac{\overline{\neg\forall x\neg(C(x))\wedge\forall x\forall y((C(x)\wedge C(y))\rightarrow\neg\neg(x=y)).}}{\neg\forall x\neg(C(x))\wedge\forall x\forall y((C(x)\wedge C(y))\rightarrow(x=y)).^{8}}$$

where "C(x)" and "V" abbreviate "x has ten coins in his pocket" and "The position will stay vacant", respectively.

If σ is Smith's cognitive state, ρ the mind-reading cognitive state of the reporter, and "α" the translation of (4), we realize that the evidential factor that makes evident α in σ is the same as the one that makes it evident in ρ: exactly one applicant's having ten coins in his pocket; hence σ is a $\mathcal{C}$-truth-ground of α relatively to ρ: $j_{\rho}(\mathrm{Kn}(\mathcal{S},\mathrm{that}\alpha)) = 1$. Notice that the identity of the evidential factor in σ and in ρ depends precisely on the fact that, in both cognitive states, α does not speak of a specific person, but of a set containing exactly one element.

8.2.5 Conclusion

I conclude with two remarks. First, I remind the reader that I observed in Chap. 7 (Sect. 7.4.1) that cases like **Gettier 1** and its variant offer good reasons to treat the ESN distinction as a semantical rather than as a pragmatical distinction. I should now add that they also offer a further reason, besides the ones given in Sect. 7.2.1, to consider the TO distinction and the ESN distinction as different distinctions.

---

[6] Notice that $(\neg(\alpha\vee\beta)\rightarrow(\neg\alpha\wedge\neg\beta))$ is intuitionistically valid.

[7] Notice that $\neg\exists x\exists y(\alpha(x,y)\wedge\neg(x=y))\rightarrow\forall x\forall y(\alpha(x,y)\rightarrow\neg\neg(x=y))$ is intuitionistically valid.

[8] The last step is intuitionistically legitimate whenever identity is decidable.

The second remark is more general. Let us start from the observation that Gettier's problems can be seen as a *sophisticated* version of the sceptical objection to the possibility of knowledge: not only do they, like all sceptical objections, point to the possibility that situations that appear to be cases of knowledge are in fact only cases of justified belief, but among these situations of only apparent knowledge they include new ones, apparently even more like cases of knowledge: cases of belief that is not only justified but also *true*. Seen in this way, Gettier's problems strengthen the sceptical position. Let us now observe that the essential ingredient of my proposal for solving Gettier's problems consists in introducing the author of each problem as a reporter of the relevant attribution of belief. While on the one hand this strategy is a consequence of my proposal to represent the TO ambiguity by means of the Reporter/Believer Ambiguity (see Chap. 7, Sect. 7.2.8), on the other hand it has the merit of highlighting a neglected aspect of all Gettier problems: that the problem author, unless (s)he is God, can herself/himself be a victim of a Gettier situation.[9] This possibility is usually hidden by the fact that, in describing the case, the author refers to the way things are 'in reality', implicitly suggesting (but in no way arguing)) that her/his description of the facts is the definitive one, so to speak. On the contrary, if the author of the problem becomes one of the actors in the situation (s)he is describing, the crucial point comes to the fore: strictly speaking, the author of a Gettier problem does not tell us how things are 'in reality', but how they are on the basis of better information than that available to the believer; the author, as a reporter and thus herself/himself a knowing subject, is not in a position to rule out the possibility that things are other than how (s)he says they are.

As we have seen, this observation opens the way to a solution of the problems, as it highlights that the real terms of the problem are the information available to the subject and that available to the reporter, not the information available to the subject and 'reality'. But what I want to emphasise here is that the sophisticated sceptical objection implicit in Gettier's problems is based on a realist assumption: that it makes sense to refer to how things are 'in reality', in a reality that may be inaccessible. Conversely, this assumption of a reality that may be inaccessible seems to be the very origin of the sceptical attitude. If this impression is correct, then the simple anti-realist move of saying that if things stand in a way that is inaccessible to any knowing subject, then they stand in no way - i.e. that a transcendent truth simply does not exist - permits to overcome the sceptical objection.

8.3 Excursus. A Comparison with Pollock's Solution

The analysis of atomic Gettier problems I have proposed has much in common with John Pollock's solution (Pollock, 1986), but there are also significant differences; a detailed comparison may be interesting. In order to state Pollock's definition of knowledge some preliminary definitions are necessary.

---

[9] I thank an anonymous referee for this observation.

Definition 5 A belief β is a reason$_{P}$ for a person $\mathcal{S}$ to believe α iff it is logically possible for $\mathcal{S}$ to become justified in believing α by believing it on the basis of β.[10]

Definition 6 If β is a reason$_{P}$ for $\mathcal{S}$ to believe α, δ is a defeater for β iff δ is logically consistent with β and (β∧δ) is not a reason$_{P}$ for $\mathcal{S}$ to believe α.

A reason$_{P}$ β is defeasible iff there can be defeaters for β.

The following definition amounts to a drastically simplified, but I hope faithful, version of Pollock's analysis of knowledge:

Definition 7 $\mathcal{S}$ knows that α if and only if

(i) $\mathcal{S}$ is (subjectively) justified in believing α;
(ii) necessarily, if $\mathcal{S}$ believed all the truths, and believed α for the same reason$_{P}$, then $\mathcal{S}$ would still be (subjectively) justified in believing α[11];
(iii) necessarily, if $\mathcal{S}$ believed all the truths socially sensitive for $\mathcal{S}$[12] and believed α for the same reason$_{P}$, then $\mathcal{S}$ would still be (subjectively) justified in believing α.

Let us see how this definition can be applied to the solution of atomic Gettier problems.

Gettier 1. According to Definition 7, Smith does not know that (4) is true, since condition (ii) is violated: there is a truth (i.e., that Smith himself will get the job and has ten coins in his pocket) such that, if Smith believed it, then his original reason$_{P}$ for believing (4) would no longer justify him in believing (4).

The Fake Barns. According to condition (ii) of Definition 7, as for intuition, Tim's true and justified belief is not knowledge: again, there is a truth (i.e., that there are many more barn facades over there than real barns) such that, if Tim believed it, then

---

[10] Notice:

(1) α and β are entities of a different nature: α is a proposition («Objects of belief are called "propositions"»; Pollock, 1986: 29), β is a mental state of belief; Pollock seems not to be interested at the distinction: «I believe that at considerable stylistic expense, all reference to propositions [...] could be replaced by talk of "possible beliefs".» (Pollock, 1986: 30); presumably for this reason he refers to them as if they were both propositions (for instance, he applies the logical constants to them). Later in the book, however, Pollock modifies his definitions of reason$_{P}$ and defeater in order to respect the distinction between believed propositions and mental states (Pollock, 1986: 176); but the reason of that modification is that he wants to make a distinction between two kinds of mental states, i.e. beliefs and reasons: reasons are mental states, but not necessarily belief states.

(2) The basing relation, i.e. the relation existing between a belief α and a belief β iff α is based on β, can be taken here as primitive.

[11] Pollock'S formulation is the following (Pollock, 1986: 185):

(ii') There is a set X of truths such that, for every set of truths Y such that X⊆Y, necessarily, if $\mathcal{S}$ believed the truths in Y and believed α for the same reason, then $\mathcal{S}$ would still be (subjectively) justified in believing α.

The two formulations are equivalent.

[12] «[A] proposition is *socially sensitive for* $\mathcal{S}$ if and only if it is of a sort $\mathcal{S}$ is expected to believe when true» (Pollock, 1986: 192).

his original reason$_P$ for believing (5) would no longer justify him in believing (5): even if he knew that it is really a barn, his original reason$_P$ for believing it (his seeing a building looking like a barn) would no longer justify his belief.

Tom Grabit. Abbreviations:

- p_1 is the true piece of information that Keith has seen Tom Grabit steal a book;
- p_2 is the true piece of information that Tom's mother has said that it was not Tom but John who was in the library at the time;
- p_3 is the true piece of information that Tom's mother's statement is a lie.

Let us apply Definition 7: if Keith believed p_1–p_3, and believed (6) on the basis of p_1, then Keith would still be (subjectively) justified in believing (6); therefore Keith knows (6), in accord with our intuition.

The role of condition (iii) of Definition 7 can be evinced by considering

The civil-rights leader assassination. Abbreviations:

- p_1 is the true news (7), read by Tom it in a reliable newspaper;
- p_2 is the true piece of information that the assassination has been denied by eyewitnesses;
- p_3 is the true piece of information that the point of the denial was to avoid a racial explosion.

Notice that condition (ii) of Definition 7 is met: if Tom believed all the truths, in particular p_1-p_3, and believed (7) for the same reason$_P$, then Tom would still be justified in believing (7); so the condition (ii) alone does not make the right prediction, since, intuitively, Tom lacks knowledge. However, condition (iii) is not met: if Tom believed all the truths *socially sensitive for him*, i.e. p_1-p_2, and believed (7) for the same reason$_P$, then he would not be justified in believing (7).

Norman & Mary 1. Abbreviations:

- p_1 is the true piece of information that Norman's office said (8);
- p_2 is the true piece of information that there is a letter in the mail in which Norman writes that he is in San Francisco;
- p_3 is the true piece of information that Norman's letter was written with the intention to deceive.

The case is similar to the preceding one: condition (ii) is met: if Mary believed all the truths, in particular p_1-p_3, and believed (8) for the same reason$_P$, then Mary would still be justified in believing (8); but condition (iii) is not met: if Mary believed all the truths socially sensitive for her, i.e. p_1–p_2, and believed (8) for the same reason$_P$, then she would not be justified in believing (8). As Pollock remarks, p_2 is socially sensitive because «We are 'socially expected' to be aware of various things. We are expected to know what is announced on television, and we are expected to know what is in our mail» (Pollock, 1986: 192).

Norman & Mary 2. Abbreviations:

- p_1 and p_3 are like in the analysis of **Norman & Mary 1**;
- p_2 is the true piece of information that there is a letter under Mary's doormat in which Norman writes that he is in San Francisco.

Condition (ii) is met for the same reason as in **Norman & Mary 1**, while condition (iii) is met because p_2 is not socially sensitive.

Norman & Mary 3 will be discussed below.

I shall now try to highlight some analogies and differences between Pollock's approach and mine.

An important analogy is that in both approaches the definition of knowing that α does not contain the condition that α is true; as a consequence they are not exposed to the difficulty of the Platonic definition of knowledge pointed out in Chap. 3. On the other hand, two basic differences must be stressed at once. First, the strategy Pollock adopts to define knowledge essentially consists in imposing constraints on justified belief; the basis of his approach is therefore the notion of justifiedness of a belief, i.e. of doxastic or *ex post* justification. My approach is based on the notion of propositional or *ex ante* justification, and the strategy consists in imposing constraints on propositional justification in order to arrive at a definition of truth-grounds. Second, Pollock conceives reasons as mental states, and I suggest to conceive justifications (not reasons: reasons in my sense, i.e. evidential factors, are something very different from reasons_P) as cognitive states; there is an analogy, but it is superficial, and it conceals a more significant difference: reasons_P are *tokens* of mental states (since they have causal effects),[13] while justifications, and more generally cognitive stress as I conceive them, are *types* of mental states, which affect other mental states only through the existence of programs.

Let us consider the fine structure of the two approaches. The general idea underlying Pollock's approach is that a subject $\mathcal{S}$ knows that α when $\mathcal{S}$'s reason to believe that α is not defeated under full information; this idea yields the key to the solution of **Gettier 1**, **The Fake Barns**, and **Tom Grabit**. But it cannot be applied to **The civil-rights leader assassination** and **Norman & Mary 1**: here the situation is exactly analogous to that of **Tom Grabit**, in the sense that p_2 is a true defeater of p_1, and p_3 is a true defeater of p_2, so the application of the general idea would predict that the believer has knowledge, contrary to our intuition. To handle these cases Pollock introduces condition (iii), which is (vacuously) met in **Tom Grabit** and not met in **The civil-rights leader assassination** and **Norman & Mary 1**, since in both the latter cases there is a truth socially sensitive for the believer (i.e. p_2) such that, if the believer believed it and believed (7) (resp., (8)) for the same reason_P (i.e. p_1), (s)he would not be justified in her/his belief. The fact that, in **Norman & Mary 2**, condition (iii) is met, so that the believer has knowledge, in accordance with intuition, indicates, according to Pollock, «that it is social sensitivity and not mere

[13] «[The basing relation between beliefs] is in some loose sense a causal relation» (Pollock, 1986: 35).

ready availability that enables a truth to defeat a knowledge claim» (Pollock, 1986: 192–193).

My objection to Pollock's condition (iii) is, first, that it introduces an intractable factor of vagueness into the analysis: how to give a general and sufficiently precise characterization of the class of propositions a subject is expected to believe? Second, the case of **Norman & Mary 3** seems to indicate that social sensitivity (let alone readily availability) is not sufficient a condition to circumscribe the class of propositions whose truth can defeat a knowledge claim.

Let us introduce the following abbreviations:

- p_1 is like in the analysis of **Norman & Mary 1**;
- p_2 is the true piece of information that Norman has written that he is in San Francisco in his private diary;
- p_3 is the true piece of information that Norman's annotation in his diary was written with the intention to deceive.

The situation is analogous to **The civil-rights leader assassination** and **Norman & Mary 1**, but nothing prevents condition (iii) from being met, since p_2 is not socially sensitive for the believer: Mary is not expected to read her husband's private diary.

I have adopted a different strategy to approach this sort of cases; it can be described as a modification of Definition 7 consisting in suppressing condition (iii) and replacing condition (ii) with the following:

(ii') necessarily, if a subject $\mathcal{S}$' believed all the truths and were justified in believing α, $\mathcal{S}$' would believe α for the same reason as $\mathcal{S}$.

It is not difficult to verify that condition (ii') makes predictions in accord with our intuitions in all the Gettier cases examined above. This new condition requires a finer analysis of the reasons of belief and of their identity conditions, which is just the one proposed in my approach. Moreover, it requires a clear distinction between the reasons of a belief and the justification for its content (whereas Pollock sometimes seems to assume that reasons$_P$ themselves justify belief): in my approach, given a cognitive state σ, the reason of a belief is the the evidential factor of the proposition believed that makes it evident,[14] while its justification must be equated to the whole cognitive state σ. For example, in **Gettier 1** the reason of Smith's belief that (4) is the pair <Q, R>, where Q = "Why did the president say that Jones would in the end be selected in contrast to not saying it?" and R = "Because the president knew that Jones would in the end be selected", while the justification for (4) Smith has is Smith's global cognitive state σ; it could not be identified with any one of the single pieces of information available to Smith, nor with their conjunction: for many other things are necessary in order that those pieces of information justify Smith's belief—for instance, that Smith knows that the president is trustworthy and well-informed, that Smith does not know that he himself has ten coins in his pocket, and so on.

Condition (ii') is intended to account for one essential difference between the two approaches. A second difference depends on the fact that knowledge, when defined

[14] See (2) and Definitions 1–4.

according to Pollock's Definition 7, is not epistemically transparent. The reason is that condition (ii) of Definition 7 contains a counterfactual whose antecedent makes reference to an ideal cognitive state, and the property of a cognitive state of being ideal is not epistemically transparent, as I have argued in Chap. 4, Sect. 4.4.2. The problem is therefore whether and how this reference to an ideal cognitive state can be avoided.

The starting point is the remark that our goal is to define not the realistic truth-conditions of epistemic reports, but the notion of $\mathcal{C}$-justification for them, where a $\mathcal{C}$-justification for "$\mathcal{S}$ knows that α" is in general a cognitive state of the reporter's. As a consequence, every reference to truth in condition (ii) can be replaced with a reference to the cognitive state of the reporter. This permits to eliminate the counterfactual in (ii): instead of making reference to an ideal cognitive state, we can take into account the relation "σ′ assigns to σ the status of truth-ground of α", which is epistemically transparent. The result of these transformations is essentially Definition 8 of $\mathcal{C}$-truth-ground of an atomic statement given in Chap. 4.

8.4 Assertion

In Chap. 5 I have proposed a definition of $\mathcal{S}$-validity for sentences of $\mathcal{L}$ according to which α is $\mathcal{S}$-valid iff there is a cognitive state σ of $\mathcal{S}$ such that it is a truth-ground of α relatively to all cognitive states of $\mathcal{S}$that are better than σ. A natural question arising at this point is whether $\mathcal{S}$-validity can be conceived as the formal counterpart of the intuitive notion of assertibility of α by the subject $\mathcal{S}$.

An objection to this equation is that in the definition of $\mathcal{S}$-validity a universal quantification over cognitive states of $\mathcal{S}$ occurs which cannot be understood intuitionistically, because when α is an atomic empirical sentence its evidential factors may vary in a non-monotonic way[15]; as a consequence, the $\mathcal{S}$-validity of α is in general not epistemically transparent. On the other hand, as I have argued in Chap. 2, Sect. 2.4.3.1, to be intuitively warranted *in believing* that α one must not only have evidence for α, but also base one's belief on that evidence; and to base one's belief on that evidence, one must *know* that it is evidence for α; and the same holds for acts like assertion and inference. In other terms, if we want to conceive the practice of assertion as determined by computations, it is necessary that the assertibility of a sentence α is transparent. The question I address in this section is how to conceive $\mathcal{S}$-assertibility of empirical sentences in such a way that it is epistemically transparent.

The most natural proposal is to define $\mathcal{S}$-assertibility as follows:

Definition 8 A sentence α is *assertible* by a subject $\mathcal{S}$ in the cognitive state σ iff $\mathcal{S}$ believes that α and $\mathcal{S}$'s believing that α is based on σ's being a truth-ground of α relative to σ.

[15] See Chap. 5, fn. 19.

A possible objection to this definition is based on the remark that every justification σ for α is a truth-ground of α relative to σ, since the betterness relation between cognitive states is obviously reflexive; so why does the definition require that σ is a truth-ground of α relative to σ and not simply a justification for α? The answer I suggest has to do with an essential characteristic of the notion of truth-ground. As I observed in Chap. 4, Sect. 4.4.2, the relation "σ′ assigns to σ the status of $\mathcal{S}$-truth-ground of α" concerns two cognitive states, one belonging to the (virtual) subject of belief, the other to the (virtual) reporter of the belief. Clearly, σ and σ′ may be the same cognitive state, and, when this turns out to be the case, a $\mathcal{C}$-justification for α is automatically a $\mathcal{C}$-truth-ground of α. But it is equally clear that σ and σ′ may be different; moreover, it seems to me plausible to assume that every subject is aware of this very fact, hence also of the possibility that one cognitive state σ′ of hers/his, better than another cognitive state σ of hers/his that was a $\mathcal{C}$-justification for α, does not assign to σ the status of $\mathcal{C}$-truth-ground of α. More generally, it seems to me plausible to assume that the practice of assertion presupposes such awareness, on the part of subjects, of the possible discrepancy between justifications and truth-grounds, hence a critical attitude towards their own cognitive states, on account of which every subject judges that having a *prima facie* justification to believe α is not enough to assert α: it is necessary to acquire an optimal cognitive state from which to evaluate one's own justification.

Let us now define the notion of justification for first-person epistemic reports, i.e. reports of the form "I know that α", through a chain of definitions resulting from Definitions 1–4 by identifying the reporter with the believer:

Definition 9 Let ρ be an arbitrary mind-reading cognitive state for $\mathcal{L}_{\text{B+Kn}}$ of a subject $\mathcal{R}$ such that $[\![\text{Kn}]\!]_\rho = f_{\text{Kn}}$, $\mathcal{S}$ is both the believer $\mathcal{B}e\ell$ and the reporter $\mathcal{R}ep$, $[\![I]\!]_\rho = \mathcal{S}$, $[\![\text{that}\ \alpha]\!]_\rho = [\![\text{that}_{\ \mathcal{B}e\ell}\ \alpha]\!]_\rho = [\![\alpha]\!]_{\mathcal{S}}$; then the set $\text{EF}_{\text{Kn}(I,\ \text{that}\ \alpha),\rho}$ of the *evidential factors* of Kn(I, that α) in ρ is $f_{\text{Kn}}(I, [\![\alpha]\!]_S) \cup \mathcal{A} = \{$<Q, B($\mathcal{S}$, thatα>|Q is a why-question arising in ρ and $f_{\text{Kn}}(I, [\![\alpha]\!]_S) =\ _\rho 1$ is a potential answer to Q$\}$.

Definition 10 Let a ∈ $\text{EF}_{\text{Kn}(I,\text{that}\alpha),\rho}$; then a makes evident Kn(I, thatα) in ρ (in symbols a$\vDash_\rho$Kn(I,thatα) iff

- a $= f_{\text{Kn}}(I, [\![\alpha]\!]_S)$ and $f_{\text{Kn}}(I, [\![\alpha]\!]_S) =\ _\rho 1$; or
- a ∈ $\mathcal{A}$ and $f_{\text{Kn}}(I, [\![\alpha]\!]_S) =\ _\rho 1$ is the best answer to Q in ρ.

Definition 11 j_ρ(Kn(I, thatα)) = 1 iff there is an a ∈ $\text{EF}_{\text{Kn}(}\text{EF}_{\text{Kn}(I,\ \text{that}\alpha),\rho}$ such that a$\vDash_\rho$Kn(I, thatα).

Definition 12 An *$\mathcal{R}$-justification for the sentence* Kn(I, that α) is a mind-reading cognitive state ρ of the subject $\mathcal{R}$ such that j_ρ(Kn(I,thatα)) = 1.

We can see that

Remark 2 A sentence α is assertible by a subject $\mathcal{S}$ in the mind-reading cognitive state σ iff σ is an $\mathcal{S}$-justification for the sentence Kn(I, that α).

Suppose α is assertible by $\mathcal{S}$ in σ; by Definition 8 $\mathcal{S}$ believes that σ and $\mathcal{S}$'s believing that α is based on σ's being a truth-ground of α relative to σ; since σ is a mind-reading cognitive state, the roles of believer and reporter are assigned to the same subject $\mathcal{S}$; moreover, the property of occupying a cognitive state is transparent, the property of being a truth-ground relative to a given cognitive state is transparent, and the relation of semantic expression is transparent; hence conditions (i)–(iv) mentioned in (2) are satisfied, therefore $f_{\mathrm{Kn}}(I, [\![\alpha]\!]_{\mathcal{S}}) = {}_{\sigma}1$. Conversely, suppose that σ is an $\mathcal{S}$-justification for $\mathrm{Kn}(I, \text{that}\,\alpha)$, i.e. a mind-reading cognitive state such that $j_{\sigma}(\mathrm{Kn}(I, \text{that}\,\alpha)) = 1$; by Definition 11 there is an $\mathrm{a} \in \mathrm{EF}_{\mathrm{Kn}(I,\text{that}\,\alpha),\sigma}$ such that $\mathrm{a} \vDash_{\rho} \mathrm{Kn}(I, \text{that}\alpha)$; hence the roles of believer and reporter are assigned to the same subject $\mathcal{S}$, and the presence of features (i)-(iv) mentioned in (2) has been verified by the feature-checking algorithm; as feature-checking is assumed to be factive,[16] conditions (i)–(iv) are satisfied, and α is assertible by $\mathcal{S}$ in σ.

A consequence of Remark 2 is that the assertibility conditions of α are the same as the assertibility conditions of "I know that α"; since possession of $\mathcal{C}$-justifications is epistemically transparent, assertibility is as well.[17]

8.5 Williamson's Argument Against Transparency

An argument against the epistemic transparency of mental states has been stated by Timothy Williamson in Williamson (2000). In this section I will question its soundness. While I will give reasons to resist some of the premises of Williamson's argument, it should be added that Williamson's claim that there are non-transparent (or non-luminous) mental states that are cases of knowledge does not contradict, by itself, the possibility that justifications for sentences, as I have proposed to define them, are transparent mental states and that, consequently, their possession and the assertibility of the relative sentences are transparent as well.[18]

I will examine the two versions of Williamson's argument, whose general structure is expounded in Sect. 8.5.1. The first version derives the conclusion from an assumption about our limited discrimination capacities whose truth is questioned in Sect. 8.5.2. The second version relies on a principle about reliability to which a counterexample has been given in Berker (2008); in Sect. 8.5.3 the counterexample is defended not in terms of a constitutive dependence of certain conditions on beliefs (as in Berker's paper), but in terms of the (absolute) reliability of certain belief-forming methods.

---

[16] See (2): the manageability of the concept f_{Kn} requires the checking of a feature configuration, and such a checking is assumed to be factive, in the sense that if the feature configuration is checked to be present, then it is present.

[17] An argument for the co-assertibility of α and "I know that α" can be found in Hambourger (1987).

[18] However, Williamson argues that his argument generalizes to assertibility conditions and to proof possession (Williamson, 2000: 111).

8.5.1 The Margin-For-Error Principle

Let us introduce the terminology used by Williamson (2000: 52 and 95). A *case* is a possible total state of a system, the system consisting of an agent at a time paired with an external environment. A *condition* is specified by a "that"-clause, and either obtains or fails to obtain in each case. A condition C is *luminous* iff

(LUM): For every case α, if in α C obtains, then in α one is in a position to know that C obtains.

The notion of being in a position to know has been explained in Chap. 2, fn. 15. Williamson's argument consists in deriving an absurd conclusion from true assumptions and the hypothesis that condition C is luminous; in such a derivation he uses the following crucial *margin-for-error principle* (where v(α) and v(β) are the values of some chosen parameter in the cases α and β):

(MAR): For all cases α and β, if $|v(\alpha) - v(\beta)| < c$ and in α one is in a position to know that C obtains, then C obtains in β.

Here is the derivation:

(12) In α C obtains and $|v(\alpha) - v(\beta)| < c$.

[Assumption]

(13) In α one is in a position to know that C obtains.

[from (12) by the hypothesis that C is luminous]

(14) In β C obtains.

[from (12) and (13), by (MAR)]

(15) For all cases α and β, if $|v(\alpha) - v(\beta)| < c$ and C obtains in α, then C obtains in β.

[from (12), (13), and (14) by First Order Logic]

If we introduce the usually uncontentious assumption that the parameter v varies continuously:

(16) For all non-negative real numbers u there is a case α such that $v(\alpha) = u$,

we can derive (Williamson, 2000: 129) from (15) the absurd conclusion

(17) For all cases α and β, if C obtains in α, then C obtains in β.

Since the argument is uncontroversially valid and (16) seems acceptable, the crucial question is whether (MAR) holds. Williamson argues for it in two distinct ways, in Chaps. 4 and 5 of Williamson (2000), respectively. I will start from the argument in Chap. 5.

8.5.2 *First Argument*

The general idea is to derive the margin-for-error principle from some principles concerning reliable belief that seem to be intuitively valid. Here is an intuitive characterization of reliable belief:

> Reliability and unreliability, stability and instability, safety and danger, robustness and fragility are modal states. They concern what could easily have happened [[19]]. They depend on what happens under small variations in the initial conditions. [...]
>
> Reliability resembles safety, stability, and robustness. [...] For present purposes, we are interested in a notion of reliability on which, in given circumstances, something happens reliably if and only if it is not in danger of not happening. That is, it happens reliably in a case α if and only if it happens (reliably or not) in every case similar enough to α [[20]]. In particular, one avoids false belief reliably in α if and only if one avoids false belief in every case similar enough to α. (Williamson, 2000: 123–124)

From this characterization the following principle can be extracted:
(REL) For all cases α and β, if β is close to α and in α one knows that C obtains, then, if in β one believes that C obtains, then in β C obtains.

The second crucial assumption concerns the limited discrimination capacities of subjects of belief: for some small positive real number c,
(LIM) For all cases α and non-negative real numbers u, if in α one believes that C obtains and $|u - v(\alpha)| < c$ then, for some case β close to α, $v(\beta) = u$ and in β one believes that C obtains.

Intuitively, if one has a belief, one has that belief even when the relevant parameter takes a slightly different value.

Further, Williamson introduces some 'background assumptions' that should encounter no resistance. First, a principle according to which the obtaining of a condition in a case depends only on the value of the relevant parameter:

(18) For all cases α and β, if $v(\alpha) = v(\beta)$ then C obtains in α iff C obtains in β.

Second, a stipulation concerning the meaning of "being in a position to know", obvious in view of the preceding explanation of this notion:

(19) For all cases α, if in α one is in a position to know that C obtains then, for some case β, $v(\alpha) = v(\beta)$ and in β one knows that C obtains.

Last, the trivial assumption that knowledge implies belief:

(20) For all cases α, if in α one knows that C obtains then in α one believes that C obtains.

---

[19] «The notion of what could easily happen behaves like the dual of safety; 'It could easily have been F′' is close to 'It was not safely not F′'» (Williamson, 2000: 125).

[20] «The relevant similarity is in the initial conditions, not in the final outcome [...]. 'Initial' here refers to the time of the case, not to the beginning of the universe [...]. Just how similar the case must be to one in which an event of type E occurs for the term 'danger' to apply depends on the context in which the term is being used.» (Williamson, 2000: 124).

From (REL), (LIM) and (18)–(20) the margin-for-error principle can be derived in the following way:

(21) $|v(\alpha) - v(\beta)| < c$ and in α one is in a position to know that C obtains

[Assumption]

(22) For some case α′, $v(\alpha) = v(\alpha')$ and in α′ one knows that C obtains

[from (21) by (19)]

(23) For some case α′, $|v(\alpha') - v(\beta)| < c$ and in α′ one believes that C obtains

[from (21) and (22) by (20)]

(24) For some case β′ close to α′, $v(\beta') = v(\beta)$ and in β′ one believes that C obtains

[from (23) by (LIM)]

(25) C obtains in β′

[from (22) and (24), by (REL)]

(26) C obtains in β.

[from (25) and (24), by (18)]

The problem, now, is whether (REL), (LIM) and (18)–(20) are acceptable. Let us concede that the background assumptions (18)–(20) are, and concentrate on (REL) and (LIM). (REL) seems to be very plausible, since it is essentially extracted from the preceding intuitive characterization of reliable belief: a belief is knowledge in α only if it is reliable, and this simply means that it is true in all cases similar enough to α.

The questionable principle is (LIM). Williamson concedes that it is «not obvious» for some conditions; the main reason for this is, according to him, that for such conditions

> the underlying parameter itself constitutively depends on one's belief, as some philosophers postulate for phenomena that they would classify as response-dependent. For example, they hold that the intensity of one's pain constitutively depends on one's beliefs about the intensity of one's pain. (Williamson, 2000: 130)

If I understand correctly, the underlying reasoning is the following. If there is a relation of constitutive dependence between a condition C and the belief that C obtains, the following is true:

(27) For every case α, if one has done in α what one is in a position to do to decide whether C obtains, and one believes in α that C obtains, then C obtains in α.

Of course for most conditions (27) does not hold, but for some it is arguable that it does, and feeling pain, or cold, are cases in point. (27) is in tension with (LIM) because by using the two principles we can infer, from the assumption that one truly believes in α, the implausible conclusion that one *truly* believes in some case similar enough to α.[21] So, when (27) holds for C, (LIM) can be questioned, and a different argument for the non-luminosity of C is required; we will consider it in the next section.

However, it seems to me that the conditions for which (LIM) is not obvious, or better not acceptable, are much more numerous than Williamson is ready to concede. Let me preliminarily illustrate a distinction between computational modeling and computational explanation introduced by G. Piccinini:

> In *computational modelling* (as I'm using the term), the outputs of a computing system *C* are used to describe some behaviour of another system *S* under some conditions. The explanation for *S*'s behaviour has to do with *S*'s properties, not with the computation performed by the model. *C* performs computations in order to generate subsequent descriptions of *S*. The situation is fully analogous to other cases of modelling: just as a system may be modelled by a diagram or equation without being a diagram or equation in any interesting sense, a system may be modelled by a computing system without being a computing system in any interesting sense.
>
> In *computational explanation*, by contrast, some behaviour of a system *S* is explained by a particular kind of process internal to *S* — a computation — and by the properties of that computation. (Piccinini, 2007: 96)

A *computation*, in turn, can be characterized as a transformation of inputs to outputs according to some general rule. Inputs and outputs must be of a special sort, i.e. strings of digits.

> A *digit* is a particular or a discrete state of a particular, discrete in the sense that it belongs to one (and only one) of a finite number of types. Types of digits are individuated by their different effects on the system, that is, the system performs different functionally relevant operations in response to different types of digits. A *string* of digits is a concatenation of digits, namely, a structure that is individuated by the types of digits that compose it, their number, and their ordering (i.e., which digit token is first, which is its successor, and so on). (Piccinini, 2007: 107)

It is a matter of fact that several psychological phenomena admit a computational explanation: visual perception, linguistic perception, categorization, reasoning, and so on and so forth. So, suppose that we are concerned with the states of belief of humans about the obtaining of a certain condition C, and that both the obtaining of C and the belief that C obtains can be computationally explained in the sense

---

[21] Assume that

(a) in α one believes that C obtains, $|u-v(\alpha)|<c$, and in α C obtains.

For (LIM), (a) entails that there is a β such that

(b) β is close to α, $v(\beta) = u$, and in β one believes that C obtains.

(a) and (b) entail that there is a β such that

(c) $|v(\beta)-v(\alpha)|<c$, β is close to α, and in β one believes that C obtains.

For (27), (c) entails that there is a β such that

(d) $|v(\beta)-v(\alpha)|<c$, β is close to α, in β one believes that C obtains, and in β C obtains.

described. This means that, in particular, the inputs and outputs of the system S under investigation (for instance, the perceptual apparatus of a human subject) are strings of digits; therefore their totalities form a *discrete* space. As a consequence, if the obtaining or not obtaining of a condition C in a case depends on the values of some parameter v in that case, the set of arguments and values of this parameter will form discrete spaces, in the sense that between two contiguous arguments (values) of v there is no other argument (value). When v takes arguments and values in discrete sets, (LIM) must be phrased as follows:
(LIM') For all cases α and natural numbers n, if in α one believes that C obtains and $|n - v(\alpha)| < c$ then, for some case β close to α, $v(\beta) = n$ and in β one believes that C obtains.

For many conditions (LIM') is not true. Suppose for example that the human sensory apparatus is structured in such a way that, as the values of v lower, a smaller and smaller number of neurons is activated, with a sudden decrement at k, as shown in Fig. 8.1; and that the human belief system concerning C has a sort of threshold as indicated in the Figure: above 10 activated neurons the subject believes that C obtains, below the subject does not believe that C obtains. Let α be a case in which one believes that C obtains and $v(\alpha) = k$: since for all $n < k$ the number of activated neurons is less than 10, there is no β close to α in which one believes that C obtains.[22] Instances of conditions satisfying the preceding assumptions are all the conditions such that (i) their obtaining or not obtaining is best explained in computational terms, and (ii) for which threshold effects are known.[23] There is a lot of such conditions; to quote only one example, all speech perception is categorical in adults, in the sense that a stimulus that is variyng continuously, for instance between [pa] and [ba], is perceived categorically as either [pa] or [ba], with a sudden phase transition around a natural perceptual boundary (Liberman et al., 1967).

8.5.3 *Second Argument*

Let us now consider the argument in Chap. 4, a subtler argument required by the fact that principle (LIM) may not be acceptable for conditions—like feeling cold—such that the underlying parameter constitutively depends on one's belief.[24] Williamson envisages the following situation:

> Consider a morning on which one feels freezing cold at dawn, very slowly warms up, and feels hot by noon. One changes from feeling cold to not feeling cold, and from being in a

[22] This is not true in a continuous case space, of course.

[23] Cp. Dubucs (2002) and references therein.

[24] The relation between the first and the second argument is characterized by Williamson by saying that the second «depends on applying reliability considerations in a subtler way to degrees of confidence», while the first «models those considerations under highly simplified assumptions, which permit us to restrict our attention to the binary contrast between believing and not believing.» (Williamson, 2000: 127).

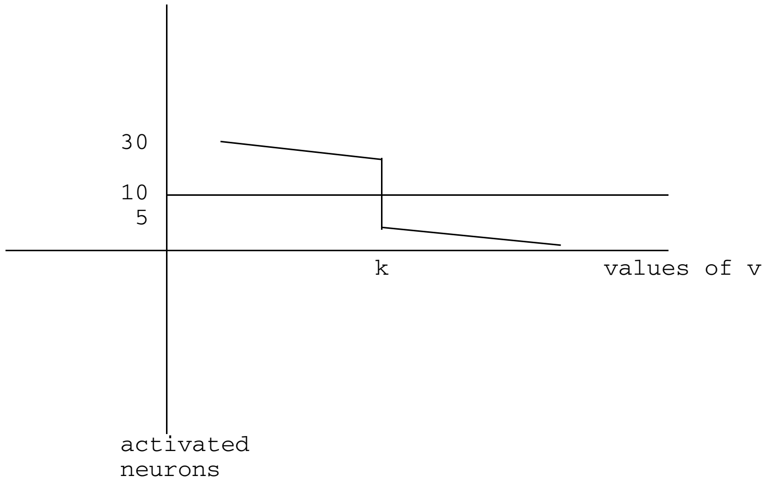


Fig. 8.1 .

> position to know that one feels cold to not being in a position to know that one feels cold. If the condition that one feels cold is luminous, these changes are exactly simultaneous. Suppose that one's feelings of heat and cold change so slowly during this process that one is not aware of any change in them during a millisecond. Suppose also that throughout the process one thoroughly considers how cold or hot one feels. One's confidence that one feels cold gradually decreases. (Williamson, 2000: 96–97)

Let t_0, t_1, …, t_n be a series of times at one millisecond intervals from dawn to noon, and α_i be the case at t_i ($0 \leq i \leq n$). The argument that the condition that one feels cold is not luminous is based on the following conditional:

(28i) If in α_i one knows that one feels cold, then in α_{i+1} one feels cold.

Let us compare (28i) with (MAR). The main differences are that (MAR) is a general principle concerning an arbitrary condition C, while (28i) concerns stage i of a process concerning the condition of feeling cold; and that the former contains in the antecedent the clause $|v(\alpha) - v(\beta)| < c$, which is satisfied by the cases α_i and α_{i+1} occurring in the latter. It is therefore not surprising that (28i) plays, in the derivation of a false conclusion from (LUM), essentially the same crucial role as (MAR) in the derivation, presented in Sect. 8.5.1, of the absurd (17)[25]: (28i) is a margin-for-error

[25] Here is the derivation:

(a) In α_i one feels cold. [Assumption].
(b) In α_i one knows that one feels cold. [From (a) and the hypothesis that feeling cold is luminous), by First Order Logic].
(c) In α_{i+1} one feels cold. [From (b) and (28i), by MP].
(d) In α_0 one feels cold. [For α_0 is at dawn].
(e) In α_n one feels cold. [From (d), by repeating (a)–(c) n times].

But (e) is false, for α_n is at noon.

conditional. The question is therefore how does Williamson argue for (28i); here is the crucial passage:

> Consider a time t_i between t_0 and t_n, and suppose that at t_i one knows that one feels cold. Thus one is at least reasonably confident that one feels cold, for otherwise one would not know. Moreover, this confidence must be reliably based, for otherwise one would still not know that one feels cold. Now at t_{i+1} one is almost equally confident that one feels cold, by the description of the case. So if one does not feel cold at t_{i+1}, then one's confidence at t_i that one feels cold is not reliably based, for one's almost equal confidence on a similar basis a millisecond earlier that one felt cold was mistaken. In pictoresque terms, that large proportion of one's confidence at t_i that one still has at t_{i+1} is misplaced. Even if one's confidence at t_i was just enough to count as belief, while one's confidence at t_{i+1} falls just short of belief, what constituted that belief at t_i was largely misplaced confidence; the belief fell short of knowledge. (Williamson, 2000: 97)

Let me compress the statement of the crucial step of this argument in the following principle (C is the condition that one feels cold):

(REL*)

If

(a) at t_i one's confidence that C obtains is of degree k,
(b) at t_i one's confidence that C obtains is reliably based,
(c) at t_{i+1} one's confidence that C obtains is of degree k–δ (for some small δ),

then

(d) at t_{i+1} C obtains.

From (REL*) and

(LIM*)

For every integer i ($0 \leq i \leq n$), if at t_i one's confidence that C obtains is of degree k, then one's confidence that C obtains at t_{i+1} is of degree k-δ (for some small δ)

we can infer

(29) For every integer i ($0 \leq i \leq n$), if at t_i one knows that C obtains, then at t_{i+1} C obtains.

Clearly (29) is a margin-for-error principle; the question is therefore whether (REL*) and (LIM*) can be accepted as valid. Notice that the two principles play, in the derivation, roles analogous to the roles played, respectively, by (REL) and (LIM) in the derivation of (MAR): (REL*), like (REL), exploits some features that one is supposed to associate to the intuitive notion of reliability, and (LIM*), like (LIM), concerns our limited discrimination capacities. However, the intuitive plausibility of the two principles is reversed with respect to (REL) and (LIM). (LIM) was a proposal about how to model the variation of belief states—an inadequate one, I argued, in the case of belief states about conditions that can be explained in computational terms. On the contrary, (LIM*) concerns degrees of confidence, not states of belief, and it says that they vary continuously from time to time: it is a correct generalization to a condition C of a feature that is empirically true of feeling cold in the situation envisaged by Williamson.[26] Consider now (REL) and (REL*); (REL) was obviously true, since it was simply extracted from the intuitive characterization of reliable *belief*; (REL*) is a much stronger principle, since it derives essentially the same conclusion as (REL)—that at t_{i+1} C obtains—not from the premise that at t_i one (reliably) *believes* that C obtains, but from the much weaker premise that at t_i one's (reliable) confidence that C obtains is of degree k, where k can be *below the cut-off point* between the cases in which one does and the cases in which one does not (reliably) believe that one feels cold. Williamson gives no independent reason to accept this principle; in particular, no reason for asserting that, if the premises (a) and (c) of (REL*) are true, and the conclusion (d) is false, then the premise (b) is false, i.e. at t_i one's degree-of-confidence that C obtains is not reliably based, or, more simply, not reliable.[27] As a matter of fact, it is not clear what does Williamson mean by "reliable" as applied to degrees-of-confidence.

He cannot mean a simple generalization to degrees-of-confidence of the notion of reliability applied to belief-states, from which the principle (REL) was extracted, since this would beg the question: (REL) was acceptable as applied to belief states, but (REL*) cannot be proposed as an innocuous generalization of (REL) in a context in which precisely the acceptability of (REL*) is in question.

Berker (2008) gives an interesting counterexample to (REL*) and then, discussing a possible objection to it on behalf of Williamson, implicitly suggests a meaning Williamson might be assigninig to "reliable". Since I will need to discuss some aspects of the counterexample, it is useful to recall it briefly. First Berker introduces a scenario in which one's subjective feelings of temperature vary from t_0 to t_n, taking values from 50 'freezons'[28] at t_0 to −50 freezons at t_n; simultaneously also one's degrees of confidence that one feels cold vary from t_0 to t_n, taking values from 1 to 0, according to the following equation (where $f(t_i)$ is the measure in freezons of one's subjective feeling of temperature at t_i and $c(t_i)$ is the measure in real numbers of one's degree of confidence at t_i that one feels cold):

(30) $f(t_i) = 100c(t_i) - 50$ freezons.

What is now necessary to do is to define the conditions at which one *believes* that one feels cold, and the conditions at which one *feels* cold; Berker's crucial assumption is that the following *penumbral connection*[29] obtains between the two kinds of conditions:

(31) One believes that one feels cold iff one feels cold.

[26] Of course, the conditions which it is meant to apply to cannot be computationally explained in the above sense, but at most computationally modelled.

[27] S. Berker convincingly argues (Berker, 2008: 9–10) that shifting from "reliable" to "reliably based" makes no difference as to the possibility of deriving a margin-for-error principle from a safety requirement. Analogous considerations apply to the possibility of arguing for (REL*); I shall therefore concentrate on "reliable".

[28] Freezons are introduced by Berker as units of measurement of the intensity of one's subjective feelings of cold.

[29] Berker borrows this name from Fine (1975).

In this scenario, if one's degree of confidence at t_i that one feels cold is reliably based, then (REL*) is false, since it may happen that at t_i one knows that one feels cold, whereas at t_{i+1} one's degree of confidence that one feels cold is only slightly smaller, but enough for *not* believing that one feels cold, so that at t_{i+1} one does not feel cold. For example, suppose that the threshold of what counts as believing that one feels cold is at 0.8:

- if $c(t_i) > 0.8$, then one believes that one feels cold,
- if $c(t_i) \leq 0.8$, then one does not believe that one feels cold.

From this and from equation (30) it follows that

- if $f(t_i) > 30$ freezons, then one feels cold,
- if $f(t_i) \leq 30$ freezons, then one does not feel cold.

Now let $c(t_i) = 0.8 + \delta$ (for some small real number $\delta > 0$) and $c(t_{i+1}) = 0.8$. Under the hypothesis that $c(t_i)$ is reliable, (REL*) is false: (REL*)(a)–(c) are true, but (REL*)(d) is false, because at t_{i+1} one does not believe that one feels cold, and therefore, for (31), one does not feel cold.[30]

It might be retorted, on behalf of Williamson, that $c(t_i)$ is *not* reliable; of course, this move is possible only if "reliable" is understood in some definite sense; Berker assumes that it means something like "similar enough to the one assigned by the ideal degree-of-confidence profile",[31] and convincingly argues that even understanding it in this sense (REL*) cannot be justified. I will not enter into the details of his reasoning; what is relevant to my point is only that the extreme attempt at rescuing Williamson's argument is blocked by Berker by invoking the possibility that there is a sort of *generalized* relation of constitutive dependence holding not only between one's feeling cold and one's state of belief that one feels cold, but more generally between the intensity of one's feeling cold and one's degree-of-confidence that one feels cold. In general, the justification of (REL*) is blocked, according to Berker, whenever the condition C is such that such a generalized relation of constitutive dependence can be invoked; the reason, if I understand him correctly, is that when that relation subsists the following is true:

(32) There is a one-to-one monotonic function f from degrees-of-confidence that C obtains to intensities of C's obtaining such that

(i) for every case α, if one has done in α what one is in a position to do to establish the degree of C's obtaining in α, and one's degree-of-confidence in α that C obtains is k, then the degree of C's obtaining in α is $f(k)$;

[30] It may be useful to repeat that the hypothesis that $c(t_i)$ is reliably based is not falsified by simply adopting the characterization of reliability quoted above from page 124 of Williamson (2000): in order to avoid false belief reliably in α one must avoid false belief in every case similar enough to α, but nothing, in that characterization, prevents one from having possibly 'false' degrees-of-confidence lower than belief in some case similar enough to α.

[31] The ideal profile is shown through a figure, but it is not necessary to give it here.

(ii) one's degree-of-confidence k in α that C obtains is above the threshold of one's belief that C obtains if and only if *f*(k) is above the threshold of C's obtaining.

Then, by (32)(ii),

This fact would be in tension with (REL*) for reasons analogous to the ones for which (3) is in tension with (LIM).[32]

It seems to me that the possibility of giving Berker-type counterexamples to (REL*) is not limited to the case of conditions for which a (generalized) relation of constitutive dependence can be invoked, and that consequently the significance of such counterexamples is much more general than Berker is willing to admit. If we analyse his counterexample, we can see that what is required in order to construct it is only the truth of the following:

(33) For every case α, one believes in α that C obtains if and only if C obtains in α.

Of course, if a relation of constitutive dependence exists between either the obtaining of C and one's believing it to obtain or between the intensity of C's obtaining and one's degree-of-confidence that C obtains, then (33) is true; but (33) may be true also when such a relation does not exist. In fact, the truth of (33) may be understood as a consequence not of the nature of condition C, but of the method employed to acquire beliefs about C's obtaining. Let us call a method M forming beliefs about a condition C *absolutely reliable* iff, for every case α, if one believes in α that C obtains on the basis of M, then C obtains in α; and call M *complete* iff, for every case α, if C obtains in α and one has applied M in α, then one believes in α that C obtains. Obviously, if M is absolutely reliable and complete every belief formed through it satisfies (33).

But do absolutely reliable belief-forming methods exist? Consider the condition that the temperature outside is less than 30°, and suppose our method M to acquire beliefs about that condition consists in consulting a thermometer; if the thermometer works and is sensible enough, M is absolutely reliable and complete, so (33) is true. On the other hand, whatever a relation of constitutive dependence exactly is, nobody

---

[32] Analogous, but not identical, since, as we have seen, (LIM) and (REL*) have a different significance. Presumably, the reason of the tension would be that (REL*) and (32) are inconsistent. Here is why:

(a) Assume that k is the threshold of one's belief that C obtains is k, and that

- at t_i one's confidence that C obtains is of degree k,
- at t_i one's confidence that C obtains is reliably based, and
- at t_{i+1} one's confidence that C obtains is of degree k–δ, (for some small δ);

then, by (REL*),

(b) at t_{i+1} C obtains, i.e., the intensity *f*(k) at t_{i+1} of C's obtaining is above the threshold of C's obtaining.

Then, by (32)(ii),

(c) one's degree-of-confidence at t_{i+1} that C obtains is above the threshold of one's belief that C obtains,

in contradiction with our last assumption in (a).

Notice that the "if" part of (32)(ii) is essential to the derivation of the contradiction.

would say that it holds between one's believing that the temperature outside is less than 30° and the temperature outside being less than 30°.

8.5.4 Conclusion

According to Berker, what ultimately blocks Williamson's arguments is the possibility that there is a relation of constitutive dependence or of generalized constitutive dependence between the obtaining of a condition C and one's believing it to obtain or between the degree to which C obtains and the degree to which one believes that C obtains. As the existence of such a relation is arguable in the case of feeling cold and believing that one feels cold, and in similar cases of beliefs about certain mental states, Berker concludes that margin-for-error principles must be rejected in those cases; on the contrary, they are obviously acceptable in many other cases:

> Few would doubt that a version of Williamson's argument can establish that an external-world conditions such as *that the temperature outside is less that 90°F* is not luminous. (Berker, 2008: 18)

However, my remarks in Sect. 8.5.3 suggest a different analysis. Berker argues in the following way for the thesis that the condition that the temperature outside is less that 90 °F is not luminous:

> We can imagine a morning on which the outside temperature starts at 70°F and then slowly warms up to 110°F, all while a given subject does nothing but carefully attend to whether the temperature outside feels to be less than 90°F. In this case the analogue of [a margin-for-error principle] obviously holds: [...] given the inexactness of our abilities to detect the external temperature around us, there must be some sufficiently similar situations in which one does not stop believing (or in which one's degree of confidence does not significantly drop) at that first point at which the outside temperature stops being less than 90°. (Berker, 2008: 19)

My objection is that the argument supports the non-luminosity of the piece of knowledge that the temperature outside is less that 90 °F *whenever this knowledge is obtained by the method in question*, not in general. So, the condition that the temperature outside is less that 90 °F is not luminous, indeed, but only relatively to the method described by Berker; and the reason of this non-luminosity is obvious: the method described by Berker is blatantly unreliable and non-complete; if we employed a reliable method, for example the one described above, Berker's counterexample could be immediately reconstructed.

Of course, even when a method is reliable the inexactness of our abilities (or of our instruments) is always present; but once we have established the degree of approximation we can tolerate, it becomes utterly irrelevant: if we have previously decided that the measurement of a temperature is acceptable with an approximation of one millionth of degree, the proposition "The temperature outside is 90 °F" will be *true* if the temperature outside is 89.999999 °F.

Concluding, the margin-for-error principles are valid, and therefore luminosity is false, whenever it is relevant to keep into account the

> inexactness of our abilities to detect [the obtaining of a condition C, because] there must be some sufficiently similar situations in which one does not stop believing (or in which one's degree of confidence does not significantly drop) at that first point at which [the condition C stops obtaining]. (*Ibid.*)

But in the cognitive situations in which it is appropriate to speak of scientific knowledge it is relevant *not* to keep into account such inexactness; in those situations it is necessary to assume that the methods employed are absolutely reliable, and therefore the margin-for-error principles are not valid. A scientific theory is precisely a context in which the limits of *tolerability* of the unavoidable unreliability of the belief-forming methods employed are fixed in advance: within these limits, it is relevant *not* to take into account such unreliability, and the margin-of-error principle is consequently false.

The discussion of Williamson's second argument points out a reason for considering irrelevant the limits of our discriminating capacities that is much more general than the one produced in the discussion of Williamson's first argument: there it was the exigency of giving a computational explanation of certain psychological phenomena, here the exigency of giving a scientific account of any sort of phenomena.

8.6 Conclusion

We have seen in Chap. 3, Sects. 3.2.1 and 3.2.3, that the key notion of a theory of meaning must be simultaneously epistemically transparent, in order to explain understanding, and capable to explain assertion conditions. Assertion conditions are normally explained in terms of the factiveness of the key notion, but this generates a tension with the requirement of transparency; I have therefore suggested to explain assertion conditions in terms of k-factiveness, i.e. in terms of knowledge. In this chapter I have defined a notion of $\mathcal{C}$-justification for "$\mathcal{S}$ knows that α" intended to be adequate to the intuitive notion of (propositional) knowledge; if the definition is adequate, the notion of truth-ground of α at the same time is the k-factive notion we are looking for and yields a solution to Gettier problems. But a last, crucial, step is required: we must show that once knowledge, defined in the terms proposed in this chapter, replaces realist truth in the explanation of meaning, understanding and assertion, no independent notion of truth is required for a theory of meaning. This is the object of the next chapter.

References

Berker, S. (2008). Luminosity regained. *Philosophers' Imprint*, *8(2)*. https://www.philosophersimprint.org/008002/

Dubucs, J. (2002). Calculer, percevoir et classer. *Archives de philosophie, 65*, 335–355.

Fine, K. (1975). Vagueness, truth, and logic. *Synthese, 30*, 265–300.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis, 23*, 121–123.
Goldman, A. I. (1976). Discrimination and perceptual knowledge. *The Journal of Philosophy, 73*, 771–791.
Hambourger, R. (1987). Justified assertion and the relativity of knowledge. *Philosophical Studies, 51*, 241–269.
Harman, G. (1968). Knowledge, inference, and explanation. *American Philosophical Quarterly, 5*, 164–173.
Harman, G. (1980). Reasoning and evidence one does not possess. *Midwest Studies in Philosophy, 5*, 163–182.
Liberman, A., et al. (1967). Perception of the speech code. *Psychological Review, 74*, 431–461.
Lehrer, K., & Paxson, T. (1969). Knowledge: Undefeated justified true belief. *The Journal of Philosophy, 66*, 225–237.
Piccinini, G. (2007). Computational modelling versus computational explanation: Is everything a turing machine, and does it matter to the philosophy of mind? *The Australasian Journal of Philosophy, 85*, 93–115.
Pollock, J. L. (1986). *Contemporary theories of knowledge*. Rowman & Littlefield.
Shope, R. K. (1983). *The analysis of knowing: A decade of research*. Princeton University Press.
Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.
Zagzebski, L. (1994). The inescapability of gettier problems. *The Philosophical Quarterly, 44*, 65–73.

Chapter 9
The Paradox of Knowability

Abstract In this chapter the significance of the Paradox of Knowability is discussed with respect to the question of how to conceive truth within an anti-realist conceptual framework. In Sect. 9.1 the Paradox is introduced; Sect. 9.2 articulates the intuitionistic equation of truth with knowledge, first by putting into evidence (Sects. 9.2.1–9.2.3) the conditions at which the equation is acceptable: transparency of knowledge and 'disquotational property' of truth; then by showing (Sects. 9.2.4 and 9.2.5) how the charge of inconsistency can be resisted. In Sect. 9.3 the neo-verificationist approaches to the Paradox are discussed, and it is shown how the Paradox hits the neo-verificationist idea of the necessity of a notion of truth irreducible to proof possession. In Sect. 9.4 the Dummettian problem is discussed of how a debate between alternative logics can be rationally shaped.

Keywords Intuitionism · Knowability paradox · Anti-realistic theory of meaning · Truth notions · Internal truth · BHK-Explanation · Neo-Verificationism · Philosophy of logic

The main objection raised by the realists to the 'epistemic' conceptions of truth is the Paradox of Knowability, an argument for the conclusion that there are unknowable truths. In this chapter I will discuss the significance of the paradox with respect to the question of how to conceive truth within an anti-realist conceptual framework. We have seen in Chap. 6 that a major divide separates intuitionists from anti-realists of other sorts concerning the issue of truth: for the former truth amounts to knowledge, for the latter truth cannot be reduced to knowledge, and what characterizes truth is its knowability. It is precisely this feature of anti-realistic truth that the paradox is intended to hit.

After having introduced the Paradox (Sect. 9.1), in Sect. 9.2 I articulate the intuitionistic equation of truth with knowledge, first by putting into evidence (Sects. 9.2.1–9.2.3) the conditions under which the equation is acceptable: transparency of knowledge and 'disquotational property' of truth, then by showing (Sects. 9.2.4 and

G. Usberti, *Meaning and Justification. An Internalist Theory of Meaning*, Logic, Epistemology, and the Unity of Science 59,
https://doi.org/10.1007/978-3-031-24605-0_9

9.2.5) how the charge of inconsistency can be resisted. In Sect. 9.3 I discuss the neo-verificationist approaches to the Paradox, and in Sect. 9.4 I tackle the Dummettian problem of how a debate between alternative logics can be rationally shaped.

The theory of meaning outlined in this book shares with intuitionism the attitude towards truth, on the one hand, and the conception of knowledge of proofs as epistemically transparent; since these are the two aspects involved in the analysis of the paradox, in this chapter I shall make reference chiefly to intuitionism, and only occasionally to my own theory.

9.1 Introduction

The Paradox of Knowability[1] is an argument that from the principle of Knowability

(K) Every truth is knowable,

derives the principle of Strong Verificationism

(SV) Every truth is known.

The argument has the following structure.[2] First, (K) and (SV) are formalized by the two following schemas, respectively:

(1) $\alpha \supset \Diamond K\alpha$
(2) $\alpha \supset K\alpha$.

Then replace α in (1) with the sentence "$q \,\&\, - Kq$"; you obtain the following instance:

(3) $(q \,\&\, - Kq) \supset \Diamond K(q \,\&\, - Kq)$,

from which it is not difficult to derive, by means of intuitively acceptable principles, the unacceptable (2). The principles are the following:

(4) $\Box(K(\alpha \,\&\, \beta) \supset (K\alpha \,\&\, K\beta)$
(5) $\Box(K\alpha \supset \alpha)$
(6) $\Box((\alpha \,\&\, - \alpha) \supset \bot)$
(7) $- (\alpha \,\&\, - \beta) \supset (\alpha \supset \beta)$,[3]

---

[1] The paradox is usually ascribed to F. Fitch, but is due in fact to A. Church. For its history see Salerno (2009a).

[2] As anticipated in the Preface, I shall use the symbols $\&$, $+$, $\supset$, $\equiv$, $-$, $\prod$, $\sum$ for the classical logical constants; and $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$, $\neg$, $\forall$, $\exists$ for the intuitionistic ones. The symbol "K" is used throughout this chapter as an operator, to be read as "It is (presently) known that"; if it were used as a predicate, $K\alpha$ would be an abbreviation of $\exists$ xKn(x,that α).

[3] This principle is not intuitionistically valid.

and this is the derivation:

```
      [q & − Kq]¹
      −−−−−−−−−−−− by (3)
      ◇K(q & − Kq)
      −−−−−−−−−−−− by (4)
      ◇(Kq & K − Kq)
  (8) −−−−−−−−−−−− by (5)
      ◇(Kq & − Kq)
       −−−−−−−− by (6)
           ◇⊥          − ◇⊥
 −−−−−−−−−−−−−−−−−− ⊃ ∃
         ⊥
           −−−−−−−− I, 1
     −(q& − Kq)
                −−−−−− by (7)⁴
                  q ⊃ Kq
```

The argument is not a paradox for the realist, who refuses (K), nor for the intuitionist, who accepts (SV): «there are no non-experienced truths», to quote Brouwer's famous dictum (Brouwer, 1949: 488). The argument is a paradox for the weak verificationist, who identifies truth with knowability, because knowability is intuitively distinct from knowledge; Dummett, for instance, finds (2) «contrary to our strong intuition» (Dummett, 2009: 51); for Martin-Löf it is «very counter-intuitive to say that a proposition becomes true when it is proved» (Martin-Löf, 1991: 142); Prawitz judges «a strange and unfortunate use» the intuitionists' use of "true" as «synonymous with the truth as known» (Prawitz, 1980: 8). The argument seems to show that the intuitive distinction between knowability and knowledge is untenable.[5] Contraposing if one does not want to accept (SV), one must reject (K) as well. The problem, for the intuitionist, is whether his/her position is consistent. In this chapter I shall argue that it is.

9.2 The Intuitionistic View

The first step towards an explanation of the intuitionistic view of the paradox consists in remarking that the formalizations of (K) and (SV) by (1) and (2), respectively, acquire a meaning very different from the intuitive one if the logical constants occurring in them are understood according to Heyting's explanation, or according to the

---

[4] Of course this last step is not intuitionistically valid.

[5] Since (7) is intuitionistically invalid, an intuitionist might accept the negation of (K), i.e. the penultimate step of derivation (8), without accepting that there are unknowable truths: $\neg\forall x\,\alpha \rightarrow \exists x\neg\alpha$ is not intuitionistically valid.

Proof Explanation, or even according to the definition of justification for α given in Chaps. 4 and 5; and that the intuitionistic formula corresponding to (2), namely
d

(2′) $\alpha \rightarrow K\alpha$,

is valid, regardless of Church-Fitch's argument (whose last step is not intuitionistically valid).

9.2.1 (2′) is Intuitionistically Valid

The intuitionistic meaning of the logical constants, as applied to mathematical statements, is elucidated by Heyting's Explanation, given in Chap. 2, (20), and made more precise by the Proof Explanation (Chap. 2, (21)), which for convenience I reproduce in the following table:

Definition 1

a proof of	IS
$\alpha \wedge \beta$	given by presenting a proof of α and a proof of β
$\alpha \vee \beta$	given by presenting either a proof of α or a proof of β (plus the stipulation that we want to regard the proof presented as evidence for $\alpha \vee \beta$)
$\alpha \rightarrow \beta$	a construction which permits us to transform any proof of α into a proof of β
$\neg\alpha$	a construction which transforms any hypothetical proof of α into a proof of a contradiction. Absurdity ⊥ (contradiction) has no proof
$\forall x\alpha$[a]	a construction which transforms a proof of $d \in D$ into a proof of $\alpha[\underline{d}/x]$[b]
$\exists x\alpha$[a]	given by providing $d \in D$, and a proof of $\alpha[\underline{d}/x]$[b]

[a] Where x varies on D
[b] $\underline{d}$ is a name of d

As for the explanation proposed in Chap. 5 of this book, inspired by Heyting's Explanation and intended to be applied to empirical as well as to mathematical statements, it can be summarized through the following table, which facilitates the comparison with the intuitionistic explanation. (In the table, $a \vDash_\sigma \alpha$ is to be read as "a makes evident α in σ".)

Definition 2

a justification for	is a cognitive state σ in which one knows
$\alpha \wedge \beta$	a pair $\prec a, b \succ$ such that $a \vDash_\sigma \alpha$ and $b \vDash_\sigma \beta$

(continued)

(continued)

a justification for	is a cognitive state σ in which one knows
$\alpha \vee \beta$	a procedure p whose execution yields an a such that, for every cognitive state $\sigma' \geq \sigma$, either $a \vDash_{\sigma'} \alpha$ or $a \vDash_{\sigma'} \beta$
$\alpha \rightarrow \beta$	a constructive function f such that, for every a such that $a \vDash_{\sigma} \alpha$, $f(a) \vDash_{\sigma} \beta$
$\perp$	a contradiction, i.e. a * such that not $* \vDash_{\sigma} \perp$
$\forall x \alpha$[a]	a constructive function f such that, for every $d \in D_{\sigma}$, $f(d) \vDash_{\sigma} \alpha[\underline{d}/x]$[b]
$\exists x \alpha$[b]	a procedure p whose execution yields a pair $\prec d, a \succ$ such that, for every cognitive state $\sigma' \geq \sigma$, $d \in D_{\sigma}$ and $a \vDash_{\sigma'} \alpha[\underline{d}/x]$[b]

[a] Where x varies on D_{σ}
[b] $\underline{d}$ is a name of d

In order to define the meaning of (2′), we must ask how to define a proof of Kα, when α is a mathematical sentence, and a justification for Kα, when α is either an empirical or a mathematical sentence. There is no official intuitionistic answer, and there are several possibilities, according to the intended intuitive reading of K: "α is presently known by someone", "α is known by someone at some time", "I know that α", "α is presently known", and so on.[6] I shall choose the last reading as the most congenial to intuitionistic ideas, and I shall give the following two definitions, adapted to proofs of α as characterized by Definition 1, and to justifications for α as characterized by Definition 2, respectively:

Definition 3 A proof of Kα is the observation o that what one is presented with is a proof of α.

Definition 4 A justification for Kα is a cognitive state σ in which one has checked[7] that σ is a justification for α.

If the meaning of the logical constants is explained through either Definition 1 or Definition 2, (2′) turns out to be valid. I shall come back to this point in Sect. 9.2.4; here I only show this in the case α is atomic. I shall make reference only to Definition 2, since the case of Definition 1 is analogous.

To show that (2′) is valid we must show that there is a cognitive state in which one can compute a function f transforming every justification for α into a justification for Kα. If σ is a justification for α, then, as justifications for atomic sentences are epistemically transparent, there is a checking c that σ is a justification for α, hence a cognitive state $\sigma' \geq \sigma$ in which c has been performed and $c \vDash_{\sigma'} \alpha$ is a justification for K(α), and we can define $f(\sigma) = \sigma'$.

In conclusion (2′), far from saying that every *intuitive* truth is known, says that proofs, or justifications, are epistemically transparent, and is therefore obviously true.

[6] In some of these readings knowledge is expressed by a predicate, in others by an operator. In this chapter I shall disregard this difference, and treat "K" as an operator.

[7] Through a process of feature-checking; see Chap. 8, (2).

T. Williamson has raised the following objection to the validity of (2′). He first argues that a proof of $\alpha \rightarrow \beta$ should be conceived by intuitionists as a function *f* from proof-tokens to proof-tokens

> that is *unitype* in the sense that if *p* and *q* are proof-tokens of the same type then so are *f(p)* and *f(q)*. (Williamson, 1988: 430)

Then, under the assumption that

> a proof of $\alpha \rightarrow K\alpha$ is a unitype function that evidently takes any proof-token of α to a proof-token, for some time *t*, of the proposition that α is proved at *t* (*Ibid*.)[8]

he shows that, if α has not yet been decided, the function *f* that associates to every proof-token of α a proof-token of the proposition that α is proved at *t* is not unitype: if the proof-token *p* is carried out at t_1 and the proof-token *q* is carried out at t_2, where $t_1 \neq t_2$, then $f(p) \neq f(q)$, since the proposition that α is proved at t_1 is different from the proposition that α is proved at t_2. However, the quoted assumption is by no means conceptually necessary, nor is it a consequence of the general conception of proofs of conditionals as unitype functions. If we assume that a proof of $\alpha \rightarrow K\alpha$ is a unitype function that takes any proof-token of α to a proof-token of the proposition that α is proved (with no mention of the time at which it is proved), *f is* unitype.

As I said above, (2′) is intuitionistically valid independently of Church-Fitch's argument; this dispenses us from the necessity of investigating about the validity of the intuitionistic formula corresponding to (1), namely

(1′) $\alpha \rightarrow \Diamond K\alpha$,

and about the meaning of the possibility operator within an anti-realist conceptual framework. These questions, however, are crucial for the supporter of weak verificationism; I shall therefore discuss them in Sect. 9.3, when I will analyze the neo-verificationist approaches to the Paradox.

9.2.2 *Truth Notions*

The second step towards an explanation of the intuitionistic way of viewing the paradox consists in looking for a plausible *intuitive* sense of (SV), according to which (SV) becomes acceptable. Of course, there is a sense in which (SV) is *not* acceptable (namely when "truth" is given its usual, realist, sense); my question is whether there is a sense in which it is. A first component of such a sense has already been made explicit: it consists in reading the logical constants according to their intuitionistic meaning, specified through either Definition 1 or Definition 2. The second component is of course the concept of truth, which (SV) explicitly refers to. It is at this point that a question becomes crucial: under which conditions is a notion a notion *of truth*?

---

[8] For uniformity with the main text, in the quoted passage I have replaced "P" with "α".

First, let me explain why, exactly, the question is crucial. If we read a formula of the language of classical propositional logic (CPL), it is natural and correct to read an occurrence in it of a propositional letter, say p, as "p is true"; for example, the intuitive reading of an instance of the schema $\alpha + - \alpha$ would be, "Either p is true or $-$ p is true" (which, given the definition of "p is false" as "$-$ p is true", is equivalent to "Either p is true or p is false"). This is correct because the key notion of the realistic explanation of the meaning of the logical constants is the realistic (i.e. bivalent) notion of truth; but it is no longer legitimate when we consider a formula of the language of IPL, since the key notion of the intuitionistic explnation(s) is *not* the (bivalent) notion of truth. As a consequence, the simple occurrence of p will not be sufficient to make reference to the truth of p: in order to make reference to the truth of p it will be necessary to use a truth-predicate, or a truth-operator. Notoriously, the choice between expressing truth with a predicate or an operator has an impact on many other things, in particular on the possibility of expressing semantic paradoxes; since the questions discussed in this chapter are independent of such a possibility, I shall choose the simpler alternative of expressing truth by means of an operator. The question arises at this point: what makes an operator a *truth* operator?

A plausible answer to this question is offered by Tarski's Convention T, in the case truth is expressed by a predicate. Tarski has proposed to consider a definition of truth as materially adequate if it entails every sentence of the form

(9) N is true if and only if t,

where N is the name of a sentence of the object language and t is a translation of that sentence into the metalanguage. Since "materially adequate" means faithful to our intuitions about the notion of truth, we can take the validity of (9) as a criterion for a formally defined predicate to be a truth-predicate, i.e. a predicate defining a notion we are intuitively prepared to consider a notion of truth.[9] From this we may easily extract an analogous condition for an operator: an operator O can be seen as a truth-operator if it is defined in such a way that it entails every sentence of the form

(10) $O\alpha$ if and only if t,

where α is a sentence of the object language and t is a translation of that sentence into the metalanguage. Finally, if we make the further simplifying assumption that the metalanguage is an extension of the object language, (10) is equivalent to

(11) $O\alpha$ if and only if α,

which is the usual version of what I shall call "The (T) Schema". So, my proposal is that an operator is to be considered as a truth operator if its meaning is defined in such a way as to satisfy the (T) Schema.

Before going on, let me examine an objection to this proposal raised by Dummett. In *The Logical Basis of Metaphysics* he writes:

---

[9] If I understand it correctly, Milne (1999: 148) makes essentially the same point.

> It is sometimes alleged that what makes a given notion a notion *of truth* is that it satisfies all instances of the (T) schema. This is wrong [...]. If a constructivist proposes that the only intelligible notion of truth we can have for mathematical statements is that under which they are true just in case we presently possess a proof of them, he is offering a characterisation of truth for which the (T) schema fails, since truth, so understood, does not commute with negation. (Dummett, 1991: 166)

Let me try to make the argument explicit. Dummett is envisaging the case of a constructivist who equates the truth of a (mathematical) statement α with the actual possession of a proof of α. The intuitionist may be seen as a case in point, and in a moment I myself shall explicitly endorse this view. At this point Dummett, assuming that a consequence of the (T) schema is that the following principle is valid:

(12) $T\neg\alpha$ if and only if $\neg T\alpha$,

remarks that (12) is invalid when truth is equated to the actual possession of a proof, since from the fact that one does not possess a proof of $T\alpha$ it does not follow that one possesses a proof of $T\neg\alpha$; he concludes, by contraposing, that the (T) schema is not valid. Here is the derivation of (12) from the (T) schema:

(13) (i) $T\neg\alpha$ iff $\neg\alpha$ [from (11), replacing α with $\neg\alpha$]

(ii) $\neg\alpha$ iff $\neg T\alpha$ [from (11), by contraposition]

(iii) $T\neg\alpha$ iff $\neg T\alpha$ [from (i) and (ii), by transitivity].

What is not clear in the passage just quoted is what Dummett means when he says that «truth, so understood, does not commute with negation». A first possibility is that he means that (12) does not hold. But this interpretation does not square with what Dummett says in this other passage from *Elements of Intuitionism*:

> If we regard a mathematical statement as becoming true only when it is proved, then the predicate "... is true" is significantly tensed; the stament "π is transcendental", for example, has been true since 1882 and was not true before that; and, for that reason, when "... is true" is understood in this way, a mathematical statement A will not be equivalent to "It is true that A", and an attribution of truth-value to a mathematical statement will not itself be a mathematical statement. [...] the 'not' which occurs in 'is not true' or 'was not true' [...] is an emprirical type of negation, not the negation that occurs in statements of intuitionistic mathematics. (Dummett, 2000: 234–235)

I agree that, if mathematical statements are significantly tensed and truth is identified with knowledge, then the 'not' which occurs in 'is not true' is an empirical kind of negation. If we denote empirical negation with " ~ ", this means that the formalization of "It is not true that α" is " $\sim T\alpha$", while the intuitionistic formalization of "α is false" is, when α is a mathematical sentence, "$T\neg\alpha$"; but then Dummett's remark that truth, understood as present possession of a proof, does not commute with negation is unjustified, since ~ and $\neg$ are not the same negation! More precisely: in order to test whether truth commutes with negation, what should be verified is the intuitionistic validity either of the biconditional

(14) $\neg T\alpha \longleftrightarrow T\neg\alpha$,

or of the biconditional

(15) $\sim T\alpha \longleftrightarrow T \sim \alpha$

(where " ~ " denotes empirical negation), *not* the validity of the biconditional

(16) $\sim T\alpha \longleftrightarrow T\neg\alpha$.

On the other hand, it would be meaningless to verify the validity of (15), because it is meaningless, for an intuitionist, to apply empirical negation to a mathematical sentence; as for (14), it is intuitionistically valid, as we will see in Sect. 9.2.4.

The only biconditional that is not intuitionistically valid is precisely (16); its invalidity is presupposed by Heyting when he explains the difference between "not valid" and "false" from an intuitionistic standpoint:

> The difference between "not valid" and "false" can be clarified as follows. By "false" we mean the mathematical negation based on contradiction [...]; by "not valid" we mean the negation of ordinary speech, which does not at all imply a contradiction. In mathematical statements only the former occurs, but in statements about mathematics the latter cannot be avoided. (Heyting, 1958: 108)[10]

Hence, the passage from Dummett (1991) can only be understood as referring to the invalidity of (16); but—I repeat—that does not mean that truth, understood as present possession of a proof, does not commute with negation. In conclusion, Dummett's argument against the use of the (T) schema as a criterion for being a truth-operator (or a truth-predicate) is not convincing.

9.2.3 Internal and Intuitive Truth

The next question to consider is whether the validity of the (T) Schema picks out a unique notion of truth. Tarski seems to hold that it does. In Tarski (1944) he expresses the conviction that the material adequacy condition imposed onto the definition of truth is capable to select the classical Aristotelian notion of truth as correspondence. The conviction is not explicitly stated, but it can be inferred from the following facts:

(i) In Sect. I.3. Tarski expresses an intention:

> We should like our definition to do justice to the intuitions which adhere to the *classical Aristotelian conception of truth* [...] we could perhaps express this conception by means of the familiar formula:
>
> *The truth of a sentence is its agreement with (or correspondence to) reality.* (Tarski, 1944: 342–343)

[10] This passage has been quoted also in Chap. 5, Sect. 5.2.2.

(ii) In Sect. I.4. the same intention is made precise by requiring that the definition satisfies the material adequacy condition. Hence, the material adequacy condition 'does justice' to the intuitive notion of truth as correspondence. The question whether the intuitive notion of truth as correspondence is bivalent is not explicitly addressed by Tarski; however, an affirmative answer is suggested by the fact that in Tarski (1935) he derives the principle of bivalence from the (materially adequate) definition of truth (Tarski 1935: 197–198).[11] The derivation crucially uses Excluded Middle, as is made clear by the following steps:

(17) (i) $\alpha + - \alpha$ [the law of Excluded Middle]

(ii) $T\alpha \equiv \alpha$ [the (T) Schema]

(iii) $T\alpha + T - \alpha$ [from (i) and (ii) by Replacement].

So, Tarski's conviction is correct only under the premise that the metalanguage is associated to a metatheory whose semantics validates Excluded Middle.

On the other hand, Göran Sundholm has remarked that

> The definition of Tarski's truth-predicate is 'logically neutral'. By this I mean that Tarski's recursive definition, on its own, is not enough to yield the *Principle of Bivalence*
>
> $\forall x(\text{Sentence}(x) \rightarrow (\text{True}_L(x) \vee \neg \text{True}_L(x)))$.
>
> On the contrary, the extension taken by the predicate True_L depends on what logical inferences are allowed in the meta-theory in which the Tarskian definition has been framed. If they be classical, bivalence *will* hold, and otherwise not. More precisely, classical logic is not needed for the derivation of (the instances of) the familiar 'T-schema':
>
> $(\text{True}_L(\underline{\varphi}) \Leftrightarrow \varphi)$ is true. (Sundholm, 2004: 403)

It is therefore possible, and necessary, to introduce a clear distinction between the condition under which an operator is a truth operator and the condition under which an operator reflects our realistic intuitions about the notion of truth; the former consists in the validity of the (T) schema,[12] the latter may be epitomized into the slogan of truth as correspondence and consequently into the validity of the law of bivalence. Any notion satisfying the former condition is capable to play (at least some of) the *roles* of the notion of truth; truth as correspondence constitutes our predominant common-sense notion of truth. There is therefore a variety of theoretical notions of truth, which I shall call "internal" to stress the fact that each of them is capable to play the, or at least some of the, conceptual roles of the notion of truth within the framework of the related theory of meaning and of the formal semantics that adopts it. In this terminology we can say that bivalent truth is both the predominant intuitive notion of truth and the internal truth notion of classical logic; and that, besides it, there are several other internal notions of truth, among which existence of a proof/verification, but also actual possession of a proof/justification.

[11] The principle of bivalence is called by Tarski "The principle of excluded middle".

[12] The claim that an operator O is a truth operator iff it satisfies the schema (11) should not be confounded with the minimalist claim that (11) is a good *definition* of the meaning of O. The former claim is perfectly compatible with the idea, embraced above, that the validity of (11) is not the definition, but the material adequacy condition of the definition, of O.

Concluding, it is true both (i) that the validity of the (T) schema expresses our essential intuition about the notion of truth, and (ii) that our most common intuitive notion of truth is realistic; but the reason why (ii) holds is bivalence, not the (T) schema: the validity of the (T) schema is neutral among different intuitive notions of truth. In this connection it is interesting to note that, as Sundholm has shown,

> also Heyting's proof-explanations are completely neutral with respect to the induced logic for the notion of truth. In particular, these explanations in no way force constructivist logic upon us. Indeed, if we are prepared to employ also non-constructive means of reasoning in the meta-theory, then all instances of the Law of Bivalence
>
> $(A \vee \neg A)$ is true
>
> are readily made true by suitable proof-objects. (Sundholm, 2004: 404–405)[13]

9.2.4 Knowledge as Intuitionistic Internal Truth

We have seen in Chap. 6, Sect. 6.3.1, that according to many neo-verificationists truth should be conceived as atemporal existence of a proof; in the terminology just introduced, we can say that atemporal existence of a proof is the internal notion of truth of many neo-verificationist theories of meaning; I have discussed it in the same chapter, and I will examine below the impact of the Paradox of Knowability on it. Here I want to articulate the notion of truth internal to intuitionism outlined in Chap. 6, according to which truth *is* knowledge.

Let us adopt a metatheory in which the logical constants are read according to either Definition 1 or Definition 2. Define the meaning of T in the following way:

Definition 5

$$T\alpha =_{def} K\alpha,$$

where the meaning of K is defined either by Definition 3 or by Definition 4. In order to show that Definition 5 is materially adequate, in the sense that all the equivalences.

(18) $\alpha \longleftrightarrow T\alpha$

are logical consequences of it, it is sufficient to prove the validity of the biconditional.

(19) $\alpha \longleftrightarrow K\alpha$,

by induction on the logical complexity of α; again, I shall make reference only to Definition 2.

- α is atomic and $\neq \bot$.

The left-to-right half of (19) is (2′), and its validity has been shown in Sect. 9.2.1. As to the right-to-left half, i.e.

---

[13] See also Sato (1997), Troelstra and van Dalen (1988: 9, 32–33).

(20) $K\alpha \to \alpha$,

we must show that there is a function g associating to every justification σ for $K\alpha$ a justification $g(\sigma)$ for α. If σ' is a justification for $K(\alpha)$, a checking c has been performed in σ' that σ is a justification for α; as checking is factive,[14] σ is a justification for α, and we can define $g(\sigma') = \sigma$.

- α is $\bot$.

Then α may have some evidential factor *, i.e. a contradiction; since contradictions are transparent and checking of the presence of contradictions is factive, α and $K\alpha$ have the same evidential factors, and functions f and g from justifications for α to justifications for $K\alpha$ and viceversa can be the same as for other atomic sentences.

- α is $\beta \wedge \gamma$.

We must show that there is a cognitive state in which one can compute a function f transforming every justification for $\beta \wedge \gamma$ into a justification for $K(\beta \wedge \gamma)$; and a cognitive state in which one can compute a function g transforming every justification for $K(\beta \wedge \gamma)$ into a justification for $\beta \wedge \gamma$. If σ is a justification for $\beta \wedge \gamma$, there is a pair $\prec$b, c$\succ$ such that b $\vDash_{\sigma}\beta$ and c $\vDash_{\sigma}\gamma$; since $\prec$b, c$\succ$ is transparent, there is a checking c that σ is a justification for $\beta \wedge \gamma$, hence a cognitive state $\sigma' \geq \sigma$ in which c has been performed and c $\vDash_{\sigma'}\beta\wedge\gamma$ is a justification for $K(\beta \wedge \gamma)$, and we can define $f(\sigma) = \sigma'$. Conversely, if σ' is a justification for $K(\beta \wedge \gamma)$, a checking c has been performed in σ' that σ is a justification for $\beta \wedge \gamma$; since c is factive, there is a pair $\prec$b, c$\succ$ such that b $\vDash_{\sigma}\beta$ and c $\vDash_{\sigma}\gamma$, hence σ is a justification for $\beta \wedge \gamma$, and we can define $g(\sigma') = \sigma$.

- α is $\beta \vee \gamma$.

If σ is a justification for $\beta \vee \gamma$, a procedure p is known whose execution yields an a such that, for every $\sigma' \geq \sigma$, either a $\vDash_{\sigma'}\beta$ or a $\vDash_{\sigma'}\gamma$. Since p is transparent, there is a checking c that σ is a justification for $\beta \vee \gamma$, hence a cognitive state $\sigma' \geq \sigma$ in which c has been performed and c $\vDash_{\sigma'}\beta \vee \gamma$ is a justification for $K(\beta \vee \gamma)$, and we can define $f(\sigma) = \sigma'$. Conversely, if σ' is a justification for $K(\beta \vee \gamma)$, a checking c has been performed in σ' that σ is a justification for $\beta \vee \gamma$; since c is factive, a procedure p is known whose execution yields an a such that, for every $\sigma' \geq \sigma$, either a $\vDash_{\sigma'}\beta$ or a $\vDash_{\sigma'}\gamma$; hence σ is a justification for $\beta \vee \gamma$, and we can define $g(\sigma') = \sigma$.

- α is $\beta \to \gamma$.

If σ is a justification for $(\beta \to \gamma)$, one can compute a function h such that if b $\vDash_{\sigma}\beta$ then h(b) $\vDash_{\sigma}\gamma$; since h is transparent, there is a checking c that σ is a justification for $\beta \to \gamma$; hence there is a cognitive state $\sigma' \geq \sigma$ in which c has been performed and c $\vDash_{\sigma'}\beta \to \gamma$; σ' is therefore a justification for $K(\beta \to \gamma)$, and we can define $f(\sigma) = \sigma'$. Conversely, if σ' is a justification for $K(\beta \to \gamma)$, a checking c has been performed in σ' that σ is a justification for $\beta \to \gamma$; since c is factive, one can compute a function

[14] See Chap. 8, Remark 1 and footnote 16.

h such that if b $\vDash_\sigma \beta$ then h(b) $\vDash_\sigma \gamma$; hence σ is a justification for $\beta \rightarrow \gamma$, and we can define $g(\sigma') = \sigma$.

The other cases are analogous. In conclusion, the operator T defined through Definition 5 is a truth operator. By using this fact, it is easy to show that T, although it does not satisfy the principle of bivalence T($\alpha \vee \neg\alpha$), does commute with the logical constants, like the truth operator of classical logic. Here are some cases; the others are analogous. The case of negation is the one discussed in Sect. 9.2.2 in connection with Dummett's objection.

(21) $K(\alpha \vee \beta) \longleftrightarrow (K\alpha \vee K\beta)$.

Left-to-right:

(a)	$\alpha \rightarrow K\alpha$	[(2′)]
(b)	$\beta \rightarrow K\beta$	[(2′)]
(c)	$(\alpha \vee \beta) \rightarrow (K\alpha \vee K\beta)$	[from (a)–(b), by IPC]
(d)	$K(\alpha \vee \beta) \rightarrow (\alpha \vee \beta)$	[(20)]
(e)	$K(\alpha \vee \beta) \rightarrow (K\alpha \vee K\beta)$	[from (c)–(d), by IPC]

Right-to-left:

(a)	$K\alpha \rightarrow \alpha$	[(20)]
(b)	$K\beta \rightarrow \beta$	[(20)]
(c)	$(K\alpha \vee K\beta) \rightarrow (\alpha \vee \beta)$	[from (a)–(b), by IPC]
(d)	$(\alpha \vee \beta) \rightarrow K(\alpha \vee \beta)$	[(2′)]
(e)	$(K\alpha \vee K\beta) \rightarrow K(\alpha \vee \beta)$	[from (c)–(d), by IPC]

(22) $K(\alpha \rightarrow \beta) \longleftrightarrow (K\alpha \rightarrow K\beta)$.

Left-to-right:

(a)	$K\alpha \rightarrow \alpha$	[(20)]
(b)	$\beta \rightarrow K\beta$	[(2′)]
(c)	$(\alpha \rightarrow \beta) \rightarrow (K\alpha \rightarrow K\beta))$	[from (a)–(b), by IPC]
(d)	$K(\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \beta)$	[(20)]
(e)	$K(\alpha \rightarrow \beta) \rightarrow (K\alpha \rightarrow K\beta)$	[from (c)–(d), by IPC]

Right-to-left:

(a)	$\alpha \rightarrow K\alpha$	[(2′)]
(b)	$K\beta \rightarrow \beta$	[(20)]
(c)	$(K\alpha \rightarrow K\beta) \rightarrow (\alpha \rightarrow \beta))$	[from (a)–(b), by IPC]
(d)	$(\alpha \rightarrow \beta) \rightarrow K(\alpha \rightarrow \beta)$	[(2’)]
(e)	$(K\alpha \rightarrow K\beta) \rightarrow K(\alpha \rightarrow \beta)$	[from (c)–(d), by IPC]

(23) $K\neg\alpha \longleftrightarrow \neg K\alpha$.

Left-to-right:

(a)	$K\neg\alpha \to \neg\alpha$	[(20)]
(b)	$K\alpha \to \alpha$	[(20)]
(c)	$\neg\alpha \to \neg K\alpha$	[from (b), by IPC]
(d)	$K\neg\alpha \to \neg K\alpha$	[from (a), (c), by IPC]

Right-to-left:

(a)	$\neg\alpha \to K\neg\alpha$	[(2′)]
(b)	$\alpha \to K\alpha$	[(2′)]
(c)	$\neg K\alpha \to \neg\alpha$	[from (b), by IPC]
(d)	$\neg K\alpha \to K\neg\alpha$	[from (a), (c), by IPC]

At this point it should be clear that an intuitive sense of the principle (SV), according to which it becomes acceptable, does exist: for, if the logical constants are understood according to the intuitionistic explanation, and truth is understood according to Definition 5, then (SV) is a tautology, saying that every known statement is known. The intuitionistic view of the 'paradox' is therefore that it is not a paradox, and that (SV) is obviously valid when the logical constants are understood intuitionistically and truth as internal.

The fact that K commutes with the logical constants shows that, once the metalinguistic logical constants are intuitionistically understood, knowledge is a notion of truth «more robust than the pure disquotational notion» (Dummett, 1994: 297)[15]; and the reason for this is that that notion of truth inherits all the features of the notion of knowledge present in the intuitionistic meaning of the metalinguistic logical constants, as it happens with the notion of truth defined by Tarski, which inherits its realist nature from the meaning of the metalinguistic logical constants once the metalanguage is classically interpreted.

That knowledge is a notion of truth more robust than the pure disquotational notion means that it is capable to play the essential roles of the intuitive notions of truth. I have argued for this at many places of this book; here I would like to mention a further role, whose importance has been rightly pointed out and analyzed by P. Casalegno. I am speaking of the role of permitting the acquisition and transmission of true beliefs by means of what Casalegno calls *truthfulness attributions*, i.e. assertions of the form "What $\mathcal{S}$ says (said/will say) in such and such circumstances is (was/will be) true". Casalegno argues that truth cannot be replaced by knowledge in this role:

> The reason is that I may be in a position to know that what X asserts is true in cases in which there is no guarantee that the content of X's assertion has been or is or will ever be an object of knowledge for somebody.

[15] In the quoted passage Dummett does *not* ascribe to the intuitionist such a more robust notion.

Here is an example.

How many people went into the library between 3 p.m and 4 p.m, before Smith was murdered? Let us suppose that: (i) both I and the detective know (and I know that he knows) that whoever went into the library must have been one of the following: Ned, Tom, Ann, Susan; (ii) I know that the detective has asked Ned, Tom, Ann and Susan whether they went into the library and has convinced himself that they have told him the truth; (iii) I know that Ned and Tom have actually told him the truth, although (iv) I have no idea whether they went into the library; (v) I know that Ann told the detective that she went into the library whereas in fact she didn't; (vi) I know that Susan told the detective that she did not go into the library whereas in fact she did. It is easy to see that, in the situation just described, the detective's belief as to the number of people who went into the library is true. Therefore, if in addition I know that the detective is willing to disclose his mind to you, I am in a position to tell you: "When you ask the detective how many people went into the library, what he will tell you will be true". But I cannot convey the same information by saying that what the detective will tell you has been or is or will be an object of knowledge. By (iv) I do not know at present how many people went into the library, and perhaps I will never know (perhaps the murderer has just chosen me as his next victim). The detective's true belief does not amount to knowledge, for by (v) and (vi) it is based on false premises. The people who went into the library—we may assume—went there at different times, and none of them knows anything about the others. Finally, you would know how many people went into the library if you trusted me and if, as a consequence, you accepted what the detective will say. But, of course, you might decide not to trust me and not to accept what the detective will say! (Casalegno, 2005: 293–294)

The example is ingenious, but it doesn't seem to me to support the conclusion. First, let us compare the following four sentences:

(a) When you ask the detective how many people went into the library, what he will tell you will be true;
(b) When you ask the detective how many people went into the library, what he will tell you has been or is or will be an object of knowledge;
(c1) if you trust me, when you ask the detective how many people went into the library, what he will tell you will make you know it;
(c2) when you ask the detective how many people went into the library, what he will tell you will allow/permit you to know it.

(a) is the standard truthfulness attribution; (b) the alternative suggested and criticized by Casalegno; (c_1) and (c_2) two revised alternatives I am proposing for consideration. Casalegno's example is intended to show that (a) and (b) do not convey the same information: if you decide not to trust me and not to accept what the detective will say, (b) is false, while (a) is true. However, neither (c_1) nor (c_2) present this drawback, since they are both true if you don't trust me: Casalegno's objection does not seem decisive. Therefore, his argument does support the conclusion that a notion of truth satisfying Tarski's biconditional is indispensable, but not the conclusion that such a notion must satisfy the principle of bivalence: knowledge seems to be a viable alternative.

9.2.5 *Is the Equation of Truth with Knowledge Inconsistent?*

Is the idea of equating truth with knowledge consistent? There is an argument—called by J. Murzi "The Standard Argument" (Murzi, 2010: 275)— for the conclusion that it is not. It consists in the following derivation of $\exists\alpha(\alpha \wedge \neg K\alpha)$ from the assumptions $p \vee \neg p$ and $\neg Kp \wedge \neg K\neg p$[16]:

(24)

```
            ¬Kp ∧ ¬K¬p                 ¬Kp ∧ ¬K¬p
           ––––––––––                 ––––––––––
     [p]¹        ¬Kp         [¬p]²       ¬K¬p
   –––––––––––––––––         ––––––––––––––––––
      p ∧ ¬Kp                      ¬p ∧ ¬K¬p
     –––––––––––––––––         –––––––––––––––––––
p ∨ ¬p   (p ∧ ¬Kp) ∨ (¬p ∧ ¬K¬p)   (p ∧ ¬Kp) ∨ (¬p ∧ ¬K¬p)     [p ∧ ¬Kp]³   [¬p ∧ ¬K¬p]⁴
–––––––––––––––––––––––––––––––––––––––––––––––––––––––– 1, 2   –––––––––   –––––––––
      (p ∧ ¬Kp) ∨ (¬p ∧ ¬K¬p)                          ∃α(α ∧ ¬Kα)  ∃α(α ∧ ¬Kα)
––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––– 3, 4
                         ∃α(α ∧ ¬Kα)
```

Now, if we observe that there are statements p that the intuitionist acknowledges as being decidable (i.e. such that $p \vee \neg p$ is assertible), and that, as a matter of fact, are unknown (i.e., such that $\neg Kp \wedge \neg K \neg p$ is true),[17] we obtain that $\exists\alpha(\alpha \wedge \neg K\alpha)$ is assertible.

My answer consists in observing that the argument is valid but unsound, since $\neg Kp \wedge \neg K\neg p$ is intuitionistically inconsistent. Assume that $\neg Kp \wedge \neg K\neg p$ is assertible, and reason in the following way:

(25)

```
[¬Kp ∧ ¬K¬p]¹   [p]²   [¬Kp ∧ ¬K¬p]¹   [¬p]³
–––––––––––     –––    ––––––––––––    –––
    ¬Kp          Kp       ¬K¬p          K¬p
–––––––––––––––––––––   ––––––––––––––––––––
     ⊥                        ⊥
–––––––––––––––––– 2    ––––––––––––––––––– 3
     ¬p                      ¬¬p
––––––––––––––––––––––––––––––––––––––––––––––
```

[16] Keeping present Dummett's remark, quoted above, that «the 'not' which occurs in 'is not true' […] is an emprirical type of negation», one might wonder whether the second assumption should not be $\sim Kp \wedge \sim K \neg p$. However, notice that, as I remarked in Chap. 5, $\neg$ can be applied also to certain classes of empirical sentences; since in Murzi (2010) "K α" reads "someone at some time knows that α", it seems appropriate to conceive the negation of K α as intuitionistic.

[17] An example is "Prime(n)", where n is some very large number.

$$\frac{\bot}{\neg(\neg Kp \wedge \neg K\neg p)}\ 1$$

The formula $\exists\alpha(\alpha \wedge \neg K\alpha)$ ⟨ may therefore be false.[18] In order to show that it is actually false, let us wonder whether there could be a proof of it, i.e. a procedure k whose execution yields, after a finite time, a pair $\prec c, \pi\succ$, where c is a proposition and π is a proof of $c \wedge \neg Kc$. A proof of $c \wedge \neg Kc$ is a pair $\prec\pi_1, \pi_2\succ$, where π_1 is a proof of c and π_2 is a proof of $\neg Kc$; such a pair cannot exist, on pain of contradiction: being presented with π_1, one can effect the observation that what one is presented with is a proof of c, thereby obtaining a proof π_3 of Kc; coupling π_3 with π_2 we obtain a proof of $Kc\wedge\neg Kc$: a contradiction; k cannot therefore exist. In conclusion, the intuitionist cannot assert $\exists\alpha(\alpha \wedge \neg K\alpha)$ ⟨, and the idea of sentences that, being unknown, are not yet true nor false is not inconsistent.

The intuitionistic inconsistency of $\neg Kp \wedge \neg K\neg p$ may sound unacceptable from the intuitive standpoint, since it seems to conflict with the idea, which also an intuitionist should accept, that there are undecided, hence unknown, statements. Here it is important, again, to pay attention to the intuitionistic meaning of the logical constants, in particular of negation. The assertibility of $\neg(\neg K\alpha \wedge \neg K \neg\alpha)$ means that a method is known to transform every proof of $\neg K\alpha \wedge \neg K \neg\alpha$ into a contradiction, hence that a logical obstacle is known to the possibility that there is a *proof* of $\neg K\alpha \wedge \neg K \neg\alpha$; it does not exclude the *fact* that neither α nor $\neg\alpha$ are known. We will see in a moment whether the existence of such a fact can be acknowledged within the intuitionistic conceptual framework. Before, I want to comment upon the existence of a logical obstacle to the possibility that there is a *proof* of $\neg K\alpha \wedge \neg K \neg\alpha$. This is neither unacceptable nor unexpected if we keep present that the operator K is, in intuitionistic logic, a truth-operator; for it is a principle valid in general, i.e. for *every* internal notion of truth, that the formula expressing the proposition "p is neither true nor false" is inconsistent. Take for instance the formula $-T\alpha\ \&\ -T-\alpha$, expressing the same proposition within classical logic, and reason exactly in the same way as in (25), simply replacing $\neg$ with $-$, and $\wedge$ with &. The crucial step is the inference of $T\alpha$ from α; in other terms, the inconsistency of the formula expressing the proposition "α is neither true nor false" depends on the validity of the principle $\alpha \to T\alpha$ (together with propositional laws that are common to classical and intuitionistic logic). We have seen that the reason why that principle is intuitionistically valid is the assumption that proofs are epistemically transparent; of course this very assumption may be questioned, but the issue of its truth or falsity is utterly different from the question whether there are intuitive truths that, as a matter of fact, are unknown.

[18] As noted by an anonymous referee, this is not unsurprising, since in (25) the principle (2′) is used, which is itself inconsistent with the conclusion of (24). What my answer stresses is only that, if we reason in an intuitionistic metatheory, the conclusion of the Standard Argument is not true.

9.2.6 Unknown Statements

I have said that the assertibility of $\neg(\neg K\alpha \wedge \neg K \neg\alpha)$ does not exclude the existence of the *fact* that neither α nor $\neg\alpha$ are known. Can the intuitionist *assert* the existence of such a fact? I think not, and in this section I shall try to motivate this opinion.

Let me remind the reader that "Kα", when it is understood according to Definition 3 or 4, i.e. as expressing actual possession of a proof or of a truth-ground, is an *empirical* statement, and that its negation is empirical negation, as stressed by both Heyting and Dummett. I have proposed in Chap. 5 to identify empirical negation with Nelson's strong negation ~. So, let us consider the language $\mathcal{L}_{\sim,K}$; we must define the notions of proof of $\sim K\alpha$ and of justification for $\sim K\alpha$, according as we explain the logical constants according to Definition 1 or to Definition 2.

Here is my proposal:

Definition 6 A proof of $\sim K\alpha$ is the observation o that what one is presented with is not a proof of α.

Definition 7 A justification for $\sim K\alpha$ is a cognitive state σ in which one has checked that σ is not a justification for α.

One might wonder whether Dummett's remark—that intuitionistic truth, when it is equated with the actual possession of a proof, does not commute with negation—is correct when it is understood as referring to strong negation, thereby causing trouble to the idea that the validity of the (T) schema is a criterion for being a truth operator. For example, the observation that what one is presented with is not a proof that it is raining is not the same thing as the observation that what one is presented with is a proof that it is not raining. However, it should be remembered that, as noted in Chap. 5, $\sim$ is not a logical constant, but a notion of (empirical) falsity; we can therefore accept Dummett's remark without giving up the (T) schema as a criterion. Moreover, if $\sim$ were considered as a logical constant, the argument (13) would no longer be valid: the second step is an application of contraposition, and contraposition is not valid for strong negation. As a consequence, the fact that strong negation does not commute with truth would not entail the invalidity of the (T) schema.

It seems to me that, if the existence of undecided statements can be expressed at all in an intuitionistic language, it should be expressible in $\mathcal{L}_{IPL\sim,K}$. Take for instance g, Goldbach's Conjecture: for the statement

(26) Goldbach's conjecture is undecided (unknown)

the following formula seems to be the only plausible formalization in $\mathcal{L}_{IPL\sim,K}$:

(27) $\sim Kg \wedge \sim K\neg g$.

Williamson argues against this formalization:

> if $\sim$ is to count intuitionistically as any sort of negation at all, $\sim$A should at least be inconsistent with A in the ordinary intuitionistic sense. (Williamson, 1994: 139)

In other words, the schema

(28) $\sim \alpha \rightarrow \neg\alpha$;

should be valid; then, from (27) one could derive $\neg\mathrm{Kg} \wedge \neg\mathrm{K}\,\neg\mathrm{g}$, which, by (2) and (21), is equivalent to $\neg\mathrm{g}\wedge\neg\neg\mathrm{g}$: a contradiction. However, the assumption that (28) is valid *for all* α *of* $\mathcal{L}_{\mathrm{IPL}}$ ~, K seems to be a sort of *petitio principii*, since, on the one hand, it almost amounts to assuming what one wants to conclude, i.e. that (27) is inconsistent, and, on the other hand, the motivation for it seems insufficient. Notice that (28) *is* valid for all α belonging to $\mathcal{L}_{\mathrm{IPL}\sim}$ (Gurevich, 1977), so, according to Williamson's criterion, $\sim$ does count intuitionistically as a sort of negation; the possible invalidity of (28) when α contains occurrences of K can therefore be imputed to the interplay between the meanings of K and $\sim$. On the other hand, (28) is clearly invalid when α contains occurrences of K. Consider the instance

(29) $\sim \mathrm{Kg} \rightarrow \neg\mathrm{Kg}$:

it asserts the existence of a function *f* associating to every proof of the antecedent a proof of the consequent; a proof of the antecedent is the observation o that what one is presented with is not a proof of g, and of course there is no way of transforming o into a proof of $\neg$Kg, i.e. into a method to transform every observation that what one is presented with is a proof of α into a contradiction.

If we look at the interplay between intuitionistic logical constants, strong negation and K from the standpoint of Kripke semantics, the intuitionistic assertibility of (27) seems to be out of the question. A *Kripke model* for $\mathcal{L}_{\mathrm{IPL}\sim}$ is a quadruple $\mathcal{M} = \prec \mathrm{W}, \leq, \mathrm{D}, \mathrm{V} \succ$, where W is a non-empty set (of nodes), $\leq$ is a reflexive partial order on W, V is a partial function from atomic formulas and nodes to {0,1} satisfying the following conditions:

- if $\mathrm{V}(\mathrm{p}, \mathrm{w}) = 0$ and $\mathrm{w} \leq \mathrm{w}'$, then $\mathrm{V}(\mathrm{p}, \mathrm{w}') = 0$; if $\mathrm{V}(\mathrm{p}, \mathrm{w}) = 1$ and $\mathrm{w} \leq \mathrm{w}'$, then $\mathrm{V}(\mathrm{p}, \mathrm{w}') = 1$ (monotonicity);
- For every $\mathrm{w} \in \mathrm{W}$, $\mathrm{V}(\bot, \mathrm{w}) = 0$; for every $\mathrm{w} \in \mathrm{W}$, $\mathrm{V}(\sim\bot, \mathrm{w}) = 1$.

The notion $\models_{\mathrm{w}}\alpha$ (α is true at w) is defined by induction on α as follows:

Definition 8
$\models_{\mathrm{w}}\mathrm{p}$ iff $\mathrm{V}(\mathrm{p}, \mathrm{w}) = 1$
$\models_{\mathrm{w}}\sim\mathrm{p}$ iff $\mathrm{V}(\mathrm{p}, \mathrm{w}) = 0$
$\models_{\mathrm{w}}\sim\bot$
$\models_{\mathrm{w}}\alpha \wedge \beta$ iff $\models_{\mathrm{w}}\alpha$ and $\models_{\mathrm{w}}\beta$
$\models_{\mathrm{w}}\sim(\alpha \wedge \beta)$ iff $\models_{\mathrm{w}}\sim\alpha$ or $\models_{\mathrm{w}}\sim\beta$
$\models_{\mathrm{w}}\alpha \vee \beta$ iff $\models_{\mathrm{w}}\alpha$ or $\models_{\mathrm{w}}\beta$
$\models_{\mathrm{w}}\sim(\alpha \vee \beta)$ iff $\models_{\mathrm{w}}\sim\alpha$ and $\models_{\mathrm{w}}\sim\beta$
$\models_{\mathrm{w}}\alpha \rightarrow \beta$ iff, for every $\mathrm{w}' \geq \mathrm{w}$, either not $\models_{\mathrm{w}'}\alpha$ or $\models_{\mathrm{w}'}\beta$
$\models_{\mathrm{w}}\sim(\alpha \rightarrow \beta)$ iff $\models_{\mathrm{w}}\alpha$ and $\models_{\mathrm{w}}\sim\beta$

Now, if we add the operator K to $\mathcal{L}_{\mathrm{IPL}}$~, the only definition I can see that is faithful to Definitions 6 and 7 is the following:

(30) $\vDash_w K\alpha$ iff $\vDash_w \alpha$
$\vDash_w \sim K\alpha$ iff not $\vDash_w \alpha$.

Call any model of $\mathcal{L}_{IPL\sim}$ in which these clauses hold a model of $\mathcal{L}_{IPL\sim}$, K. Now, let $\mathcal{M}$ be a model of $\mathcal{L}_{IPL\sim}$, K such that $W = \{w, w_1, w_2\}$, $w \leq w_1$, $w \leq w_2$, and V(g, w) is undefined, $V(g, w_1) = 1$ and $V(g, w_2) = 0$. In $\mathcal{M} \vDash_w \sim Kg \wedge \sim K\neg g$, i.e. $\vDash_w \sim(Kg \vee K\neg g)$[19]; but the constraint of monotonicity is not met, since not $\vDash_{w1} \sim Kg$ and not $\vDash_{w2} \sim K\neg g$. This strongly suggests that the existence of undecided statements cannot be expressed at all in an intuitionistic language: absence of knowledge decreases as knowledge increases.

9.3 Neo-Verificationist Approaches

Does the paradox of knowability threaten the neo-verificationist, who normally equates truth with knowability rather than with knowledge?

The preliminary question is how to conceive knowability from a neo-verificationist point of view. There are different approaches, hence different ways of tackling the paradox.

Within the framework of Prawitz's theory of meaning, C. Cozzo has suggested (Cozzo, 1994b) to identify the knowability of α with the atemporal existence of a proof of α, and to formalize this idea through the equivalence

(31) $\alpha \longleftrightarrow \exists\sigma(\text{proves}(\sigma, \alpha))$[20];

the proposed formalization of (K) is therefore the left-to-right half of (31):

(1′) $\alpha \rightarrow \exists\sigma(\text{proves}(\sigma, \alpha))$.

Cozzo's way out consists in remarking that from the assumption

(32) $q \wedge \neg Kq$

we may infer, by (1′),

(33) $\exists\sigma(\text{proves}(\sigma, q \wedge \neg Kq))$,

and, by distribution,

(34) $\exists\sigma(\text{proves}(\sigma, q)) \wedge \exists\sigma(\text{proves}(\sigma, \neg Kq))$,

but (34) is not a contradiction because $\exists\sigma(\text{proves}(\sigma, \neg Kq))$ does not imply $\neg\exists\sigma(\text{proves}(\sigma, q))$.

However, when the metalinguistic existential quantifier is understood intuitionistically, (33) means that there is a pair $\sigma = \prec\sigma_1, \sigma_2\succ$, where σ_1 proves q and σ_2 proves

[19] Moreover, several other intuitively false formulas are false at w; for instance $\sim Kg \rightarrow \neg Kg$; $\sim Kg \rightarrow \sim g$; $\sim Kg \rightarrow \neg g$; $\sim K\neg g \rightarrow \neg\neg g$.

[20] This is not literally Cozzo's formalization, but differences are not relevant in the present context.

$\neg$Kq. Since σ_1 proves q, and proofs are epistemically transparent, it is possible to perform the observation σ_3 that σ_1 proves q, and this observation is a proof of Kq; then

(35) proves(σ_3, Kq);

if we now construct the pair $\sigma' = \prec\sigma_3, \sigma_2\succ$, we have that

(36) proves(σ', (Kq $\wedge$ $\neg$Kq)),

hence

(37) $\exists\sigma$(proves(σ, $\perp$));

on the other hand, the meaning of $\perp$ is characterized by saying that there is no proof of $\perp$, hence the formula

(38) $\neg\exists\sigma$(proves(σ, $\perp$))

is assertible, in contradiction with (37). Therefore, discharging assumption (32),

(39) $\neg$(q $\wedge$ $\neg$Kq).

This is not q $\rightarrow$ Kq (which does not follow intuitionistically from (39) (footnote 3 and 4), but is still an unwelcome consequence of (1$'$) for the atemporalist, for whom it is possible that q is true and unknown.

Within the framework of his intuitionistic Type Theory, Martin-Löf proposes to rephrase the principle (K) in the following way (Martin-Löf, 1998: 113):

(40) If a judgement of the form 'α is true' is correct, then the proposition α can be known to be true,

which may be rephrased in turn in the following way, if we accept the principle that a judgement of the form 'α is true' is correct if and only if the proposition α is really true:

(K$'$) If a proposition is really true, then it can be known to be true.

It might be tempting to formalize (K$'$) by (1), but this is exactly what *cannot* be done if Martin-Löf's distinction between judgements and propositions is accepted: the conditional (K$'$) is not an implication, because "α is true" is a judgement, not a proposition, and judgements cannot be combined by means of the logical constants.[21] Göran Sundholm has proposed (Sundholm, 2014: 20) to conceive it as a rule of inference, suggesting the following semi-formalization[22]:

[21] See Chap. 2.

[22] *Semi*-formalization because «the meaning-theoretical issue about the propositionality of K and $\Box$» (Ibid.), hence of an intuitionistic account of these notions, is not tackled. On Sundholm's solution see also Klev (2016: 365–367).

$$(1'')\quad \frac{\vdash \alpha \text{ is true}}{\vdash \text{it is possible to know that } \alpha \text{ is true}};$$

this rule is valid, according to him, while the implication (1), or rather the semi-formal implication

(41) α is true → it is possible to know that α is true

is not valid. Let us consider first the latter claim. The starting point is the remark that

> The BHK-explanation in general presupposes that proofs may be open, since the meaning of a logical connective is certainly independent of whether a proposition containing it is asserted to be true categorically or hypothetically. (Klev, 2016: 365)

The crucial step of Sundholm's argument is that if we assume to have an *open*, or hypothetical, proof of α, we have no guarantee that a *closed*, or actual, proof of α can be found, i.e. that α can be known:

> An assumption that x is a hypothetical proof of A does not guarantee that A true can be known, that is, that an actual proof-object a of A can be found. (Sundholm, 2014: 20–21)

The argument seems to me unconvincing for two reasons. First, I do not think that, within Heyting's conceptual framework, the distinction between closed and open proofs is meaningful, since—as I have argued in Chap. 2—the notion of proof Heyting is defining is evidential, not inferential, proof. Second, even if the difference is admitted as meaningful, it does not block the argument for the validity of (2′) (hence of (41)) I gave in Sect. 2.1; for, suppose that x is an open proof of α; since presumably also open proofs are epistemically transparent, it is possible to perform the mental act o_x of observing that x is an open proof of α; then we can define a function *f* associating to every open proof x of α the relative observation o_x (which is always possible) and propose *f* as a proof of $\alpha \to K\alpha$.

Therefore it seems to me that the reason why (41) would not be valid, were it well-formed, cannot be the distinction between assuming and asserting: the only reason to refuse (41), as well as (1), is just that they are not well-formed expressions of the language of the theory of types, because they do not respect the proposition/judgement distinction[23]; but, on the one hand, such a distinction is not present within Heyting's conceptual framewok; on the other, there are serious reasons to reject this distinction itself (Martino & Usberti, 1991).

---

[23] This conclusion is implicitly confirmed by Klev's following remark:

> In order to make good sense of the knowability principle one must distinguish [...] the notions of proposition and judgement as well as the appropriate notions of correctness corresponding to these[...]. (Klev, 2016: 366, footnote 20)

Another possible way out, within the framework of Prawitz's theory of meaning, would consist in giving up the intuitionistic idea that proofs are epistemically transparent; in this way the step from (34) to (35) is blocked. But, as I have argued in Chap. 2, the price to pay is very high: as the notion of proof, or more generally of verification, is the key notion of a neo-verificationist theory of meaning, the non-transparency of proofs/verifications would create the same difficulties the neo-verificationists impute to the realist theory of meaning because of the non-transparency of truth-conditions (essentially, the non-satisfiability of either the observability or the specifiability requirement imposed onto knowledge of meaning, as explained in Chap. 1).

A further way out has been proposed by Dummett. As a matter of fact, Dummett has tackled the paradox in two papers,[24] suggesting two different answers; since the former has been explicitly withdrawn by him,[25] I will consider only the latter. Dummett's solution consists in accepting

(56) $\alpha \rightarrow \neg\neg K\alpha$,

rejecting at the same time (2).[26] This is legitimated, firstly, by the fact that only (56), not (2), follows intuitionistically from (1)[27]; secondly, by the fact that, if one reads negation intuitionistically,

> '$\neg\neg K\alpha$' means 'There is an obstacle in principle to our being able to deny that α will ever be known', in other words 'The possibility that α will come to be known always remains open'[,]. (Dummett, 2009: 52)

which is precisely what the neo-verificationist believes to hold good for every true α.

Dummett does not explain why (2) should be rejected; he only remarks that what (2) says is «contrary to our strong intuition» (Dummett, 2009: 51). As I remarked at the beginning of this chapter, what (2) says is not contrary to our intuition if (2) is read intuitionistically; on the contrary, it certainly *is* contrary to our intuition if what it says is that either the fact that α does not obtain, or the fact that α is (or will ever be) known obtains; but this is precisely the classical reading of the implication occurring in (2). Hence, Dummett is reading classically the implication in (56), intuitionistically the double negation. Such a hybrid reading is not justified; as a consequence, Dummett's solution seems quite ad hoc.

I conclude with a more general remark about the neo-verificationist approaches. For the atemporalist neo-verificationist it is essential to keep potential truth distinct from actual truth, or, equivalently, atemporal existence of a proof from its temporal existence (i.e. from actual possession of it). We have seen that the Paradox of Knowability shows the untenability of this distinction. It seems to me that the same untenability is pointed out not only by the Paradox, but also by a careful consideration

---

[24] Dummett (2001), (2009).

[25] «I do not stand by the resolution of this paradox I proposed in "Victor's Error", a piece I wrote in a mood of irritation with the paradox of knowability.» (Dummett, 2007: 348).

[26] Dummett credits Bernhard Weiss for the idea of preferring (56) to (1) as a formalization of (K).

[27] Notice that (56) is intuitionistically equivalent to $\neg(\alpha \wedge \neg K\alpha)$.

of the intuitionistic meaning of the logical constants; more precisely, I will argue that, if there is room to distinguish truth from knowledge, then the metalinguistic logical constants are not construed according to their intuitionistic meaning; or that, by contraposition, if the metalinguistic logical constants are construed according to their intuitionistic meaning, there is no room to distinguish truth from knowledge.

A canonical passage in which Dummett introduces the debate about how to conceive truth from the neo-verificationist point of view is the following:

> If [...] we allow that a statemet is true when we possess merely a demonstration of it, then truth will not distribute over disjunction: we may possess a demonstration of A∨B without having a demonstration either of A or of B. (Dummett, 1975: 243)

Dummett is assuming, for the sake of discussion, that "possessing a demonstration" (i.e. a non-canonical proof) means that «we are aware that we have an effective means of obtaining a canonical proof» (Ibid.); under this interpretation, he says, the implication.

(57) $\text{True}(\alpha \vee \beta) \rightarrow (\text{True}\alpha \vee \text{True}\beta)$

is not valid; for example, we may be aware that we have a primality test for *n*, hence a proof of "prime(n) ∨ ¬prime(n)", without having applied the test, hence without being aware neither that we have a (non-canonical) proof of "prime(n)" nor that we have a (non-canonical) proof of "¬prime(n)".

It seems to me that Dummett is here understanding classically the metalinguistic implication in (57), in the sense that for it to be true it is required that, if we are aware at a time t that we have the primality test, then either we are aware at t that n is prime or we are aware at t that n is not prime. On the other hand, if → is understood intuitionistically, what is required is that we know at t a function *f* transforming every observation o that we have the primality test into either the observation that (we have a proof that) n is prime or the observation that (we have a proof that) n is not prime; and we do know *f* already at t, in the sense that we can define it, by cases: *f*(o)= the observation o_1 that n is prime, if the execution of the test terminates with "prime(n)"; *f*(o)= the observation o_2 that n is not prime, *f* if the execution of the test terminates with "¬prime(n)". To show that *f* exists it is sufficient to define it: it is not necessary to perform the primality test (hence to know whether n is prime or not).

Let us see more in detail how it can be shown that truth commutes over disjunction. Let us observe, first of all, that, even within the intuitionistic conceptual framework, it would be possible to suggest a definition of truth different from Definition 5. For this purpose, let us emend Definition 1 by replacing, in it, the clauses for $\alpha \vee \beta$ and $\exists x\alpha$, respectively, with the following ones:

(58) given by presenting a procedure p whose execution yields, in a finite time, either a proof of α or a proof of β.

(59) given by providing a procedure p whose execution yields, in a finite time, a d ∈ D, and a proof of $\alpha[\underline{d}/x]$.

We have seen in Chap. 2, footnote 80, that the possibility of this emendation of the intuitionistic explanation has been suggested by Dummett himself (I have incorporated it into Definition 2).

Then let us define truth in the following way:

Definition 9

$$TR\alpha =_{def} \exists x(\text{proves}(x, \alpha)),$$

and assume that the metalinguistic existential quantifier is understood according to (59): a proof of ∃x(proves(x, α)) is a procedure p whose execution yields, in a finite time, a pair ≺d, π≻, where d ∈ D[28] and π is a proof of "$\underline{d}$ proves α". Since the 'witness' d proves α, it is an intuitionistic proof of α, as defined by the (emended) Definition 1; at this point, π can only be an observation o that d is a proof of α, i.e. a proof of Kα.

Let us prove that Definition 9 is materially adequate, i.e. that there is a function *f* transforming every proof of α into a proof of ∃x(proves(x, α)), and that there is a function *g* transforming every proof of ∃x(proves(x, α)) into a proof of α. Define *f* in the following way: if π is a proof of α, *f*(π) is the procedure p consisting in effecting the observation o that π proves α and in constructing the pair ≺π, o≻. Since proofs are epistemically transparent, the observation o terminates after a finite time, and the pair ≺π, o≻ is therefore a proof of ∃x(proves(x, α)). Conversely, define *g* in the following way: if p is a procedure whose execution yields, after a finite time, a pair ≺d, π≻, where d is a construction and π is a proof of "d proves α", then stipulate that *g*(p) is π: whichever d is, π is a proof of α.

At this point it is possible to show that TR distributes over disjunction, i.e. that

(60) $TR(\alpha \vee \beta) \rightarrow (TR\alpha \vee TR\beta)$

is intuitionistically valid. Given an arbitrary proof of TR(α ∨ β), obtain, by the function *g* defined in the proof of material adequacy, a proof of α ∨ β, i.e. a procedure p such that its execution yields, in a finite time, either a proof π of α or a proof π′ of β; then define the following function *f*: if the execution of p terminates with π, *f*(p) = the observation o that π proves α; if the execution of p terminates with π′, *f*(p) = the observation o′ that π′ proves β; the pair ≺π, o≻ is a proof of TR(α), the pair ≺π′, o′≻ is a proof of TR(β); since both a proof of TRα and a proof of TRβ are, a fortiori, proofs of TRα ∨ TRβ, in both cases *f*(p) is a proof of TRα ∨ TRβ.

The gist of this small proof is that the function *f* is intuitionistically *well-defined*, even if it is a definition by cases, because the antecedent of (59) guarantees that α ∨ β is decidable; this is what is required by the intuitionistic meaning of →. In other terms: under the assumption that α ∨ β is decidable, there is (i.e. it is possible to define) a function transforming every proof of α ∨ β into a proof of α or into a proof of β.

---

[28] D is, in this case, a domain of mental constructions.

In conclusion, Dummett's objection to actual possession of proof as a truth-notion is based on a classical construal of implication, hence it is not justified, once the metalinguistic logical constants are intuitionistically construed.

Vice versa, it is for the notion of truth proposed by Dummett, and a fortiori for the notion of atemporal (potential) truth proposed by Prawitz and Martin-Löf, that there is no room within the intuitionistic conceptual framework; not in the sense that it produces inconsistencies, but that it forces a non-intuitionistic construal of the metalinguistic logical constants. Consider Dummett's proposal:

> It is therefore tempting to go one step further, and say that a statement is true provided that we are in fact in possession of a means of obtaining a canonical proof of it, whether or not we a aware of the fact. (*Ibid.*)

The truth of α is therefore not an epistemic state, but merely factual accessibility to a cognitive state in which a proof of α is possessed. As a consequence, also the relation between, for instance, having a (non-canonical) proof π of "prime(n) ∨ ¬prime(n)" and having a proof of "prime(n)" (or of "¬prime(n)") is purely factual, not epistemic; it is not a relation between cognitive states, but between states of affairs: if it is a mere *fact* that π terminates with a proof of "prime(n)", then "(prime(n) ∨ ¬prime(n)) → prime(n)" is true, otherwise it is false. At this point, "→" can only denote material implication.

As we saw in Chap. 2, the essential characteristic of intuitionistic logic, as Heyting conceives it, is its being a *logique du savoir*, opposed to classical logic as a *logique de l'être*[29]; this entails that the intuitionistic meaning of the logical constants, implication in particular, must be explained in terms of cognitive states instead of facts and relations between facts. This is the content of Heyting's *principle of positivity*: «Every mathematical or logical theorem must express the result of a mathematical construction» (Heyting, 1958: 108; see also Heyting, 1956: 231). In the case of implication, in particular, Heyting holds that, within classical logic,

> there is no room for implication as such, since every proposition is true or false, and it is not conceivable how its truth could depend on that of other propositions. (Heyting, 1956: 226)[30]

On the contrary,

> it is only natural that the proof of a proposition should depend on the proof of another proposition. (Heyting, 1956: 233)[31]

The fact that, if the metalinguistic logical constants are construed according to their intuitionistic meaning, there is no room to distinguish truth from knowledge may be more vividly highlighted by remarking that the following formula is intuitionistically valid, if the meaning of K is defined by Definition 3:

[29] «Heyting (1956) has opposed intuitionistic logic as the logic of knowledge (*logique du savoir*) to classical logic of existence (*logique de l'être*)» (Heyting, 1958: 107).

[30] «[I]l n'y a pas de place pour une implication proprement dite, car chaque proposition est vraie ou fausse, et on ne conçoit pas comment sa vérité pourrait dépendre de celle d'autres propositions.»

[31] «[I]l est tout naturel que la démonstration d'une proposition dépende de la démonstration d'une autre proposition.»

(61) $\exists\sigma(\text{proves}(\sigma, \alpha)) \longleftrightarrow K\alpha$,

since both subformulas are equivalent to the same formula α.

As a consequence, if one wanted to define a notion of ‘epistemic’ truth by means of Definition 9, one would face a dilemma: either to accept (61), whose validity follows from the fact that the biconditional is read intuitionistically, giving up the possibility of expressing the fact that there are true but unknown statements; or to insist that there are intuitionistically true but unknown statements, giving up the intuitionistic reading of the logical constants occurring in the semantical metalanguage.

The moral drawn from this dilemma by the realist is clear: there are statements that are (intuitionistically) true but unknown; hence, as shown by the paradox, there are also statements that are (intuitionistically) true but unknowable; therefore (K) must be rejected. Equally clear is the moral drawn by the intuitionist: both linguistic and metalinguistic logical constants must be read intuitionistically, hence (61) is valid. There is, however, a third answer that can be, and has been, proposed—an answer I would call “hybrid”: it consists in defining truth by Definition 9, in insisting that there are intuitionistically true but unknown statements, in giving up the intuitionistic reading of the metalinguistic logical constants, and in accepting (K). This position is instantiated by whoever accepts (K) rejecting at the same time (SV); for the reason why (SV) is judged unacceptable can only be that it is understood as expressing the thought that every α is either false or known, i.e. is understood on the basis of a non-intuitionistic construal of the implication occurring in its formalization. Among the supporters of the hybrid position there are many neo-verificationists, I surmise.

It may be interesting to wonder which is the non-intuitionistic construal of the metalinguistic logical constants the supporter of the hybrid position is implicitly appealing to. In Chap. 6, Sect. 6.3.2.1, I have argued that, when the supporter of the hybrid position is an atemporalist about truth, the principle of valence becomes intelligible and valid in its potential reading, in which it simply means that all sentences are atemporally determinate. It is therefore legitimate to conjecture that the meaning of the metalinguistic logical constants presupposed by the atemporalist is their potential meaning.

9.4 How is a Rational Discussion Possible?

An essential ingredient of the solution I have proposed is the remark that, when the logical constants are understood intuitionistically, the formalization (2′) of (SV) becomes perfectly acceptable. On the other hand, when the logical constants are understood classically, (2) is utterly unacceptable. This situation is far from surprising; on the contrary, it illustrates a general truth recalled above: the classical meaning of the logical constants is deeply different from their intuitionistic meaning. Consider for instance the schema “α or not α”: classically understood (i.e., formalized as $\alpha + -\alpha$) it expresses the intuitively true principle that every proposition is either true or false (*intuitively* true because our common-sense or pre-theoretic intuitions are undoubtedly realistic), whereas intuitionistically understood (i.e. formalized as $\alpha \vee \neg\alpha$) it expresses the intuitively false principle that every proposition is decidable in the sense that there is either a proof or a refutation of it.

However, this situation generates a serious problem: the problem whether a rational discussion between supporters of classical logic and supporters of intuitionistic logic is possible at all. How is it possible that there is real disagreement or real agreement between them, given that both disagreement and agreement about a principle presuppose that the same meaning is assigned to it by both parties, while, as we have just seen, the meaning of one and the same formula drastically changes across classical and intuitionistic construals?

It seems to me that there are at least two alternative strategies to tackle the problem.

9.4.1 The First Strategy

The first consists in placing the discussion between the two parties *before* the formalization of the intuitive notions (as the logical constants, the notion of truth, and so on) into a formal language. The discussion, in this case, concerns questions like the following:

(i) Which intuitive notions should be formalized? For instance: inclusive or exclusive disjunction? Which notion of implication? Which notion of truth?
(ii) Which intuitive notion should be chosen as the key notion of the theory of meaning, i.e. as the notion in terms of which the meaning of the expressions of the formal language (in particular of the logical constants) is to be characterized? For instance: (bivalent) truth (as the realist claims), or knowability/existence of a proof (as the neo-verificationist claims), or knowledge (as the intuitionist claims)?

In this case the problem can be solved, *provided that* each party accepts the *intelligibility* of the key notion adopted by the other party; for only in this case is a rational discussion possible: the same intuitive notions are accessible to both parties, and the disagreement concerns the legitimacy, the adequacy, the fruitfulness, etc. of adopting one notion or another as the key notion. From this standpoint, Brouwer's idea that such classical notions as bivalent truth or actual infinity are unintelligible is to be abandoned, in favour of a slightly different claim: that those classical notions, precisely because they are intelligible, turn out to be uncapable to play the foundational role the realist gives them. Of course, such a claim should be motivated by rational arguments; which means that a rational discussion would be possible. Some examples of such arguments have been given in this book: Chomsky's criticism of the externalist notion of reference (Chap. 1); Dummettian criticism of classical truth-conditions as epistemically transcendent (Chap. 1); criticism of the use of 'absolute' semantical notions for the semantics of propositional attitudes (Chap. 7); critique of the use of the classical existential quantifier to account for the ESN distinction (Chap. 7); and so on.

9.4.2 The Second Strategy

The second strategy consists in placing the discussion between the two parties *after* the formalization of the intuitive notions. In this case the problem of course arises, owing to the fact that the choice of different key notions for the theory of meaning induces differences in the meaning of the logical constants. However, there may be tactics to solve it.

I hold the first alternative is better, but I have not an a priori argument; I will argue for my thesis by considering what seems a very plausible tactics and explaining why, in my opinion, it is not viable.

The tactics is based on the idea of translating one logic into the other, analogously to the case of the translation of a language into another. As a matter of fact, there are several so-called 'translations', both of classical logic/mathematics into intuitionistic logic/mathematics—the so-called negative translations (by Kolmogorov, Gödel, Gentzen, Kuroda and others); and of intuitionistic logic/mathematics into extensions of classical logic/mathematics (Shapiro, 1985, Horsten, 1998; Artemov & Protopopescu, 2016).

I shall not enter here into a detailed discussion of this tactics. I want only to stress an obvious fact: that the so-called 'translations' are not translations at all. A translation, in general, must be correct, and it is correct if it is meaning-preserving, i.e. if, for every expression E of $\mathcal{L}$ (the language to be translated), its translation Tr(E) into $\mathcal{L}'$ has the same meaning as E (whatever meaning is). But there is no reason to believe that the 'translations' mentioned above are meaning-preserving. Consider for instance the BHK clause for implication; Shapiro himself admits that the notion of "transformations of proofs" cannot be captured in the language of Epistemic Arithmetic, and C. Smorynski has observed that the 'translation' of intuitionistic logic into epistemic logic «does not capture the full flavor of talk about methods» (Smorynski, 1991: 1497).[32] To make another example, Kuroda's negative translation is based on a simple idea: that intuitionistic double negation is a sort of 'equivalent' of classical truth; this is surely true if one aims at a faithful 'immersion' of classical logic into intuitionistic logic (i.e. at a representation preserving theoremhood), but not if one aims at a genuine translation, for the classical truth of α, expressed by its occurrence within any formula, is something very different from the existence of an obstacle in principle to our being able to deny that α, expressed by $\neg\neg\alpha$. Moreover, there seems to be a conceptual reason for the impossibility of a genuine translation of one logic into another: on the one hand, a translation is correct only if it is meaning-preserving; on the other hand, classical logic explains the meaning of the logical constants in terms of a notion (bivalent truth) the intuitionist considers unintelligible or illegitimate, and also the converse is true (the classicist finds mysterious the intuitionistic notion of general method or effective function): so it seems unlikely that one of them finds in one's own language an expression with the same meaning of an expression of the other's language.

[32] Cp. Horsten (1998), 9.

In this context many of Dummett's remarks about classical and intuitionistic implication are valuable and convincing; here is an example:

> In some very vague intuitive sense one might say that the intuitionistic connective → was stronger than the classical →. This does not mean that the intuitionistic statement A→B is stronger than the classical A→B, for, intuitively the antecedent of the intuitionistic conditional is also stronger. The classical antecedent is that A is *true*, irrespective of whether we can recognize it as such or not. Intuitionistically, this is unintellegible: the intuitionistic antecedent is that A is (intuitionistically) *provable*, and this is a stronger assumption. We have to show that we could prove B on the supposition, not merely that A happens to be the case (an intuitionistically meaningless supposition), but that we have been given a *proof* of A. Hence intuitionistic A→B and classical A→B are in principle *incomparable* in respect of strength. (Dummett, 2000: 11)

It is therefore difficult to accept the following idea, expressed by Dummett himself in "The Philosophical Basis of Intuitionistic Logic":

> the desire to express the condition for the intuitionistic truth of a mathematical statement in terms which do not presuppose an understanding of the intuitionistic logical constants as used within mathematical statements is entirely licit. Indeed, if it were impossible to do so, intuitionists would have no way of conveying to platonist mathematicians what it was that they were about [...]. That we are not in this situation is because intuitionists and platonists can find a common ground, namely statements, both mathematical and non-mathematical, which are, in the view of both, decidable, and about whose meaning there is therefore no serious dispute and which both sides agree obey a classical logic. Each party can accordingly, by use of and reference to these unproblematic statements, explain to the other what his conception of meaning is for those mathematical statements which are in dispute. Such an explanation may not be accepted as legitimate by the other side (the whole point of the intuitionist position is that undecidable mathematical statements cannot legitimately be given a meaning by laying down truth-conditions for them in the platonistic manner): but at least the conception of meaning held by each party is not wholly opaque to the other. (Dummett, 1975: 237–238)

It is difficult to see why, from the fact that the statements of a certain class are decidable for both the intuitionist and the platonist, it would follow that about their meaning there is no serious dispute: although both sides agree that they obey the excluded middle, it is still true that they give it utterly different meanings. Contrarily to what Dummett claims, it is just because the two sides give to this and to many other principles different meanings (bivalence and decidability, in the case of the excluded middle) that they accept them as valid.

Concluding, the existence of a 'neutral' common ground invoked by Dummett seems to me a myth, presumably generated by the legitimate desire to warrant the possibility of a rational discussion between the two parties.

9.5 Conclusion

The argument usually called the Paradox of Knowability is a paradox, if an intuitionistic construal of the logical constants of the language is mixed with a classical construal of the logical constants in the metalanguage; it is not, if the logical constants are construed in the same way at the two levels. As a consequence, while the argument seems to be capable to refute weak verificationism in the sense that it shows that it is not weaker than strong verificationism, it yields no reason either to refuse (K)—as the realist is willing to do—or to accept (SV)—as the intuitionist is ready to do.

The traditional 'refutation' of the intuitionistic idea that «there are no non-experienced truths» is essentially the Standard Argument discussed in Sect. 9.2.5, supplemented by some example of unknown truth:

> For example, either my office contains an even number of books at noon on 11 October 1999 (time t) or it does not. I could find out by counting whether it contains an even number of books at t. But I will not count them; nor will anyone else. As a matter of contingent fact, no one will ever know whether my office contains an even number of books at t. Thus either it is an unknown truth that my office contains an even number of books at t or it is an unknown truth that my office contains an odd number of books at t. Either way, there is an unknown truth; strong verificationism is false. (Williamson, 2000: 272)

But this is not an example of one specific unknown truth, which is impossible[33]; what can be done, and what Williamson does with his example , is

> the next best thing; we can know of two propositions that one or other of them is an unknown truth; we just cannot know which. (*Ibid*.)

However, such a "next best thing" is not good at all for the intuitionist: what is needed to convince her/him that there are unknown *truths*, as opposed to merely undecided propositions, is just to exhibit a specific one, according to the intuitionistic meaning of the existential quantifier.

Analogously, the intuitionistic argument I suggested in Sect. 9.2.1 for the validity of the formalization of (SV) cannot convince the realist, because it is based on a systematic neglect of the distinction (s)he deems essential between truth and proof.

I have tried to show that the paradox completely vanishes when the logical constants occurring in its formalization are understood according to the BHK explanation, since it is now necessary to distinguish two notions of truth: internal intuitionistic truth, which coincides with knowledge, and intuitive truth, essentially consisting in correspondence to external reality; in the former sense it is obvious that every truth is known, in the latter it is equally obvious—also for the anti-realist—that not every truth is known, and also that not every truth is knowable.

Summing up, the Paradox seems to be a powerful argument against neo-verificationism, whose weak point is the attempt to give a hybrid construal of the

[33] The reason why it is impossible for the intuitionist has been stated in the discussion of the Standard Argument. But it is impossible also for the realist, as it can be seen through the following argument, using the principles (4)–(6):

(a) $\Box(K(q \,\&{-}\, Kq) \supset (Kq \,\&\, K{-}Kq))$ by (4) (substituting q for α and $-Kq$ for β);
(b) $\Box(K{-}Kq \supset {-}\, Kq))$ by (5), substituting $-$ Kq for α;
(c) $\Box(K(q \,\&{-}\, Kq) \supset (Kq \,\&{-}\, Kq))$ from (a) and (b), by transitivity;
(d) $(K(q \,\&\, -\, Kq) \supset \bot)$ from (c), by (6);
(e) $-\, K(q \,\&{-}\, Kq)$ from (d), by PC;
(f) $-\, \Diamond K(q \,\&\, -\, Kq)$ from (e), by definition of $\Diamond$.

logical constants: intuitionistic at the linguistic level, classical or rather potential at the metalinguistic level; on the contrary, it leaves the debate between realists and intuitionists at the same point it was before its discovery. The crucial point of the debate is which notion between truth and evidence should be adopted as the key notion of the theory of meaning, or—if we accept the (in my opinion misleading) idea that meaning is to be explained in any case in terms of truth-conditions—which notion of truth, between bivalent and non-bivalent truth, the theory of meaning should be built on; in this case, the criterion for distinguishing realism from anti-realism cannot be the acceptance or refusal of the intuitive principle (K), but the acceptance or refusal of the principle of bivalence, at all levels.

References

Artemov, S., & Protopopescu, T. (2016). Intuitionistic epistemic logic. *The Review of Symbolic Logic, 9*(2), 266–298.

Brouwer, L. E. J. (1949). Consciousness, philosophy and mathematics. In *Proceedings of 10th international congress on philosophy* (pp. 1235–1249). (Now in Brouwer 1975: 480–494).

Casalegno, P. (2005). Truth and truthfulness attributions. *Proceedings of the Aristotelian Society, 105*(1), 279–304.

Cozzo, C. (1994). What can we learn from the paradox of knowability? *Topoi, 13*(2), 71–78.

Dales, H. G. & Oliveri, G. (Eds.) (1998). *Truth in mathematics.* Oxford University Press.

Dummett, M. (1975). The philosophical basis of intuitionistic logic. In H. E. Rose & J. C. Sheperdson (Eds.), *Logic Colloquium'73* (pp. 5–40). North Holland. [Now in Dummett (1978) (pp. 215–247)].

Dummett, M. (1978). *Truth and other enigmas.* Duckworth.

Dummett, M. (1991). The logical basis of metaphysics. Duckworth.

Dummett, M. (1994). Reply to Prawitz. In McGuinness & Oliveri (1994) (pp. 292–298).

Dummett, M. (2000). *Elements of intuitionism.* Clarendon Press. (First Edition 1977).

Dummett, M. (2001). Victor's error. *Analysis, 61*, 1–2.

Dummett, M. (2007). Reply to Wolfgang Künne. In R. E. Auxier & L. E. Hahn (Eds.), *The philosophy of Michael Dummett* (pp. 345–350). The Library of Living Philosophers, Open Court.

Dummett, M. (2009). Fitch's Paradox of Knowability. In Salerno (2009b) (pp. 51–52).

Gurevich, Y. (1977). Intuitionistic logic with strong negation. *Studia Logica, 36*(1/2), 49–59.

Heyting, A. (1956). La conception intuitionniste de la logique. *Les Études Philosophiques, 2*, 226–233.

Heyting, A. (1958). Intuitionism in mathematics. In R. Klibansky (Ed.), *Philosophy in the mid-century. A Survey* (pp. 101–115). La Nuova Italia.

Horsten, L. (1998). In defense of epistemic arithmetic. *Synthese, 116*, 1–25.

Klev, A. (2016). A proof-theoretic account of the miners paradox. *Theoria, 82*(4), 351–369.

Martin-Löf, P. (1991). A path from logic to metaphysics. In G. Sambin & G. Corsi (Eds.), *Atti del Congresso* Nuovi problemi della Logica e della Filosofia della Scienza (Vol. 2, pp. 141–149). CLUEB.

Martin-Löf, P. (1998). Truth and knowability: on the principles C and K of Michael Dummett. In Dales & Oliveri (1998) (pp. 105–114).

Martino, E. (2018). *Intuitionistic Proof Versus Classical Truth.* Springer.

Martino, E., & Usberti, G. (1991). Propositions and Judgements in Martin-Löf. In G. Usberti (Ed.), *Problemi fondazionali nella teoria del significato* (pp. 125–136). [Now in Martino (2018) (pp. 75–84)].

McGuinness, B. & Oliveri, G. (1994). *The philosophy of Michael Dummett*. Kluwer.

Milne, P. (1999). Tarski, truth and model theory. *Proceedings of the Aristotelian Society, XCIX*, 141–167.
Murzi, J. (2010). Knowability and bivalence: intuitionistic solutions to the Paradox of Knowability. *Philosophical Studies, 149*(2), 269–281.
Prawitz, D. (1980). Intuitionistic logic: a philosophical challenge. In G. H. von Wright (Ed.), *Logic and philosophy* (pp. 1–10). Nijhoff.
Salerno, J. (2009a). Knowability Noir: 1945-1963. In Salerno (2009b)(pp. 29-48).
Salerno, J. (Ed.). (2009b). *New essays on the knowability paradox*. Oxford University Press.
Sato, M. (1997). Classical Brouwer-Heyting-Kolmogorov interpretation. *RIMS Kokyuroku, 1021*, 28–47.
Shapiro, S. (1985). Epistemic and intuitionistic arithmetic. In S. Shapiro (Ed.), *Intensional mathematics* (pp. 11–46). Amsterdam: North-Holland.
Smorynski, C. (1991). Review of *Intensional Mathematics* by S. Shapiro. *The Journal of Symbolic Logic, 56*, 1496–1499.
Sundholm, G. (2004). The proof-explanation of logical constants is logically neutral. *Revue Internationale De Philosophie, 58*(4), 401–410.
Sundholm, G. (2014). Constructive recursive functions, Church's thesis, and Brouwer's theory of the creating subject: Afterthoughts on a Parisian joint session. In J. Dubucs & M. Bourdeau (Eds.), *Constructivity and computability in historical and philosophical perspective* (pp. 1–36). Springer.
Tarski, A. (1935). Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica*, *1*, 261–405. (English Tr. in Tarski (1983) (pp. 152-278).
Tarski, A. (1944). The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research, 4*(3), 341–376.
Tarski, A. (1983). *Logic, semantics, metamathematics: Papers from 1923 to 1938*. Ed. J. Corcoran. Hackett Publishing Company.
Troelstra, A. S., & van Dalen, D. (1988). *Constructivism in mathematics* (Vol. 1). North-Holland.
Williamson, T. (1988). Knowability and constructivism. *Philosophical Quarterly, 38*, 422–432.
Williamson, T. (1994). Never say never. *Topoi, 13*(2), 135–145.
Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

Bibliography

Boghossian, P. (1997). Analyticity. In B. Hale & C. Wright (Eds.), *A companion to the philosophy of language* (pp. 331–368). Blackwell.

Brouwer, L. E. J. (1908). De onbetruwbaarheid der logische principes. *Tijdschrift voor wijsbegeerte* 2, 152–158. (Engl. tr. The unreliability of the logical principles. In Brouwer 1975: 107–111).

Chomsky, N. (1992). *Explaining language use.* In N. Chomsky (2000) (pp. 19–45).

Dales, H. G., & Oliveri, G. (Eds.). (1998). *Truth in mathematics.* Oxford University Press.

Dummett, M. (1963). *Realism.* In Dummett (1978) (pp.145–165).

Dummett, M. (1974). The significance of Quine's indeterminacy thesis. *Synthese, 27*, 351–397. [Now in Dummett (1978) (pp. 375–419)].

Dummett, M. (1987). Reply to Dag Prawitz. In B. Taylor (Ed.), *Michael Dummett: Contributions to philosophy* (pp. 281–286). Nijhoff.

Fitch, F. (1963). A logical analysis of some value concepts. *The Journal of Symbolic Logic, 28*, 135–142.

Frege, G. (1923). *Compound thoughts.* In Frege (1984) (pp. 390–406).

Gödel, K. (1941). In what sense is intuitionistic logic constructive? Unpublished Lecture. In Gödel (1995) (pp. 189–200).

Gödel, K. (1995). *Collected works* (Vol. 3). Oxford University Press.

Heyting, A. (1947). Taal en teken in de wiskunde. *Algemeen Nederlands Tijdschrift Voor Wijsbegeerte En Psychologie, 40*, 121–131.

Heyting, A. (1955). *Les fondements des mathematiques. Intuitionnisme. Theorie de la demonstration.* Gauthier-Villars.

Heyting, A. (1968). L. E. J. Brouwer. In R. Klibanski (Ed.), *La philosophie contemporaine* (Vol. 1, pp. 308–15). La Nuova Italia.

Jackendoff, R. S. (1983). *Semantics and cognition.* The MIT Press.

Kreisel, G., et al. (1973). Perspectives in the philosophy of pure mathematics. In P. Suppes (Ed.), *Logic, methodology and philosophy of science IV* (pp. 255–278). North Holland.

Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy, 72*, 690–716.

Martin-Löf, P. (1984). *Intuitionistic type theory*. Bibliopolis.

Martin-Löf, P. (1987). Truth of a proposition, evidence of a judgement, validity of a proof. *Synthese, 73*, 407–420.

Martino, E. (2018). *Intuitionistic proof versus classical truth.* Springer.

Orenstein, A., & Kotatko, P. (Eds.). (2000). *Knowledge, language and logic: Questions for quine.* Kluwer.

Prawitz, D. (2018). To explain deduction. In M. Frauchiger (Ed.), *Truth, meaning, justification, and reality* (pp. 103–122). de Gruyter.

G. Usberti, *Meaning and Justification. An Internalist Theory of Meaning*, Logic, Epistemology, and the Unity of Science 59,
https://doi.org/10.1007/978-3-031-24605-0

Rogers, H. (1967). *Theory of recursive functions and effective computability*. McGraw-Hill.
Shapiro, S. (1985). Epistemic and intuitionistic arithmetic. In S. Shapiro (Ed.), *Intensional mathematics* (pp. 11–46). North-Holland.
Stainton, R. J. (2005). The context principle (2nd ed.). In K. Brown (Ed.), *Encyclopedia of language and linguistics*. Elsevier.
Sundholm, G. (2004). Antirealism and the roles of truth. In I. Niiniluoto, M. Sintonen, & J. Wolenski (Eds.), *Handbook of epistemology* (pp. 437–466). Kluwer Academic Publishers.

Author Index

G. Usberti, *Meaning and Justification. An Internalist Theory of Meaning*, Logic, Epistemology, and the Unity of Science 59,
https://doi.org/10.1007/978-3-031-24605-0

Subject Index

G. Usberti, *Meaning and Justification. An Internalist Theory of Meaning*, Logic, Epistemology, and the Unity of Science 59,
https://doi.org/10.1007/978-3-031-24605-0