UNIVERSITY OF SHEFFIELD

Intuitionism and Logical Revision

Julien Murzi

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

Faculty of Arts and Humanities
Department of Philosophy

Sheffield
October 2010

To my parents.
To Barbara.
To Bob and Dom.

Abstract

The topic of this thesis is logical revision: should we revise the canons of classical reasoning in favour of a weaker logic, such as intuitionistic logic? In the first part of the thesis, I consider two metaphysical arguments against the classical Law of Excluded Middle—arguments whose main premise is the metaphysical claim that truth is knowable. I argue that the first argument, the Basic Revisionary Argument, validates a parallel argument for a conclusion that is unwelcome to classicists and intuitionists alike: that the dual of the Law of Excluded Middle, the Law of Non-Contradiction, is either unknown, or both known and not known to be true. As for the second argument, the Paradox of Knowability, I offer new reasons for thinking that adopting intuitionistic logic does not go to the heart of the matter.

In the second part of the thesis, I motivate an inferentialist framework for assessing competing logics—one on which the meaning of the logical vocabulary is determined by the rules for its correct use. I defend the inferentialist account of understanding from the contention that it is inadequate in principle, and I offer reasons for thinking that the inferentialist approach to logic can help model-theorists and proof-theorists alike justify their logical choices. I then scrutinize the main meaning-theoretic principles on which the inferentialist approach to logic rests: the requirements of harmony and separability. I show that these principles are motivated by the assumption that inference rules are complete, and that the kind of completeness that is necessary for imposing separability is strictly stronger than the completeness needed for requiring harmony. This allows me to reconcile the inferentialist assumption that inference rules are complete with the inherent incompleteness of higher-order logics—an apparent tension that has sometimes been thought to undermine the entire inferentialist project.

I finally turn to the question whether the inferentialist framework is inhospitable in principle to classical logical principles. I compare three different regimentations of classical logic: two old, the multiple-conclusions and the bilateralist ones, and one new. Each of them satisfies the requirements of harmony and separability, but each of them also invokes structural principles that are not accepted by the intuitionist logician. I offer reasons for dismissing multiple-conclusions and bilateralist formalizations of logic, and I argue that we can nevertheless be in harmony with classical logic, if we are prepared to adopt classical rules for disjunction, and if we are willing to treat absurdity as a logical punctuation sign.

Acknowledgements

I am greatly indebted to my supervisors in Sheffield: Bob Hale and Dominic Gregory. Their help and support has been invaluable. They have always provided extremely helpful feedback on my work, and they have always responded, sometimes with a great deal of olympic patience, to my many queries. Bob's presence has been a constant encouragement for me. His constant and unfailing support, both philosophical and human, is behind every single bit of my work in Sheffield. Dom has acted as a second primary supervisor, always managing to be helpful, present, and sharp. I warmly thank them both. This thesis is for them.

Many thanks, too, to my previous supervisor in Rome, Cesare Cozzo. His analytical skills, as well as his philosophical rigour, have always been an example for me. Many thanks to Carrie Jenkins, who supervised my work at the time of my first visit to Arché, St Andrews, in the Spring of 2006, to Martin Smith, under whose keen direction I worked during my second visit to Arché, in the Spring of 2008, and to Stephen Read, with whom I had the pleasure to work during the Spring of 2008 and during the academic year 2009/10. Very special thanks are also due to Rosanna Keefe, Ian Rumfitt, and Stewart Shapiro for their very helpful feedback on my work, and for their constant encouragement; to Tim Williamson, for valuable correspondence and discussion, and for the comments he gave on my joint paper "The Paradox of Idealisation" at the Oxford Graduate Conference in November 2007; to Crispin Wright, whose comments at a seminar in St Andrews in July 2006 very much contributed to what I can refer to as my realistic turn; to Gabriele Usberti, for his very helpful and generous feedback and support, and to Frederik Stjernberg, who has given so many helpful comments on my early work.

Many of the subjects presented herein have been object of exceedingly stimulating discussions with many friends and colleagues: JC Beall, Colin Caret, Hartry Field, Salvatore Florio, Ole Thomassen Hjortland, Luca Incurvati, Marcus Rossberg, and Florian Steinberger. I am most grateful to all of them. I am also very much indebted to Simona Aimar, Catarina Dutilh Novaes, Richard Dietz, Dick de Jongh, John Divers, Michael Gabbay, Joe Salerno, Stewart Shapiro, and Elia Zardini, who have been kind enough to read and comment on my work, and to Völker Halbach, Lloyd Humberstone, Dag Prawitz, Greg Restall, and Neil Tennant for very helpful email exchanges.

I am very grateful for the feedback I received from audiences at conferences and talks in Rome, Lisbon, Nancy, St Andrews, Aix-en-Provence, Sheffield, Geneva, San Francisco, Leeds, Bristol, Norwich, St Andrews, Oxford, Baltimore, Cambridge, Canterbury, Hejnice, London, Amsterdam, Dubrovnik, Siena and Aberdeen. Roy Cook, Marcus Rossberg, Tim Williamson, Joe Salerno, Luca Incurvati and Peter Smith, and Dick de Jongh were the commentators in San Francisco, St Andrews, Oxford, Baltimore, Cambridge, and Amsterdam respectively. Many thanks to all of them.

Very special thanks are due to the Department of Philosophy in Sheffield: for giving me confidence, for providing such a friendly and stimulating environment, and for letting me teach a course of my own on some of the topics of this thesis; a challenging, but truly amazing, experience. I should also thank my students in Sheffield, both graduate and undergraduates, from whom I have learned a great deal. Special thanks are also due to the University of St Andrews, and to its research center, Arché, which provided an exceedingly stimulating environment for my work, in a very friendly and hospitable atmosphere.

I also wish to thank my philosopher friends and colleagues in Rome, Sheffield, St Andrews, and around the world. To mention but a very few: Aaron Bogart, Theodora Achourioti, Paolo Bonardi, Pablo Cobreiros, Colin Caret, Mike De, Roy Dyckhoff, Matteo Falomi, Graeme A. Forbes, Kate Harrington, Andrea Iacona, Frederique Janssen-Lauret, Peter Milne, Sebastiano Moruzzi, Sergi Oms, Francesco Orsi, Jonathan Payne, Kathy Puddifoot, Sara Protasi, Cristina Roadevin, Sam Roberts, Gil Sagi, Valerio Salvi, Andrea Sauchelli, Daniele Sgaravatti, Andreas Stökke, Margot Strohminger, and Richard Woodward.

I very gratefully acknowledge the very generous financial support of the University of Rome "La Sapienza", of the Royal Institute of Philosophy, of the University of Sheffield, of the Analysis Trust, and of the Petrie Watson Exhibition Fund. A final very big thank to Barbara and to my parents, whose love, understanding, and support made this thesis possible.

Julien Murzi
Sheffield, October 2010

Contents

Abstract ii

Acknowledgements iii

Contents v

1 Introduction 1

1.1 From metaphysics to logic . . . 2

1.1.1 The Basic Revisionary Argument . . . 3

1.1.2 The Paradox of Knowability . . . 4

1.2 Inferentialism and logical revision . . . 5

1.2.1 Logical inferentialism . . . 6

1.2.2 Proof-theoretic harmony . . . 7

1.2.3 Inferentialism and separability . . . 8

1.2.4 Classical inferentialism . . . 9

1.2.5 Conclusions . . . 10

I From metaphysics to logic **11**

2 The Basic Revisionary Argument 13

2.1 Dummett's challenges . . . 15

2.1.1 Some definitions . . . 15

2.1.2 Dummett's case against Semantic Realism . . . 18

2.2 Tennant and Salerno on logical revision . . . 27

2.2.1 Tennant on manifestation and logical revision . . . 27

2.2.2 Salerno on logical revision . . . 33

2.3 The Basic Revisionary Argument . . . 35

2.3.1 Brouwer's line of argument 35
2.3.2 From Brouwer to Wright 37
2.3.3 Introducing the Basic Revisionary Argument 41
2.4 How basic is the Basic Revisionary Argument? 47
2.4.1 How Basic is the Basic Revisionary Argument? 48
2.4.2 Objections and replies 49
2.4.3 Wright on epistemic modesty 51
2.4.4 Intuitionism and Dialetheism 55
2.4.5 How Basic is the Single Premise Argument? 56
2.5 Conclusions 58

3 The Paradox of Knowability **61**
3.1 The Church-Fitch argument 62
3.1.1 The proof 63
3.1.2 Possible ways out: a shopper's guide 65
3.2 Victor's anti-realism 67
3.2.1 Dummett on anti-realism and basic statements 67
3.2.2 A weaker Manifestation Requirement 68
3.2.3 Williamson on basic statements 68
3.2.4 Basic statements and logical revision 70
3.2.5 Summing up 71
3.3 A seemingly trivial way out 71
3.3.1 Proof-types and-proof tokens 72
3.3.2 Proofs as Aristotelian types 73
3.3.3 Truth and provability 74
3.4 The standard intuitionistic response 76
3.4.1 Dummett's new line 76
3.4.2 The Standard Argument 77
3.4.3 Knowability and bivalence 79
3.5 The Paradox of Idealisation 80
3.5.1 Hierarchical treatments 81
3.5.2 Strict Finitism and the Paradox of Idealisation 82
3.5.3 Objections and replies 85
3.5.4 Church-Fitch and empirical negation 88
3.6 Conclusions 90

II Intuitionism and logical inferentialism 91

4 Logical inferentialism 93
4.1 Logical inferentialism 94
4.1.1 The Determination Thesis 96
4.1.2 The Fundamental Assumption 97
4.1.3 From rules to truth-conditions 100
4.1.4 The Stipulation Thesis 101
4.2 Three arguments for logical inferentialism 103
4.2.1 Inferentialism and deductive knowledge 103
4.2.2 Modesty and full-bloodedness 106
4.2.3 Admissible models 109
4.2.4 Inferentialism and model-theory 112
4.3 Williamson's case against logical inferentialism 112
4.3.1 McGee, Peter, and Stephen 114
4.3.2 Knowledge of rules 117
4.3.3 Inferentialism and 'Boche' 119
4.3.4 Theoretical and radical revisionism 120
4.3.5 The Quine-Williamson challenge 121
4.3.6 Inferentialism and understanding 124
4.4 Prior's `tonk` 128
4.4.1 Liberal inferentialism and `tonk` 128
4.4.2 Towards a less liberal form of logical inferentialism 131
4.5 Conclusions 132

5 Proof-theoretic harmony 133
5.1 Two arguments for harmony 134
5.2 Intrinsic harmony 138
5.2.1 Making Gentzen's ideas more precise 139
5.2.2 Prawitz's Inversion Principle 140
5.2.3 Reduction steps 141
5.2.4 A necessary but not sufficient condition for harmony 143
5.3 Strong intrinsic harmony 144
5.4 General elimination harmony 146
5.5 Quantum disjunction 148
5.5.1 Harmony and the quantum rules 148
5.5.2 What is wrong with quantum disjunction 151
5.6 Harmony as full invertibility 153

5.6.1 Weir's account of harmony . . . 154
5.6.2 Weir's rules for disjunction . . . 156
5.6.3 Tonk strikes back . . . 158
5.6.4 Harmony as full invertibility . . . 159
5.7 The Argument from Harmony . . . 161
5.7.1 Minimal, intuitionistic, and classical logic . . . 161
5.7.2 Double Negation Elimination . . . 163
5.7.3 Classical *reductio* . . . 165
5.8 Conclusions . . . 167

6 Inferentialism and separability **169**
6.1 Meaning-constitutive rules . . . 171
6.1.1 Atomism, molecularism, and holism . . . 172
6.1.2 Purity, simplicity, single-endedness, and sheerness . . . 173
6.1.3 Rule-purism . . . 174
6.1.4 Dummett on I- and E-rules . . . 177
6.2 Global proof-theoretic constraints . . . 181
6.2.1 Separability and conservativeness . . . 182
6.2.2 Some conjectures . . . 186
6.3 The Argument from Separability . . . 188
6.3.1 The Argument from Analyticity . . . 188
6.3.2 Belnap's Argument for Structural Conservativeness . . . 190
6.3.3 Bendall's generalization of Belnap's argument . . . 190
6.3.4 Order, purity, and inferential packages . . . 192
6.3.5 Separability and understanding . . . 193
6.3.6 Logical holism and the possibility of logical atomism . . . 195
6.3.7 The Argument from Separability . . . 196
6.4 Conservativeness and higher-order concepts . . . 199
6.4.1 Truth and conservativeness . . . 199
6.4.2 Higher-order logics . . . 201
6.5 Inferentialism and the Categoricity Problem . . . 205
6.5.1 Carnap's problem and Raatikainen's diagnosis . . . 206
6.5.2 The intuitionist's response . . . 208
6.5.3 A Carnap-like problem for BHK semantics? . . . 209
6.5.4 Incurvati and Smith's objections . . . 211
6.5.5 Absurdity as a logical punctuation sign . . . 213
6.6 Conclusions . . . 215

Appendix: categoricity and conservativeness 216

7 Classical inferentialism **219**

7.1 Multiple conclusions 220

7.1.1 Sequent calculi 220

7.1.2 Sequent calculi and classical logic 221

7.1.3 Multiple conclusions and actual reasoning 224

7.1.4 Restall's interpretation 226

7.2 Bilateralism 228

7.2.1 Frege's razor and Restall's arguments 229

7.2.2 Bilateralism and classical logic 230

7.2.3 Minor adjustments 234

7.3 Bilateralism, denial, and external negation 236

7.3.1 Dummett's charge of incoherence 236

7.3.2 Correctness valuations and external negation 238

7.3.3 Assuming denials 241

7.4 Classical harmony 244

7.4.1 Classical disjunction and classical *reductio* 245

7.4.2 CR as a structural rule 249

7.4.3 **CPL** normalized 251

7.4.4 Categoricity again 254

7.5 Conclusions 256

8 Conclusions **257**

Appendices **267**

A Manifestability and decidability **267**

A.1 Williamson's first argument 267

A.2 Response to Williamson first argument 270

A.3 Williamson's second argument 271

A.4 Response to Williamson's second argument 273

B The Whole Discourse Argument **277**

B.1 The central inference 277

B.2 Tennant's proof of the central inference 278

B.3 A snappy proof of the Central Inference 279

C Empirical negation **281**
C.1 Kripke's semantics for intuitionistic logic 281
C.2 Empirical negation 282
C.3 Never say never 283
C.4 De Vidi and Solomon's reply 284
C.5 Two problems 285
C.6 Possible knowledge of actual truths? 286
C.6.1 All actual truths are knowable 286
C.6.2 Transworld knowledge and the Trivialization Objection . . . 287
C.6.3 Edgington's reply 291

D Higher-order logics **297**
D.1 Higher-order logics: a syntactic approach 297
D.2 Inferentialism and higher-order logics 299
D.2.1 Rejecting higher-order logics? 300
D.2.2 Higher-order logics and indefinite extensibility 301
D.2.3 The holist's response 301
D.2.4 Wright on higher-order logics 302
D.3 Understanding higher-order concepts 304

E Proof-theoretic consequence **307**
E.1 Prawitz's account of validity 307
E.2 Syntactic validity and higher-order logics 309

Index **312**
References 320

Philosophy would interest me much less if I did not think it possible for us eventually to attain generally agreed answers to the great metaphysical questions; but I should not have written this book unless I also thought that we should do better not to go at them bold-headed.

– Sir Michael Dummett, *The Logical Basis of Metaphysics*

Chapter 1

Introduction

It is a widespread belief that our logic is classical, at least in the following minimal respect: it validates the Law of Excluded Middle, that for every declarative sentence A, either A or $\neg A$, and the Law of Double Negation Elimination, that one may infer A from $\neg\neg A$. These logical principles are widely relied upon in mathematics departments, as well as in our everyday practice. Yet, the belief that the Excluded Middle and Double Negation Elimination are valid has been famously challenged by *mathematical intuitionists*, such as Jan Brouwer, and by so-called *semantic anti-realists*, such as Michael Dummett, Dag Prawitz, and Neil Tennant. In this thesis, I consider two families of arguments against classical logic: some *metaphysical* arguments, resting on the anti-realist claim that all truths are knowable, and a series of *semantic* arguments, to the effect that there is something amiss with classical logic itself, independently on one's commitment to the knowability of truth.[1] My immediate aim is to determine whether these arguments compel us to revise the classical canons of inference, even granting their most controversial premises. My more general—and ambitious—aim is to provide the bare bones of a framework for assessing disputes about the correct logic.

Since the publication of Kripke (1975), revisionary approaches to the semantic paradoxes have become dominant in the contemporary literature—see e.g. Graham Priest's *In Contradiction* (2006), Hartry Field's *Saving Truth from Paradox* (2008), and JC Beall's *Spandrels of Truth* (2009). We are invited to solve paradoxes such as the Liar Paradox

[1] The revisionary arguments considered in this thesis by no means exhaust the possible arguments for intuitionistic logic. Two of the most powerful and fascinating arguments have been left out, for reasons of space: Crispin Wright's argument from *vagueness*, and Michael Dummett's argument from *indefinite extensibility*. See e.g. Wright (2001), Wright (2003b), Wright (2007b), and Dummett (1991a, Chapter 24). I will very briefly say something about them—or, better, about their role—at the end of Chapter 8.

$$(\lambda)\ \lambda \text{ is not true}$$

and Curry's Paradox

$$(\kappa)\ \text{If } \kappa \text{ is true, then London is the capital of France}$$

by weakening the logic, thus preserving the consistency, or the non-triviality, of the systems in which they can be run. The logical principles we are asked to give up, however, are very basic ones. It is recommended, for instance, that we accept *modus ponens*

$$\to\text{-E}\ \frac{A \to B \qquad A}{B}$$

but we reject the so-called *pseudo modus ponens*:

$$(A \wedge (A \to B)) \to B.$$

The reason we are given is that, given certain assumptions, the former is consistent with Curry's Paradox, but the second is not.[2] On similar grounds, we are required to give up the standard introduction rules for negation and implication:

$$\neg\text{-I}, i\ \frac{\begin{matrix}[A]^i \\ \vdots \\ \bot\end{matrix}}{\neg A} \qquad\qquad \to\text{-I}, i\ \frac{\begin{matrix}[A]^i \\ \vdots \\ B\end{matrix}}{A \to B}$$

I submit, however, that it is hard to assess these suggestions without a background conception of what logic *is*. In this study, I make a case for the need of assessing competing logics against the backdrop of a general conception of logic, and I outline, and motivate, one such conception.

1.1 From metaphysics to logic

In the first part of the thesis, I consider two arguments to the effect that classical logic is inconsistent with the metaphysical belief that truth does not outstrip our capacity to know. I suggest that the first argument, Crispin Wright's Basic Revisionary Argument, requires that we already have reasons for thinking that the logical principles on which it relies are more acceptable than the classical principles it seeks to undermine. As for the second argument, Alonzo Church and Frederic B. Fitch's so-called Paradox of Knowability, I argue that it can only be made at work if intuitionists are able to define an empirical negation. In Appendix C, I critically examine a recent attempt to define an empirical negation in an intuitionistic language, and I conclude that it faces serious difficulties.

[2] See e.g. Field (2008) and Beall (2009).

1.1.1 The Basic Revisionary Argument

It is natural to think that every non-defective description of reality should be either correct or incorrect. Either things are as the description say they are, or they are not. Reality has no gaps. Natural though it may be, this thought is in tension with two further claims, one controversial, the other seemingly trivial. The first is the *anti-realist* claim that truth may not outstrip our capacity to know: at first approximation, that, if a statement is true, then it must be possible to know, at least in principle, that it is true. The second is the mere recognition that we do not presently know that every statement, or its negation, is knowable. Anti-realists typically resolve the tension by rejecting the Law of Excluded Middle, and, with it, the thought that reality is fully determinate. The upshot is usually taken to be the adoption of a weaker logic such as *intuitionistic logic*, where the Law of Excluded Middle does not hold; see e.g. Dummett (1973b), Prawitz (1980), Wright (1992), and Wright (2001). Crispin Wright dubs this the Basic Revisionary Argument.

Following Incurvati and Murzi (2008), I argue in Chapter 2 that the argument leads to conclusions that are unacceptable to classicists and intuitionists alike. The problem is that the uncontroversial claim

(EM) We do not presently know, of every statement, that either it or its negation is knowable

is classically equivalent to the claim

(EM′) We do not presently know, of every statement, that it is not the case that neither it nor its negation is knowable.

But what if the latter claim were taken as an assumption of the intuitionist's argument for logical revision? As I show, the upshot would be a parallel argument, call it the Basic Revisionary Argument*, to the effect that the *Law of Non-Contradiction*, that it is not the case that both A and its negation hold, for any A, must be rejected—a conclusion that neither classicist nor intuitionist logicians are prepared to accept. However, while *intuitionists* can distinguish between these two ways of expressing our epistemic modesty (in intuitionistic logic, the former intuitionistically entails the latter, but the converse implication does not hold), classicists cannot do so, since EM and EM′ are classically equivalent. It follows that the difference between EM and EM′, and, indeed, between the Basic Revisionary Argument and the Basic Revisionary Argument*, can only be conveyed to the classicist if the classicist is willing, at least temporarily, to abandon classical logic. However, it may be argued, an argument for the abandonment of classical logic should not itself require, as a precondition for its success, that classical logic be abandoned.

In the chapter, I consider a second related revisionary argument, which I attribute to Dummett and which, following Tennant, I call the Single Premise Argument. The argument is a *reductio* of the claim that the Principle of Bivalence holds, on the assumption that knowledge of meaning—i.e. understanding—must be manifestable in our linguistic practice. I argue that the argument incurs in the same problem which afflicts the Basic Revisionary Argument: it validates a parallel *reductio* of the claim that the *Principle of Contravalence*—that no statement can be both true and false—holds, on the assumption that understanding is manifestable in use.

The suggested upshot is that, pending independent reasons for maintaining the Law of Non-Contradiction and the Principle of Contravalence, the anti-realist claim that all truths are knowable may lead to conclusions that are unwelcome to classicists and intuitionists alike.

1.1.2 The Paradox of Knowability

How can anti-realists solve the problem? Chapter 3 considers whether one of the main objections to anti-realism, the so-called Paradox of Knowability, can be turned into an argument for rejecting classical logic. Oddly enough, some of the most eminent contemporary intuitionist and classical logicians—Dummett, on the one hand, and Timothy Williamson, on the other—agree on a positive answer to this question. The chapter argues that this revisionary path is fraught with difficulties—difficulties that are indirectly confirmed by Dummett's hesitation between radically different, and equally problematic, intuitionistic responses to the problem.

The Paradox of Knowability is an argument to the effect that the anti-realist claim that all truths are knowable is true only if all truths are *actually* known at some time (see Fitch, 1963). But, since the latter claim is clearly false, so is anti-realism itself. Or is it? As Williamson (1982) first pointed out, the argument is only *classically* valid: intuitionistically, it only leads to the conclusion that every truth is such that it is not the case that it will be forever unknown. Williamson (1982) and Dummett (2009) argue, among others, that this conclusion is not as bad as the classical one. I offer reasons for thinking that they are both wrong.

I first consider the relatively little discussed idea that, on an intuitionistic interpretation of the conditional, there is no Paradox of Knowability to start with.

I show that this proposal only works if proofs are thought of as tokens, and suggest that anti-realists have good reasons for conceiving of proofs as types. In my next step, I turn to Dummett's recent work on the problem, and argue that his proposed treatment of the Paradox does not succeed, even granting his (contentious) assumption that classical logic fails for statements that could have been verified at some time, but for which all the available evidence has now been lost. Finally, following Florio and Murzi (2009), I highlight the general form of the knowability paradoxes. A knowability-like paradox can be constructed for any property $\mathcal{P}$ such that there are truths that can only be known by agents who are $\mathcal{P}$, but there are no $\mathcal{P}$-agents. By way of example, I focus on the notion of an *ideal* agent, i.e. of an agent whose cognitive capacities exceed a certain threshold. Now let Q be some feasibly unknowable truth, some truth that can only be known by an ideal agent, and suppose that there are no ideal agents. Then, the conjunction $\ulcorner Q$ and there are no ideal agents$\urcorner$ cannot be known. I consider a few possible intuitionist counters, and I find them all wanting.

1.2 Inferentialism and logical revision

The problems faced by the two metaphysical arguments examined in the first part of the thesis suggest that disputes concerning the correct logic should be assessed against some background conception of logic. In the second part of the thesis, I examine in detail one such background conception—one that, it has been argued, has itself revisionary implications. In a nutshell, the basic idea is the semantic assumption that the meaning of a logical expression is fully determined by the rules for its correct use. There is nothing more to our understanding of 'and', at least as it is used in the context of mathematical proofs, than our willingness to infer according to its operational rules. In a natural deduction system, its introduction and elimination rules: that one may infer $\ulcorner A$ and $B\urcorner$ from $\ulcorner A\urcorner$ and $\ulcorner B\urcorner$, and *vice versa*. Similarly for the other connectives, and for the quantifiers. Call this view *logical inferentialism*.

It has been forcefully argued that, on an inferentialist approach to logic, only non-classical logics such as intuitionistic logic can be validated: the meaning of the *classical* logical constants cannot be justified on the basis of the rules for their use; see e.g. Dummett (1991b), Prawitz (1977), and Tennant (1997). More recently, however, it has been objected that these arguments are at best incomplete, since classical logic can be made consistent with the inferentialist approach of logic, given some non-standard assumptions concerning the way logic is to be formalized—see e.g. Read (2000), Rumfitt (2000), Milne (2002), and Restall (2005) among others. But are these non-standard formalizations of classical logic ultimately acceptable? And why should one adopt an inferentialist account of logic in the first place?

In the second part the thesis, I offer reasons for thinking that the inferentialist approach to logic offers an attractive account of the meaning, and of our understanding, of the logical expressions—one that is less problematic, and less radical, than it is usually thought. I introduce two inferentialist arguments against classical logic, the Argument from Harmony and the Argument from Separability, and I discuss in detail their semantic assumptions. I finally turn to the question whether classical logic is effectively undermined by these arguments, even conceding the inferentialist assumptions on which they rely.

1.2.1 Logical inferentialism

Chapter 4 introduces logical inferentialism, and some of the objections it faces. I suggest that the slogan that rules determine meanings can be interpreted in at least two ways.

On the *first interpretation*, meaning-constitutive rules determine meanings at least in the sense that they exhaust the grounds for asserting the complex statements they allow us to introduce. Michael Dummett and Dag Prawiz call this the Fundamental Assumption: introduction rules specify in principle necessary and sufficient conditions for asserting complex statements. Sometimes inferentialists further require that rules determine correct use in a stronger sense: all the correct uses of a constant $ must be derivable from its meaning-constitutive rules; meaning-constituive rules should be *complete*.

On the *second interpretation*, meaning-constitutive rules determine the satisfaction clauses of the logical operators, given minimal semantic assumptions. Thus, for instance, on the assumption that the introduction and elimination rules for conjunction are truth-preserving, one must be able to derive that a conjunction is true if and only if each of its conjuncts is also true. Similarly for the remaining logical operators.

I argue that the inferentialist approach to logic has an epistemological advantage over its non-inferentialist rivals. It allows us to solve some epistemic puzzles concerning deductive knowledge, and it offers the prospects of justifying some of our logical choices. I then consider some objections to logical inferentialism, with particular focus on Williamson's recent attacks to the inferentialist model of understanding—see e.g. Williamson (2003), Williamson (2006), and Williamson (2009b). In Williamson's view, the inferentialist account of understanding—that to understand $\$$ is to be willing to infer according to the rules for their correct use—is undermined by counterexamples. There are subjects, Williamson claims, who (i) understand logical expressions just like the overwhelming majority of competent speakers do, but (ii) are nevertheless unwilling to infer according to the rules for their correct use. I argue that Williamson's argument is ultimately question-begging.

1.2.2 Proof-theoretic harmony

Chapter 5 focuses on the proof-theoretic requirement of harmony—roughly, that introduction and elimination rules should be in balance with each other. It has long been known since the publication of Arthur Prior's *The runabout inference ticket* (Prior, 1960) that not *all* rules can be meaning-constitutive. Prior asks us to suppose we could define a connective, he calls it `tonk`, whose meaning-constitutive rules are: 'From $\ulcorner A \urcorner$, infer $\ulcorner A\ \texttt{tonk}\ B \urcorner$', and 'From $\ulcorner A\ \texttt{tonk}\ B \urcorner$, infer $\ulcorner B \urcorner$'. Then, anything would follow from anything—clearly an unacceptable result. What has gone wrong?

One standard diagnosis is that the introduction and elimination rules for `tonk` are out of balance. They are not in *harmony* between each other. More precisely, the elimination rule is too strong: it is not justified by the corresponding introduction. But what *is* harmony? And how to justify this requirement?

In the chapter, I introduce three different accounts of harmony: strong intrinsic harmony, general elimination harmony, and harmony as full invertibility. I argue that each of these accounts can be motivated by at least two kinds of considerations. The first is the epistemic requirement that logic alone should be *epistemically conservative*: roughly, logic alone should neither create nor allow us to lose knowledge. The second is the assumption, Dummett's Fundamental Assumption, that introduction rules specify in principle a complete set of instructions for asserting complex statements.

As we shall see, on any decent account of harmony, the rules for `tonk` are sanctioned as disharmonious, as it should be. But there is a potential drawback. In standard regimentations of classical logic, the rules for *classical negation* are also sanctioned as disharmonious. Hence, eminent inferentialists such as Dummett and Prawitz have concluded, classical negation *and* `tonk` are in equal standing: they are both incoherent, or perhaps even not meaningful. I argue, though, that this conclusion is over hasty. The argument merely shows that the *standard formalizations* of classical logic are not harmonious: we are not given an argument to the effect that classical logic cannot be given a harmonious formalization. All the same, the argument compels classical logicians to either reject the requirement of harmony, or provide us with a harmonious formalization of classical logic.

Along the way, I offer reasons for thinking, *pace* Dummett (1991b), that quantum disjunction does not constitute a problem for harmony, and I show that an account of classical harmony defended by Alan Weir is flawed.

1.2.3 Inferentialism and separability

In Chapter 6, I consider yet another inferentialist argument against classical logic. I introduce proof-theoretic constraints other than harmony, and I explore the relations between them. I focus on both *local* constraints on rules and on *global* constraints on logical systems. Our main focus will be on the twin global properties of *separability* and *conservativeness*.

A formal system is separable if every provable rule R can be proved by means of a proof every step of which is an application of one of the operational rules for the logical operators figuring in R (possibly together with structural rules). A rule introducing new vocabulary yields a conservative extension of a formal system if, roughly, everything that can be proved in the extended system but was not provable in the old system contains new vocabulary. As we shall see, it can be shown that, if

(i) basic logical rules only specify conditions for correct *assertion*,

(ii) logical arguments have at most one conclusion,

and

(iii) absurdity is a nullary logical operator,

classical logic *cannot* respect the requirements of separability and conservativeness. I call this the *Argument from Separability*.

The conclusion of this argument applies to a wide range of possible regimentations of classical logic. Hence, the argument is potentially stronger than the Argument for Harmony, which only applies to one formalization of a logic at the time. On the other hand, the Argument from Separability requires stronger assumptions. I individuate two:

(a) that basic inference rules must be *complete* in a strong sense, viz. that they must allow us to derive, or justify, *all* the intuitively correct logical uses of the expressions they introduce and eliminate;

(b) that the meanings of the logical constants can be learned independently of one another.

Both assumptions are problematic, or at least so I argue.

To begin with, *higher-order* logics—logics where we are allowed to quantify over sets and properties—are notoriously *incomplete*, at least in the following minimal sense: for every $n > 1 \in \omega$, the rules for the n^{th}-order quantifiers do not capture all of $\forall^n$'s and $\exists^n$'s correct uses. It follows that, provided we are willing to ascend high enough in the hierarchy of higher-order logic (at least up to level 3), and provided that rules are open-ended, i.e. provided that they apply to all possible extensions of the language, higher-order logics are not separable, and their rules are not conservative.

Second, the standard arguments for separability all assume the falsity of the very view they seek to undermine, viz. *logical holism*, the claim that the logical expressions cannot be learned independently of one another.

I argue that neither problem affects the inferentialist's argument against classical logic from separability and conservativeness. For one thing, I suggest, inferentialists have no reasons for assuming that higher-order logics are complete. Hence, they may consistently impose the requirements of separability and conservativeness for *complete* logics, e.g. first-order logic, but not for *incomplete* ones, e.g. higher-order logics. For another, it would seem that whether our understanding of the logical vocabulary is holistic or not may well be an empirical question, and that, for this reason, classical logicians with inferentialist sympathies had better be able to provide a separable formalization of classical logic.

In Appendix D, I sketch an inferentialist account of the meaning, and of our understanding, of the higher-order quantifiers.

In the last part of the chapter, I turn to a different objection to the inferentialist view, viz. that rules do not in general determine meanings in the sense of determining their standard satisfaction clauses. The problem was first raised by Rudolf Carnap, and was recently revived by Timothy Smiley, Ian Rumfitt, and Panu Raatikainen. I suggest that it does not affect the intuitionist inferentialist, contrary to what Raatikainen alleges, and I claim in Chapter 7 that it does not affect the classical inferentialist either.

1.2.4 Classical inferentialism

Chapter 7 considers three different formalizations of classical logic, all of which are harmonious, separable, and categorical, in the sense that the satisfaction clauses for each of the logical connectives can be derived from their meaning-constitutive rules. All three formalizations obtain the desired result by enhancing—in some way or other—the structural resources of the language.

Multiple-conclusions formalizations reject the standard assumption that arguments can have at most one conclusion, and allow rules to have *multiple* conclusions. *Bilateral* formalizations reject the assumption that basic logical rules only specify conditions of correct assertion, and countenance both rules for asserting and rules for *denying* complex statements. Formulae in the formal language are prefixed by *force signs*, indicating either assertion or denial.

In the chapter, I offer considerations that cast doubt on the viability of both multiple-conclusions and bilateral frameworks. In particular, I argue that, pending an adequate interpretation of bilateral rules involving discharge of assumptions, bilateralist formalizations require the speech act of denial to play the role of an external negation.

In the final part of the chapter, I introduce a novel harmonious and separable regimentation of classical propositional logic—one on which disjunction is given a classical interpretation, and the only significant departure from the standard formalizations is that absurdity is interpreted as a logical punctuation sign.

1.2.5 Conclusions

Chapter 8 offers some concluding remarks. I suggest that the inferentialist framework provides a background conception of logic against whose backdrop one can assess competing logics. I briefly focus on Hartry Field's proposed all purposes logic (see Field, 2008), and I show that, from an inferentialist perspective, Field's logic is found wanting on several counts. I close by reassessing the prospects for metaphysical arguments such as the Basic Revisionary Argument in the light of our discussion of the inferentialist approach to logic.

Part I

From metaphysics to logic

PAGINATED BLANK PAGES ARE SCANNED AS FOUND IN ORIGINAL THESIS

NO INFORMATION MISSING

Chapter 2

The Basic Revisionary Argument

There are many possible reasons why one might question the validity of the Law of Excluded Middle; in symbols:

$$(\mathsf{LEM})\ \forall\varphi(\varphi \vee \neg\varphi).$$

For a start, one might have qualms with the Principle of Bivalence, that every statement is either true or false

$$(\mathsf{BIV})\ \forall\varphi(\mathcal{T}\varphi \vee \mathcal{T}\neg\varphi),$$

where '$\mathcal{T}\varphi$' reads 'it is true that φ' and φ's falsity is interpreted, as usual, as the truth of φ's negation. On the further assumption that the Equivalence Thesis

$$(\mathsf{ET})\ \forall\varphi(\mathcal{T}\varphi \leftrightarrow \varphi)$$

holds good, worries about BIV directly transfer to LEM. Some such worries are familiar. To mention but a few: it might be argued, perhaps following Aristotle, that the unrestricted Principle of Bivalence is inconsistent with the view that the future is open in a way the past is not;[1] or it might be thought that abandoning BIV is the key for solving the Sorites Paradox.[2] Most recently, Hartry Field has argued that the best hope for solving the semantic paradoxes is to revert to a logic which does not validate LEM.[3] My interest in this chapter will be in some *metaphysical* reasons for rejecting both LEM and BIV. In particular, I will focus on a line of argument that, albeit having been the object of much recent philosophical discussion, has been very rarely analyzed in detail: the argument from semantic anti-realism, the claim that truth must be epistemically constrained, to the rejection of both LEM and BIV.

[1] See e.g. Aristotle's *De Interpretatione*, IX, in Aristotle (1961) and Thomason (1970).
[2] See e.g. Fine (1975) and Keefe (2000). Notice that supervaluationists question BIV, not LEM.
[3] SeeField (2008).

Perhaps surprisingly, anti-realists themselves disagree as to what the argument from anti-realism to the rejection of LEM and BIV should be. In *The Philosophical Basis of Intuitionistic Logic*, Dummett writes that "so far as I am able to see, there are just two lines of argument for repudiating classical reasoning in mathematics in favour of intuitionistic reasoning" (Dummett, 1973b, p. 216). The two lines of argument he is referring to are his celebrated Acquisition and Manifestation challenges. The first proceeds from the assumption that meanings must be learnable to the conclusion that they cannot be identified with potentially verification-transcendent truth-conditions. The second also aims at ruling out verification-transcendence, but starts from the assumption that understanding must be manifestable in use. On the face of it, it is not immediately clear why the result of either argument should compel one to reject LEM. In the *The Taming of the True*, Neil Tennant argues that Dummett's Manifestation Challenge involves gross logical mistakes and does not actually provide grounds for rejecting LEM and BIV. In a recent paper, Joe Salerno writes that "given the resources provided by [...] Dummett [...], choice of logic is not a realism-relevant feature—i.e., logical revision is not a consideration that is enjoined by one's stance on the possibility of verification transcendent truth" (Salerno, 2000, p. 212). In Salerno's view, Dummett's own argument for logical revision does not itself rely on semantic anti-realism as a premise, contrary to what Dummett—and anti-realists in general—claims.

I argue that these criticisms are off-target. In his reconstruction of Dummett's revisionary argument, Tennant omits one of Dummett's key premises, viz. that there are undecidable statements: statements that are not guaranteed to be decidable. As for Salerno's criticism, it rests on a mistaken reading of Dummett's text, or at least so I shall argue. Contrary to what Tennant and Salerno allege, I will show that one can find in Dummett's text a compelling argument from anti-realism to the rejection of LEM and BIV. The argument may be supported by Dummett's challenges, but does not need to be. It can be traced back to Jan Brouwer, and it has been more recently endorsed by Dag Prawitz and Crispin Wright. Wright was the first to give it a name: the *Basic Revisionary Argument*.[4] The first formal presentation of the argument was eventually offered by Luca Incurvati and the present author.[5]

The aim of this chapter is threefold. First, it is to respond to Tennant's and Salerno's criticisms. Second, it is to present the Basic Revisionary Argument. Third, it is to argue that, even granting its most controversial premise, the anti-realist

[4]See Wright (1992, Chapter 2) and Wright (2001, p. 65).

[5]See Incurvati and Murzi (2008).

claim that truth is epistemically constrained, the argument leads to a conclusion which is unwelcome to classicists and intuitionists alike. More specifically, I will contend that the Basic Revisionary Argument validates a parallel argument to the effect that the Law of Non-Contradiction

$$\text{(LNC)}\ \forall\varphi\neg(\varphi \wedge \neg\varphi)$$

is either not a logical law, or it is both known *and* not known.

The structure of this chapter is as follows. Section 2.1 briefly introduces Dummett's challenges. Section 2.2 focuses on Tennant's criticism of the Manifestation Challenge, as presented in Chapter 5 of his *The Taming of the True*, and on Salerno's objections to what he takes to be Dummett's main argument for logical revision. Section 2.3 argues that, *pace* Tennant and Salerno, the anti-realist's main argument for for rejecting LEM and BIV, the Basic Revisionary Argument, is valid, and has long been known to anti-realists. Section 2.4 raises a new challenge to the revisionary anti-realist, to the effect that the Basic Revisionary Argument validates a parallel argument for the rejection of LNC. Section 2.5 offers some concluding considerations. Two appendices explore some loose ends. Appendix A considers, and addresses, Timothy Williamson's contention that it is a consequence of Dummett's challenges that an epistemic notion of truth cannot play the semantic role key anti-realist figures, such as Dummett, Prawitz, and Wright, would like it to play. Appendix B briefly introduces Tennant's own revisionary argument, the Whole Discourse Argument, and argues that it in fact collapses on the Basic Revisionary Argument.

2.1 Dummett's challenges

Dummett has put forward at least *three* distinct lines of argument against Semantic Realism. We have already mentioned Dummett's so-called semantic challenges: the argument from acquisition and the argument from manifestation. The third argument is a charge of *circularity*: Dummett accuses realist theories of meaning of being hopelessly circular. Section 2.1.1 introduces some terminology. Section 2.1.2 presents, in turn, each of these extremely controversial arguments.

2.1.1 Some definitions

First off, some definitions. Following Dummett, I will define Semantic Realism as the thesis that it is metaphysically possible that there be epistemically unconstrained truths: statements that are true independently of our capacity to know that they are true. Formally:

$$(\mathsf{SR})\ \Diamond\exists\varphi(\varphi \wedge \neg\mathcal{E}\varphi),$$

where '$\mathcal{E}\varphi$' is an epistemic predicate of some sort, such as 'φ has a proof' or 'φ is knowable', and '$\Diamond$' expresses metaphysical possibility..[6] Semantic Anti-realism may then be defined as the claim that, of necessity, truth is epistemically constrained:

$$(\mathsf{EC})\ \Box\forall\varphi(\varphi \rightarrow \mathcal{E}\varphi),$$

where '$\Box$' expresses metaphysical necessity. For present purposes, we might identify '$\mathcal{E}\varphi$' with 'φ is possibly known by someone at some time'. Semantic anti-realism then becomes the thesis that all truths are knowable. I shall call this the Knowability Principle:

(KP) For all φ, if φ, then it is possible to know φ.

Following Williamson (2000, Chapter 12), I will refer to the principle's most common formalisation as *Weak Verificationism*:

$$(\mathsf{WVER})\ \forall\varphi(\varphi \rightarrow \Diamond\mathcal{K}\varphi),$$

where '$\Diamond\varphi$' and '$\mathcal{K}\varphi$' respectively read, as usual, 'it is possible that p' and 'it is known by someone at some time that φ'. More sophisticated formalisations of KP will be considered in Chapter 3.[7]

I will call a statement φ *decidable* if either it is possible (in principle) to know that it is true, or it is possible (in principle) to know that it is false. Formally:

$$(DEC)\ \Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\neg\varphi.^{8}$$

I take this to be equivalent to the more standard account of decidability in terms of the existence of a decision procedure whose application would enable us to know, in a finite amount of steps, whether φ is true or false. For on the one hand, neither φ nor its negation would be knowable, if there was no effective method for knowing their truth-values. On the other, if there is such a method, then either φ or its negation *is* knowable.[9] I will call a statement φ *undecidable* if it is *presently*

[6]These two formulations are not obviously equivalent, as we shall see in Chapter 3.
[7]Our points in this chapter carry over to those alternative formalisations.
[8]I am using capital italics to name properties and capital sans-serif to names theses.
[9]See Section 2.2.1 for a more detailed presentation of this argument.

not now known to be decidable. Formally:

$$(UND)\ \neg\mathcal{K}_n(\Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\neg\varphi),$$

where '$\mathcal{K}_n\varphi$' reads 'it is now known that φ'.[10] Dummett's examples of undecidable statements include:[11]

(1) A city will never be built on this spot;

(2) There are odd perfect numbers;

(3) Jones was brave.

Notice that undecidability, thus characterised, is an *epistemic* and *tensed* concept: whether a statement is undecidable depends on what we *now know*. Undecidable statements may cease to be undecidable, if, as it is in the case of Fermat's Theorem, they come to be known. Undecidability is therefore not to be conflated with *absolute undecidability*. Whereas a statement is undecidable if we have no guarantee that either it or its negation is knowable, a statement is absolutely undecidable if neither it nor its negation are knowable—or, equivalently, if there is no procedure for determining its truth and there is no procedure for determining its falsehood. Formally:

$$(UND^*)\ \neg\Diamond\mathcal{K}\varphi \wedge \neg\Diamond\mathcal{K}\neg\varphi.$$

Both undecidability and absolute undecidability must in turn be distinguished from *potential verification-transcendence*. A *statement* is potentially verification-transcendent if, for all we now know, it is absolutely undecidable. Formally:

$$(\Diamond UND^*)\ \neg\mathcal{K}_n\neg(\neg\Diamond\mathcal{K}\varphi \wedge \neg\Diamond\mathcal{K}\neg\varphi).$$

A *truth* is potentially verification-transcendent if, for all we know, it is unknowable. Formally:

$$(\Diamond VT)\ \neg\mathcal{K}_n\neg(\varphi \wedge \neg\Diamond\mathcal{K}\varphi).$$

I take present ignorance of φ's negation to express the epistemic possibility that φ.

Two potential concerns are worth mentioning. First, one might wonder whether UND really is what the anti-realist means, or should mean, by 'undecidable'. I

[10]See also Shieh (1998) for an argument to the effect that this is actually Dummett's notion of undecidability. In order to avoid confusions with what I will call below *absolute undecidability*, Tennant (1984, p. 84) suggests that undecidability be called *pro tempore* undecidability. This is a good suggestion, but unfortunately it has not been adopted.

[11]See Dummett (1959).

will argue in due course that it is.[12] Second, it might be thought that the foregoing characterisation of realism and anti-realism is in contrast with Dummett's well-known contention that the *Principle of Bivalence* is the hallmark of realism. As he puts it:

> It is difficult to avoid noticing that a common characteristic of realist doctrines is an insistence on the Principle of Bivalence [...] that every proposition, of the kind under dispute, is determinately either true or false. (Dummett, 1991b, p. 9)

This contrast is only apparent, however. Together with some plausible assumptions about our present epistemic situation, commitment to BIV does indeed enjoin commitment to the existence of possibly unknowable truths. We shall actually *prove* this claim when discussing Tennant's so-called Single Premise Argument, in Section 2.2.1. With these definitions in place, we can now introduce Dummett's main worries about semantic realism.

2.1.2 Dummett's case against Semantic Realism

Let us now turn to Dummett's challenges to semantic realism: the Acquisition and the Manifestation challenges, and Dummett's accusation that realist accounts of meaning are bound to be circular. I will briefly conclude by presenting some possible objections. A fuller discussion of a recent objection to Dummett's challenges, recently advanced by Timothy Williamson, can be found in Appendix A.

The Acquisition Challenge

Already in his early article *Truth*, Dummett accuses realist theories of meaning of giving an implausible account of the meanings we could have *learned* (see Dummett, 1959). The general idea is that it is difficult to see how we could have learned the meanings of undecidable statements, if these are construed along the lines of a realist, bivalent, theory of meaning. Dummett's main assumption is that what we learn, and can learn, when we learn the meaning of a statement, is how to *use* that statement. He writes:

> When we learn [...] expressions [...] what we learn to do is to make use of the statements of that language: we learn when they may be established [...], we learn from what they may be inferred and what

[12] See 2.2.1 (The Single Premise Argument revisited) below.

> may be inferred from them [...]. These things are all that we are shown when we are learning the meanings of the expressions of the language [...], because they are all that we can be shown: and, likewise, our proficiency in making the correct use of the statements and expressions of the language is all that others have from which to judge whether or not we have acquired a grasp of their meanings. Hence it can only be in the capacity to make a correct use of the statements of the language that a grasp of their meanings, and those of the symbols which they contain, can consist. (Dummett, 1973b, pp. 217)

In Dummett's view, our training in the use of the language consists in learning both (i) under what recognizable conditions statements can be asserted, and (ii) what may be legitimately be inferred from them. If these are the essential features of the use of a statement, and if all we learn when we learn the meaning of a statement is how to use that statement, knowledge of the meaning of a statement—our understanding—cannot but consist in a knowledge of the conditions under which it may be correctly asserted, together with a knowledge of what may be legitimately inferred from it. Hence, Dummett concludes, our understanding cannot consist of knowledge of potentially verification-transcendent truth-conditions. For these conditions transcend, at least potentially, the correct use of the statements to which they are supposed to attach. The challenge to the realist is to provide an epistemology of potentially verification-transcendent truth-conditions. But is Dummett's argument correct?

The literature on the Acquisition Argument is too vast to be even briefly reviewed here.[13] I will limit myself to mentioning one problem, viz. that, in the course of his argument, Dummett focuses only on *one* aspect of the use of a statement—its assertibility-conditions:

> What we learn to do is to accept the truth of certain sentences [...] or [...] the occurrence of certain conditions which we have been trained to recognize, as conclusively justifying the assertion of a given statement [...] and the truth of certain other statements, or the occurrence of certain other conditions, as conclusively justifying its denial. (Dummett, 1978b, p. 362)

However, it would seem, it is open to argue that knowledge of the meaning of a verification-transcendent statement could be given by a knowledge of what may

[13]For an overview of the literature, see e.g. Hale (1997) and Miller (2003).

legitimately inferred from it, i.e. by a knowledge of what Dummett *himself* takes to be one of the central features of the use of a statement. Pending further argument to the effect that the meaning of verification-transcendent statements cannot be acquired in *this* way, it seems fair to conclude that there is a gap in Dummett's argument. Can anti-realists do better?

The Manifestation Challenge

Dummett's second argument is perhaps the most famous. In outline, the argument proceeds from two main premises: that knowledge of truth-conditions must be manifestable in use, and that, by contrast, knowledge of realist truth-conditions cannot be so manifested. Dummett's suggested conclusion is that, unless the realist can indicate elements of our behaviour that would manifest knowledge of realist truth-conditions, the very notion of verification-transcendence should be regarded as a piece of metaphysical superstition. The main principle at work in the argument is the so-called Manifestability Principle, that differences in meaning must in principle be manifestable in differences in use. Dummett's principal reason for adopting the principle is that elements of meaning that could not be manifestable in use would have no function in communication, and would be, so to speak, idle. As Dag Prawitz puts it:

> The most general support of the [manifestability] principle is obtained by arguing that meaning has to be communicable and that communication has to be observable: to assume that there is some ingredient in the meaning of a sentence which cannot become manifest in the use made of it is to assume that part of the meaning cannot be communicated. This part of the meaning would then be irrelevant when the statement was used in communication. (Prawitz, 1977, p. 4)

If knowledge of meaning must be manifestable in use, the question arises as to *how* knowledge of the truth-conditions of undecidable sentences can be manifested, given that their truth-conditions are assumed to be potentially verification-transcendent.

Let *D* be a domain containing undecidable sentences, such as e.g. sentences about the past, or quantifications over potentially infinite totalities. Both the realist and her opponent agree that we know the meanings of the sentences in *D*, i.e. we know their truth-conditions. But what *constitutes* such a knowledge? And how can one *manifest* it? Let us begin with the first question.

One might think that knowledge of truth-conditions is constituted by one's capacity to restate them in a non-trivial and informative way. This is how we sometimes learn the meanings of new expressions, and, it might be thought, this may well be what constitutes knowledge of their meaning. However, Dummett points out, this cannot provide a *general* model of understanding. For if knowledge of an expression's meaning always involved knowledge of the meaning of some *other* expressions, one could not learn a language without already possessing one. We would then be involved in an infinite regress: knowledge of a language L would require a previous knowledge of a different language L_1, and so on. Dummett writes:

> To suppose that, in general, a knowledge of meaning consisted in verbalisable knowledge would involve an infinite regress: if a grasp of the meaning of an expression consisted, in general, in the ability to *state* its meaning, then it would be impossible for anyone to learn a language who was not already equipped with a fairly extensive language. Hence that knowledge which [...] constitutes [...] understanding must be implicit knowledge. (Dummett, 1973b, p. 217)

Knowledge of meaning, Dummett suggests, must be, in general, *implicit* knowledge—whatever that means more exactly. Thus, our *second* question becomes: how can implicit knowledge of the truth-conditions of undecidable statements be manifested, if, as the realist maintains, these truth-condition may obtain, or fail to do so, independently of our capacity to know, even in principle, that they obtain, or fail to obtain?

Dummett considers two possible ways of manifesting, in general, implicit knowledge a sentence φ's meaning:

(i) by applying a decision procedure for φ, thereby coming to know whether it is true or false;

(ii) by being disposed to recognize a (correct) argument for φ if presented with one.

In Dummett's view, both of these options prove problematic in the case of potentially verification-transcendent statements. The first option can immediately be set aside: it only applies to statements that are known to be decidable. The second option, Dummett claims, cannot account for undecidable statements, since, in this case, there may be no proof for us to recognize in the first place.

It might be objected that one can nevertheless be disposed to say, of any purported proof of a potentially verification-transcendent statement, that it is *not* a proof of that statement. However, Dummett could in turn retort that such discriminating abilities would not be discriminating enough: no observable behavior would be exhibited which manifests understanding of a *specific* potentially verification-transcendent statement.

A second possible objection is that, although one cannot *in general* manifest knowledge of a statement's truth-conditions by restating them in a non-trivial and informative way, this may well happen in *some* cases. For instance, it might be thought that we can manifest knowledge of the truth-conditions of undecidable statements by using their component expressions in statements knowledge of whose truth-conditions *is* manifestable. But, Dummett says, this will not do. He offers the following rather compressed argument:

> The existence of [undecidable] sentences cannot be due solely to the occurrence of expressions introduced by purely verbal explanations: a language all of whose sentences were decidable would continue to have this property when enriched by expressions so introduced. (Dummett, 1976, p. 81)

It seems to follow that knowledge of the meanings of at least some undecidable sentences must in the end be implicit. On the other hand, as we have seen, it also seems that we have no model of how implicit knowledge of the truth-conditions of undecidable statements can be manifested. Dummett's conclusion is that we are left with no account of how knowledge of potentially verification-transcendent truth-conditions can be manifested. He writes:

> If the knowledge that constitutes a grasp of the meaning of a sentence has to be capable of being manifested in actual linguistic practice, it is quite obscure in what the knowledge of the condition under which a sentence is true can consist, when that condition is not one which is always being capable of being recognized as obtaining. (Dummett, 1973b, p. 228)

Indeed, Dummett goes as far as claiming that attributions of implicit knowledge of potentially verification-transcendent truth-conditions are deprived of content:

> Whenever the condition for the truth of a sentence is one that we have no way of bringing ourselves to recognize as obtaining whenever it obtains, it seems plain that there is no content to an ascription of *implicit*

> knowledge of what that condition is, since there is no practical ability by means of which such knowledge may be manifested. An ascription of the knowledge of such a condition can only be construed as *explicit* knowledge, consisting in a capacity to *state* the condition in some non-circular manner; and that, as we have seen, is of no use to us here. (Dummett, 1976, p. 82)

The challenge to the realist is to show that attribution of understanding of sentences with potentially evidence-transcendent truth-conditions is not deprived of content, and that knowledge of such truth-conditions can be manifested in use.

The Argument from Circularity

Quite surprisingly, Dummett has recently declared that, albeit "important", the Acquisition and the Manifestation arguments are not the "central" arguments against semantic realism. He says:

> Neither the objection arising from the manifestation nor that arising from the acquistion of the knowledge [of truth-conditions] is central. The central objection is the circularity of a truth-conditional account. (Dummett, 2006, p. 55)

The point is that if a theory of meaning is a theory of understanding, and if understanding a statement is knowing its truth-conditions, then the explanation of our understanding of φ, i.e. our explanation of our knowledge of φ's meaning, cannot depend upon a prior understanding of what φ means, on pain of circularity. Presumably, Dummett is reasoning as follows. Consider the following biconditional:

$$'A' \text{ is true if and only if } P,$$

where P expresses A's truth-conditions. Dummett's thought is that, if to understand a statement A is to know its truth-conditions, we are explaining what it is to grasp a thought, the thought expressed by A, in terms of what it is to grasp another thought, *that A's truth-conditions are so-and-so.* In Dummett's words:

> we are trying to explain what it is to grasp one proposition—that expressed by the sentence—in terms of judging another—the proposition that the sentence is judged under such-and-such conditions to be true. (Dummett, 2006, p. 50)

However, Dummett objects, this is just circular. As he puts it:

> A blanket account of understanding a statement as knowing what it is for it to be true is useless, because circular: it attempts to explain what it is to grasp a thought in terms of having a thought about that thought. (Dummett, 2006, p. 78)

The key assumption here is that a theory of meaning must give a non-circular account of what it is to grasp a concept, i.e. it "must embody an explanation of all the concepts expressible in that language" (Dummett, 1976, p. 5). In Dummett's terminology, a theory of meaning must be *full-blooded*. However, Dummett thinks, to merely "show or state which concepts are expressed by which words" (Dummett, 1976, p. 5), as truth-conditional theories of meaning typically do, falls short of giving a non-circular account of what it is to grasp a concept. A *modest* theory of meaning, Dummett suggests, is not, and cannot be, a theory of understanding.[14]

One interesting question is whether the Argument from Circularity rests on weaker premises than its most famous cousin, the Manifestation Challenge. This, I take it, is difficult to assess. On the one hand, unlike the Manifestation Challenge, the argument does not directly require that meaning be manifestable in use. It rather assumes that one should be able to say, for every concept, what it is to grasp that concept, in terms that do not require an understanding of that concept. On the other hand, this assumption is quite controversial—Dummett, it may be argued, is setting himself, and philosophers of language in general, an impossible task.

This completes our brief presentation of Dummett's challenges against semantic anti-realism.

The intended output of Dummett's arguments

The above arguments invite the conclusion that verification-transcendent truth-conditions are at odds with two seemingly platitudinous principles: that we learn the meanings of the sentences of our language by learning how to *use* these sentences, and that knowledge of truth-conditions must be manifestable in use. If one wishes to maintain these platitudes, it would seem that there cannot be unknowable truths, i.e. semantic realism is bankrupt. By contrast, Dummett claims that semantic anti-realists *can* offer an account of the acquisition and manifestation

[14]On Dummett's distinction between modest and full-blooded see e.g. Dummett (1976) and, *infra*, p. 105. The distinction has been the focus of a famous and long exchange between Dummett and John McDowell. See e.g. McDowell (1981), McDowell (1987), Dummett (1987b), McDowell (1997), McDowell (2007) and Dummett (2007b).

of understanding that complies with the foregoing platitudes. For the anti-realist, we acquire knowledge of the truth-conditions of a sentence S by learning what would establish it as true, and we manifest such a knowledge by being disposed to recognize proofs of S when presented with them. Understanding a sentence, for Dummett,

> is to be able to recognize a verification of it if one is produced, without needing to have a procedure for arriving at one. (Dummett, 1993b, p. 190)

Neil Tennant writes in a similar vein:

> What [anti-realists] maintain [...] is that grasp of meaning consists in an ability to decide, of any particular presentation, whether it establishes the sentence as true or false. (Tennant, 1981, p. 115)

Thus, for instance, we understand Goldbach's Conjecture because we would recognize a proof of it, if presented with one, even if we presently lack any such proof.[15]

But here is the rub: in order for the anti-realist account of understanding to work, there must be a guarantee that true statements are always provable, and false ones are always disprovable. In slogan: all truths must be knowable. Hence the link between Dummett's challenges and semantic anti-realism. As Dummett puts it:

> If meaning is use, that is, if the knowledge in which a speaker's understanding of a sentence consists must be capable of being fully manifested in by his linguistic practice, it appears that a model of meaning in terms of a knowledge of truth-conditions is possible only if we construe truth in such a way that the principle of bivalence fails; and this means, in effect, some notion of truth under which the truth of a sentence implies the possibility, in principle, of *our* recognizing its truth. (Dummett, 1979, p. 116)

The intended output of Dummett's challenges, then, is that truth must be knowable. Unknowable truths are but a metaphysical phantasy—one that is in tension with seemingly plausible and minimal platitudes concerning the manifestability of understanding. Or so Dummett argues.

[15]Strictly speaking, I should be talking here of *canonical* proofs, where—roughly—a canonical proof for a complex statement is a proof that ends with an application of one of the introduction rules for its main logical operator. More on canonical arguments in § 4.1.2 and in Appendix C.

Some possible objections

Much could be said about Dummett's challenges, and the circularity issue we have just raised. For instance, the Acquisition Challenge assumes that we learn the meaning of statements by coming to know what would establish them as true, and what follows from them. This is controversial, however. We often learn the meanings of new statements by coming to know the meanings of their component words, as used in other statements. Crispin Wright makes the point:[16]

> But now the realist seems to have a very simple answer. Given that the understanding of statements in general is to be viewed as consisting in possession of a concept of their truth-conditions, acquiring a concept of an evidence-transcendent state of affairs is simply a matter of acquiring an understanding of a statement for which that state of affairs would constitute a truth-condition. And such an understanding is acquired, like the understanding of any unheard sentence in the language, by understanding the constituent words and the significance of their mode of combination. (Wright, 1993, p. 16)

Some realists (see e.g. Byrne, 2005) have suggested that a similar response can be devised for the Manifestation Challenge. In their view, knowledge of the meaning of a statement φ need not be manifested by a capacity to use φ itself: it may well be manifested by manifesting a capacity to use its component expressions in other statements. I find this more problematic, though. Mere compentence with the component expressions of a statement, as manifested in one's use of other statements, does not in general add up to understanding that statement—let alone manifesting such an understanding. A quick example. Consider the sentence:

(NS) I rocked a slice above the quality.

It seems that we understand the component words of this seemingly well-formed sentence—we can indeed manifest such an understanding by correctly using them in a wide range of cases. Yet, it would seem, we do *not* understand NS.

More recently, Timothy Williamson has argued that Dummett's challenges establish too strong a result, viz. that the central semantic concept of a theory of meaning must be *decidable*. Since it is agreed on all parties that truth is in general *not* decidable, he suggests, anti-realists cannot identify the meanings of statements with their truth-conditions, thus being forced to ignore the recent progress that

[16]See also Hale (1997, p. 279).

has been made in linguistics and formal semantics—both of which share a truth-conditional background.[17] In Appendix A, I argue that Williamson's objection rests on a mistaken identification of the notions of truth and of a truth-maker. Whatever Dummett's arguments may establish, they do not require the decidability of the central semantic concept of a theory of meaning. But then, what, if anything, *do* they establish?

2.2 Tennant and Salerno on logical revision

It is often thought that Dummett's challenges, if sound, require that we abandon the canons of classical logic in favour of some weaker logic: intuitionistic logic. As we already saw, Dummett once wrote that the Acquisition and the Manifestation challenges (possibly together with his argument from circularity) are the only possible reasons "for repudiating classical reasoning [...] in favour of intuitionistic reasoning" (Dummett, 1973b, p. 216). This is puzzling, however. The challenges, as presented by Dummett himself, are merely arguments for rejecting semantic realism, the claim that there can be uknowable truths. Nothing has been said so far about the necessity of abandoning the canons of classical reasoning. So what is Dummett's argument from Acquisition and Manifestation to the adoption of intuitionistic logic? We will approach this question by first looking at what I take to be some mistaken reconstructions of the argument. We will consider two. In his *The Taming of the True*, Neil Tennant, perhaps the most prominent anti-realist in North-America, devotes one entire chapter, significantly entitled 'The Manifestation Argument is Dead', to criticizing what he takes to be Dummett's argument from the Manifestability Principle to the rejection of the Principle of Bivalence. According to Tennant, Dummett's challenges do *not* lead to the rejection of classical logic. Indeed, he contends that Dummett's Manifestation Challenge embodies "a logical mistake of numbing grossness". I will argue in Section 2.2.1 that Tennant's criticism is misguided: Dummett has never claimed that manifestability *alone* leads to the rejection of LEM and BIV. I will then turn in Section 2.2.2 to Joe Salerno's criticism of what he takes to be Dummett's central argument against LEM and BIV. I will suggest that Salerno's criticism is off target but helpful: Salerno correctly individuates two of the three main premises of Dummett's argument, but fatally equivocates on the logical form of the third.

[17] See Williamson (2008, pp. 282-4).

2.2.1 Tennant on manifestation and logical revision

Let us begin, then, with Tennant's reconstruction of Dummett's revisionary argument. I will argue that it is incorrect, and that, for this reason, Tennant's criticism of Dummett's argument is off-target.

Tennant's strategy

Tennant offers the following compact formulation of Dummett's Manifestation Challenge. He asks us to consider the following three principles:[18]

(A) The meaning of a declarative sentence is its truth-conditions.

(B) To understand a sentence is to know its meaning.

(C) Understanding is fully manifestable in the public exercise of recognitional skills.

He agrees with Dummett that, if we accept these principles, we are forced to conclude that understanding of statements with potentially verification-transcendent truth-conditions cannot be manifested. Hence, if understanding *must* be manifestable, truth cannot be verification-transcendent.

The problem, in Tennant's view, is that this is as far as Dummett's argument gets:

> The manifestation challenge, in its original simplicity, is disarmingly effective, but only against the notion of *recognition-transcendent* truth. (Tennant, 1997, p. 179)

The thought is that rejecting verification-transcendence is not enough for the anti-realist's purposes, because an argument against verification-transcendence is not, by itself, an argument against *semantic realism*. The reason, Tennant argues, is that semantic realism is the conjunction of *two* distinct theses: verification-transcendence *and* the unrestricted Principle of Bivalence. According to Tennant, Dummett's original challenge undermines the first conjunct, but poses no threat to bivalence and classical logic. In his words: "[the Manifestation Challenge] does not yet touch the Gödelian Optimist", i.e. the philosopher who, perhaps following Gödel or Hilbert, rejects verification-transcendence and, at the same time, asserts that every statement is either true or false.[19] Tennant takes this is to be a bad result: Gödelian Optimism was meant to be incompatible with what he takes to be the

[18]See Tennant (1997, pp. 176-7).

[19]The terminology is Stewart Shapiro's. See Shapiro (1993).

intended upshot of Dummett's original arguments, viz. that bivalence, Dummett's "hallmark" of realism, cannot be asserted across the board. He writes:

> In Dummett's hands the manifestation argument ha[d] been supposed to establish something more—namely, the incoherence of asserting bivalence 'across the board' for the discourse in question. (Tennant, 1997, p. 160)

Can Dummett's argument be turned into an argument against bivalence? Tennant suggests a negative answer. When directed against bivalence and classical logic, he alleges, Dummett's argument is hopelessly invalid:

> Dummett's manifestation argument, *in so far as it is directed against bivalence*, is, when properly regimented, revealed as embodying a 'non-sequitur of numbing grossness'.[20] (*Ibid.*)

Tennant's suggested upshot is that

> *bivalence*, the other central strand of realism, would appear to survive the manifestation challenge. (Tennant, 1997, p. 180)

So much for the headlines. Let us now try to get clearer on the argument Tennant is attributing to Dummett, and on his reasons of dissatisfaction with it.

The Single Premise Argument

Tennant attributes to Dummett a revisionary argument in two steps. Step one aims at showing that what Tennant calls the Manifestation Requirement

(MR) $\forall\varphi\forall S$(S understands $\varphi \rightarrow$ (($\varphi \rightarrow S$ can recognize a proof of φ if presented with one) $\wedge$($\neg\varphi \rightarrow S$ can recognize a disproof of φ if presented with one)))

is incompatible with the existence of undecidable statements that we do understand. Step two is the derivation of an inconsistency between the Manifestation Requirement and the Principle of Bivalence, via a subargument to the effect that bivalence entails the existence of *undecidable statements* (more on Tennant's interpretation of the notion in a moment). Schematically, we may represent Tennant's reconstruction of Dummett's argument as follows, where D is a discourse containing undecidable statements, and both ϕ and ψ are assumed to be understood:

[20]The expression is Strawson's, as Tennant points out.

$$
\dfrac{\begin{array}{c} \text{Bivalence for } D \\ \Pi \\ \exists\varphi \in D\ \varphi \text{ is undecidable} \end{array} \qquad \begin{array}{c} \underbrace{{}^{(1)}\dfrac{}{\psi \text{ is undecidable,}} \quad \begin{array}{c}\text{Manifestation}\\ \text{Requirement}\end{array}} \\ \Sigma \\ \bot \end{array}}{\bot}\;(1).
$$

Tennant calls this the *Single Premise Argument.*[21]

A 'non-sequitur of numbing grossness'

Tennant distinguishes two distinct readings of the Single Premise Argument, depending on one's understanding of the word 'undecidable'. The first option is to define φ's undecidability as the present lack of a decision procedure for determining φ's truth-value. This is essentially *our* definition of undecidability, which Tennant calls *effective undecidability*. The second is to interpret 'undecidable' as 'verification-transcendent'. Then, the above argument becomes:

$$
\dfrac{\begin{array}{c} \text{Bivalence for } D \\ \Pi_{VT} \\ \exists\varphi \in D\ \varphi \text{ is } VT \end{array} \qquad \begin{array}{c} \underbrace{{}^{(1)}\dfrac{}{\psi \text{ is } VT,} \quad \begin{array}{c}\text{Manifestation}\\ \text{Requirement}\end{array}} \\ \Sigma_{VT} \\ \bot \end{array}}{\bot}\;(1),
$$

where 'φ is VT' reads 'φ is verification-transcendent'. Tennant argues that neither reading is successful.

On the first reading, Tennant suggests, the subargument Σ, i.e. the derivation of an inconsistency between the Manifestation Requirement and the assumption that some statement ψ that we do understand is undecidable, becomes problematic *from an anti-realist standpoint*. It would transform the argument into a Trojan horse, since, after all, "even intuitionistic arithmetic is effectively undecidable" (Tennant, 1997, p. 184). The Manifestation Argument would thus "backfire", and "the anti-realist would be hoist with his own petard" (*Ibid.*). On the second reading, Σ_{VT} is acknowledged to be "watertight". But, Tennant contends, the second subargument Π_{VT} now becomes problematic. Tennant offers two arguments for this conclusion. The first aims at showing that the subargument is fallacious, on either reading (see Tennant, 1997, § 6.6.3). The second purports to establish that independence results are of no help to Dummett (see Tennant, 1997, § 6.6.4). Tennant concludes that

[21] The terminology was first suggested by Jon Cogburn.

> all that [...] still fails to make the desired logical transition available to the Dummettian: the transition, that is, from bivalence to the existence of recognition transcendent truths. (Tennant, 1997, p. 194)

The details of these two arguments need not concern us, for a very simple reason: Dummett, and anti-realists with him, can happily grant that bivalence *alone* does not entail the existence of verification-transcendent truths.

The Single Premise Argument revisited

Tennant's reconstruction of the Single Premise Argument requires the availability of a subargument to the effect that the Principle of Bivalence entails the existence of verification-transcendent truths. However, no anti-realist, including Dummett, has ever claimed that the Principle of Bivalence *alone* entails this much. Hence, it is not surprising that Tennant has been unable to find an argument to this effect. The Principle of Bivalence entails the existence of *potentially* verification-transcendent truths *only on the further assumption that there are undecidable statements*, i.e. statements for which we lack any guarantee that either them or their negations are knowable. Dummett unmistakeably makes the point:

> It is when the principle of bivalence is applied to undecidable statements that we find ourselves in the position of being unable to equate an ability to recognize when a statement has been established as true or as false with a knowledge of its truth-condition, since it may be true in cases when we lack the means to recognize it as true or false [...]. (Dummett, 1976, p. 63)

The headlines of Dummett's argument are clear enough: if there are undecidable statements, the Principle of Bivalence entails the existence of verification-transcendent truths. We may spell out the argument in more detail as follows:[22]

> *Proof*: Assume Bivalence. By the Equivalence Thesis, the Excluded Middle holds too; that is, we can apply classical logic unrestricedly. Now

[22]Jon Cogburn (2005) offers a revisionary argument along similar lines. First, Cogburn tells us (without offering any proof) that classical truth-conditional semantics (TCS) entails the epistemic possibility of the existence of absolutely undecidable statements. Second, he points out that, if all truths are knowable, there are no absolutely undecidable statements. More formally: $\mathsf{KP} \vdash \neg(\mathsf{UND}^*)$, $\mathsf{TCS} \vdash \mathsf{UND}^*$; from which we get $\mathsf{KP}, \mathsf{TCS} \vdash \bot$. It is worth noticing, however, that the only relevant property of truth-conditional semantics Cogburn is assuming, however, is that it validates the Principle of Bivalence. Hence, Cogburn is really saying that bivalence entails the potential existence of absolutely undecidable statements, but he ultimately fails to give an argument for this claim.

assume that there are undecidable statements, and let P be one of them. In symbols: $\neg\mathcal{K}_n(\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P)$. Then assume for arrow introduction that it is presently known that there are no verification-transcendent truths. In symbols: $\mathcal{K}_n\neg\exists\varphi(\varphi \wedge \neg\Diamond\mathcal{K}\varphi)$. By the factivity of knowledge, $\neg\exists\varphi(\varphi \wedge \neg\Diamond\mathcal{K}\varphi)$ follows. However, this entails that, for some P, both P and its negation are not verification-transcendent; that is, $\neg(P \wedge \neg\Diamond\mathcal{K}P)$ and $\neg(\neg P \wedge \neg\Diamond\mathcal{K}\neg P)$ hold. These conjunctions classically entail, respectively, $P \to \Diamond\mathcal{K}P$ and $\neg P \to \Diamond\mathcal{K}\neg P$. By the Excluded Middle, we can thereby infer by disjunction elimination that $\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P$. If knowledge is closed under known entailment, we presently know that P is decidable. That is: $\mathcal{K}_n(\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P)$. By arrow introduction, we may then derive $\mathcal{K}_n\neg\exists\varphi(\varphi \wedge \neg\Diamond\mathcal{K}\varphi) \to \mathcal{K}_n(\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P)$. Now assume $\mathcal{K}_n\neg\exists\varphi(\varphi \wedge \neg\Diamond\mathcal{K}\varphi)$ for negation introduction. By arrow elimination, $\mathcal{K}_n(\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P)$ follows. Contradiction. We must therefore negate and discharge our assumption that it is presently known that there are no verification-transcendent truths. In symbols: $\neg\mathcal{K}_n\neg\exists\varphi(\varphi \wedge \neg\Diamond\mathcal{K}\varphi)$. But this says that, for all we presently know, there are verification-transcendent truths. ■

In a nutshell, we have been able to derive, assuming classical logic and the existence of undecidable statements, that, for all we know, there are verification-transcendent truths. If, however, the potential existence of verification-transcendent truths is incompatible with the Manifestation Requirement, we cannot but conclude that either this requirement is faulty, or classical logic has to go.

Two observations are in order. First, I take it that the above proof definitely settles the issue concerning the interpretation of the anti-realist notion of undecidability: only on *our* interpretation does the above argument goes through. Second, it is now clear that there is a missing premise in Tennant's Single Premise Argument. At a glance, the argument must be corrected as follows:

$$\cfrac{\begin{array}{c}\text{Bivalence for } D,\ \exists\varphi(\varphi \text{ is undecidable})\\ \Pi_{\Diamond VT}\\ \exists\varphi \in D\ \varphi \text{ is } \Diamond VT\end{array} \qquad \begin{array}{c}\underbrace{{}^{(1)}\cfrac{}{\psi \text{ is } \Diamond VT,} \quad \begin{array}{c}\text{Manifestation}\\ \text{Requirement}\end{array}}\\ \Sigma_{\Diamond VT}\\ \bot\end{array}}{\bot}\ (1),$$

where 'φ is $\Diamond VT$' reads 'φ is a potentially verification-transcendent truth'. The modified argument rests on three main premises: the Principle of Bivalence, the existence of undecidable statements, and the Manifestation Requirement.

How are we to assess this argument? The main problem, I would like to suggest, is that it may not be general enough. Anti-realism *may* be motivated by the Manifestation Requirement, but it does not need to. Hilary Putnam, in his anti-realist phase during the 80's, is a case in point of an anti-realist whose reasons for subscribing to anti-realism were arguably independent from the Manifestation Requirement.[23] Presumably, what the anti-realist is really after is an argument from *anti-realism* to the rejection of exclusively classical canons of correct inference. For this reason, we will continue our search for a revisionary argument from broadly anti-realist ideas to the rejection of classical logic. All the same, it is worth emphasizing that, *pace* Tennant, the argument we have just presented *does* justice to Dummett's claim that the manifestability of meaning enjoins a rejection of the Principle of Bivalence—on the eminently plausible assumption that there are undecidable statements. We shall return to this argument in Section 2.4.5, where I will argue that it stands, or falls, with the Basic Revisionary Argument. For the time being, let us turn to Salerno's criticism of his own reconstruction of Dummett's revisionary argument.[24]

2.2.2 Salerno on logical revision

In his article *Revising the Logic of Logical Revision*, Joe Salerno has recently argued that Dummett's case for the adoption of intuitionistic logic relies on a set of inconsistent assumptions. He writes:

> Given the resources provided by [...] Dummett [...], choice of logic is not a realism-relevant feature—i.e., logical revision is not a consideration that is enjoined by one's stance on the possibility of verification transcendent truth. In fact, it is not clear that [...] Dummett [...] provides a consistent set of anti-realist commitments from which to argue. (Salerno, 2000, p. 212)

I will first presents Salerno's criticism of his reconstruction of Dummett's argument. I will then argue that it is misses its target.

Salerno on Dummett

According to Salerno, Dummett's revisionary argument rests on three main assumptions: the Knowability Principle, that all truths are knowable, the Law of

[23]See e.g. Putnam (1980), Putnam (1981), and Putnam (1983).

[24]I briefly consider Tennant's own revisionary argument, the Whole Discourse Argument, as he calls it, in Appendix B.

Excluded Middle, and the claim that there are undecidable statements. By 'undecidable', however, Salerno really means ***absolutely undecidable.*** In symbols:

$$(\mathsf{UND}^*)\ \exists\varphi(\neg\Diamond\mathcal{K}\varphi \wedge \neg\Diamond\mathcal{K}\neg\varphi).$$

His reconstruction of Dummett's argument may thus be represented as follows:[25]

$$\begin{array}{c} \underbrace{\overline{\mathsf{LEM}}^{\,(1)} \quad \mathsf{KP} \quad \mathsf{UND}^*} \\ \vdots \\ \dfrac{\bot}{\neg\mathsf{LEM}}\,(1). \end{array}$$

It is not difficult to see that there is something wrong with premise three. As Salerno points out,

> Despite appearances, this logical strategy ends in disaster for the revisionist. As it turns out, an intuitionistically acceptable reductio exists resting merely upon KP and UND*. Importantly, the contradiction resting on KP and UND* is intuitionistically acceptable. No exclusively classical principles are employed. (Salerno, 2000, p. 214)

The problem, Salerno observes, is that the Knowability Principle (KP) and UND* are *already* inconsistent! Formally:

$$\mathsf{KP} \vdash \neg(\mathsf{UND}^*).^{26}$$

Salerno concludes that *Dummett's* own case for logical revision is fatally flawed.

Undecidability and absolute undecidability

Salerno's reconstruction of Dummett's argument is, at best, extremely uncharitable. That the Knowability Principle entails that there are no absolutely undecidable statements had long been known to anti-realists. Here is Dummett:[27]

> It is impossible [...] that we should ever be in a position to assert, of any statement A, that A is neither absolutely provable nor refutable

[25]See Salerno (2000, p. 214).

[26]*Proof*: Assume $\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P$. By two steps of conjunction elimination, $\neg\Diamond\mathcal{K}P$ and $\neg\Diamond\mathcal{K}\neg P$ hold. Now assume P. By $\forall\varphi(\varphi \rightarrow \Diamond\mathcal{K}\varphi)$, derive $\Diamond\mathcal{K}P$. Contradiction. By negation introduction, $\neg P$. By similar reasoning, show that $\neg\neg P$. Contradiction. By negation introduction and universal generalisation, $\neg\exists\varphi(\neg\Diamond\mathcal{K}\varphi \wedge \neg\Diamond\mathcal{K}\neg\varphi)$. ■

[27]See also (Brouwer, 1908, p. 108).

> [...]. [Hence] it would be a complete mistake to replace the classical dichotomy true/false by a trichotomy provable/refutable/undecidable. (Dummett, 1977, p. 17)

Why, then, saddle Dummett with a principle that is inconsistent with his own beliefs?

The answer is probably to be found in passages from Dummett's early writings such as the following:

> the [realism/anti-realims] dispute can arise only for classes of statements for which it is admitted on both sides that there may not exist evidence either for or against a given statement. (Dummett, 1963b, p. 155)

One may read this passage as saying that there may be absolutely undecidable statements: statements, for which there is no evidence for or against. But this reading would be uncharitable. The above passage is more appropriately understood as saying only that there may be statements for which *now* there is no evidence either way. Such statements are *de facto* undecidable, but—for all we know—not absolutely undecidable.

Decisively, in more recent writings Dummett more clearly asserts that there are statements for which we presently lack a guarantee that either them or their negation are knowable:

> we are not entitled to assert, of every arbitrary proposition, that it is either provable or refutable. (Dummett, 1998, p. 128)

Following Cesare Cozzo (1998), I shall sometimes refer to this as the thesis of the *Missing Guarantee*. *This* is the real premise of the revisionary argument Salerno is trying to criticize: it is not intuitionistically inconsistent with semantic anti-realism, and, as we shall see below, it explicitly figures in Dummett's own presentation of the argument.

2.3 The Basic Revisionary Argument

Tennant's and Salerno's reconstructions of Dummett's argument have a common core. They both rest on three main premises: a broadly anti-realist principle, the Manifestability Requirement or the Knowability Principle, a classical thesis, the Principle of Bivalence or the Law of Excluded Middle, and a claim to the effect that there are undecidable statements. This is, very roughly, the basic structure of the premises of the Basic Revisionary Argument. The aim of this section is to briefly trace back the history of the argument, and to offer a fully regimented presentation of it. Our starting point will be the writings of Jan Brouwer, the founder of mathematical intuitionism.

2.3.1 Brouwer's line of argument

It is difficult to attribute to Brouwer a proper argument for the revision of classical *logic*. We shall nevertheless attempt a reconstruction of a Brouwerian line of argument for the adoption of intuitionistic logic.

The unreliability of the Excluded Middle

The starting point of Brouwer's reasoning is the observation that, if a statement is true only if it is provable, commitment to LEM enjoins commitment to the controversial claim that every problem is solvable. Brouwer presents this first part of his argument in a slightly misleading way:

> The question of the validity of the *principium tertii exclusi* is equivalent to the question whether unsolvable mathematical problems can exist. (Brouwer, 1908, p. 109)

Brouwer surely cannot mean by this that the question whether LEM holds is equivalent to the question whether there can be unsolvable problems. To be sure, if one assumes, as Brouwer does, that truth is epistemically constrained, then one may read LEM as saying that every problem is solvable. But this is not necessarily equivalent to the claim that there are no unsolvable problems. After all, the equivalence between $\Diamond\mathcal{K}A \vee \Diamond\mathcal{K}\neg A$ and $\neg(\neg\Diamond\mathcal{K}A \wedge \neg\Diamond\mathcal{K}\neg A)$ is only classically valid (intuitionistically, the former entails the latter, but the converse direction does not hold). Rather, what Brouwer means here is that, if truth is epistemically constrained, LEM is equivalent to the claim that every problem is solvable, or that every statement is decidable:

$$\text{(DEC)}\ \forall\varphi(\Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\neg\varphi).$$

DEC expresses the view, held by the famous German mathematician David Hilbert, that all mathematical truths are *decidable*. As Hilbert put it in his address at the Society of German Scientists and Physicians, in 1930:

> However unapproachable these problems may seem to us and however helpless we stand before them, we have, nevertheless, the firm conviction that the solution must follow by .[...] logical processes [...] This conviction of the solvability of every mathematical problem is a powerful incentive to the worker. We hear the perpetual call: There is a problem. Seek its solution. You can find it [...] for *in mathematics there is no ignorabimus.*

Hilbert's optimism is shared, for different reasons, by Gödel (hence Stewart Shapiro's choice of the label *Gödelian Optimism*):[28]

> [T]hose parts of mathematics which have been systematically and completely developed [...] show an amazing degree of beauty and perfection. In those fields, by entirely unsuspected laws and procedures [...] means are provided [...] for solving all relevant problems [...]. This fact seems to justify what may be called 'rationalistic optimism'.

No shred of a proof

Brouwer finds Hilbert's optimism hard to swallow. Here is his famous reply to Hilbert:

> there is not a shred of a proof for the conviction which has sometimes been put forward that there exist no unsolvable mathematical problems. (Brouwer, 1908, p. 109)

For consider so-called *weak* counterexamples to the Law of Excluded Middle:[29]

(4) There are seven consecutive '7' in the decimal expansion of π.

If truth requires knowability, and if a true disjunction must have one true disjunct, then either (4) or its negation must be knowable. Yet, it would seem, we certainly have no such guarantee! So how could we make such a bold prediction? It seems to follow that, if truth requires knowability, in absence of a proof that every problem can be solved, we cannot accept LEM.

[28] See Shapiro (1993) and *supra*, § 2.2.1.

[29] They are so called in that they do not *disprove* LEM. Rather, they only (purport to) show that we are not in a position to assert LEM in our present state of information.

2.3.2 From Brouwer to Wright

In many of their writings, Dummett and Prawitz endorse the foregoing line of argument as one of their main reasons for abandoning classical logic. In his *Truth and Objectivity*, Wright also formulates a version of the argument, which he later dubs the Basic Revisionary Argument. The argument is perspicuously presented for the first time in Salerno (2000) and Wright (2001).

From Brouwer to Dummett and Prawitz

In the *Introduction* to the *Logical Basis of Methaphysics*, Dummett explicitly mentions the Brouwerian line of argument we have just depicted. He writes:

> Those who first clearly grasped that rejecting realism entails rejecting classical logic were the intuitionists, constructivists mathematicians of the school of Brouwer. If a mathematical statement is true only if we are able to prove it, then there is no ground to assume every statement to be either true or false. (Dummett, 1991b, p. 9)

The argument is elaborated in more detail in the first edition of *Elements of Intuitionism*:

> The intuitionistic reconstruction of mathematics has to question even the sentential logic employed in classical reasoning. The most celebrated principle underlying this revision is the rejection of the law of excluded middle: since we cannot, save for the most elementary statements, guarantee that we can find either a proof or a disproof of a given statement $[\neg\mathcal{K}\forall\varphi(\Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\neg\varphi)]$, we have no right to assume, of each statement, that it is either true or false $[\neg\mathcal{K}\forall\varphi(\varphi \vee \neg\varphi)]$. (Dummett, 1977, p. 8)

In keeping with Dummett's presentation, Prawitz formulates the argument in his *Intuitionistic Logic: a Philosophical Challenge* thus:

> The difference between the two principles [realism and anti-realism] boils down to this: on the platonistic principle, a truth condition for a sentence obtains or does not obtain independently of our means of recognizing that it obtains or fails to obtain, and we are then forced to admit that there may be truths that are in principle impossible to recognize (*if we are not to assert unwarrantably that all problems are in principle solvable*); on the non-realistic principle, a truth is in principle

> always possible to recognize, but we must then refrain from asserting that a truth condition either obtains or does not obtain (*again, in order not to assert that everything is solvable*). (Prawitz, 1980, p. 9; italics mine)

These quotes strongly suggest that Tennant's and Salerno's criticisms of Dummett's argument rest on a mistaken reading of Dummett's (and Prawitz's) text: the argument to which both Dummett and Prawitz are referring is quite different from the arguments Tennant and Salerno are respectively attacking. Dummett's and Prawitz's argument is, in essence, Brouwer's: anti-realism and classical logic entail that every problem is solvable; but, since we have no guarantee that it is so, classical logic must go, if anti-realism holds.

Wright's scales of in principle evidence

The foregoing argument has been recently revived by Wright, in the second chapter of his book *Truth and Objectivity* (see Wright, 1992, pp. 37-44). Again, Wright's central result is that the Knowability Principle and the Law of Excluded Middle jointly entail that every problem is solvable, i.e. that every statement, or its negation, is knowable. However, Wright writes, this is "in contradiction with the a priori unwarrantability of the claim that the scales of in principle available evidence must tilt, sooner or later" (Wright, 1992, p. 43). That is, Wright seems to suggest, it is known *a priori* that *this*, that, for any statement, it is possible to have evidence either for it, or for its negation, is something we cannot legitimately assert.[30] Wright concludes that

> unless some other way of blocking the argument is found [...], the thesis [...] that truth is essentially evidentially constrained *must* enjoin a revision of classical logic, one way or another, for all discourses where there is no guarantee that evidence is available, at least in principle, to decide between each statement of the discourse concerned and its negation. (Wright, 1992, p. 43)

In short, as soon as we admit that "not every issue can be guaranteed to be decidable" (Wright, 1992, p. 41), i.e. that we do not presently know that every statement or its negation is knowable, commitment to the Knowability Principle mandates a revision of classical logic. In his *On Being in a Quandary*, Wright

[30] Wright's claim that we can know this *a priori* is very strong. Whether it is known a priori or not, however, is irrelevant to Wright's argument. As we shall see in due course, Wright himself has later questioned the claim that we do not presently know that every statement or its negation is knowable (see *infra*, pp. 43-6).

labels the argument the *Basic Revisionary Argument*.[31] In what follows, I will adopt Wright's felicitous terminology, and call the argument the Basic Revisionary Argument.[32]

Revising the logic of logical revision

Salerno (2000) offers the first semi-formal formalisation of the Basic Revisionary Argument. At a glance, his reconstruction is as follows ('$\mathcal{K}_n\varphi$' reads 'It is presently known that φ'; see Salerno, 2000, p. 219):

$$
\dfrac{\dfrac{\dfrac{\underbrace{\dfrac{}{\mathcal{K}_n\forall\varphi(\varphi\vee\neg\varphi)}{\scriptstyle(1)} \qquad \mathcal{K}_n\forall\varphi(\varphi\to\Diamond\mathcal{K}\varphi)}_{}}{\vdots}}{\mathcal{K}_n\forall\varphi(\Diamond\mathcal{K}\varphi\vee\Diamond\mathcal{K}\neg\varphi)} \qquad \neg\mathcal{K}_n\forall\varphi(\Diamond\mathcal{K}\varphi\vee\Diamond\mathcal{K}\neg\varphi)}{\dfrac{\bot}{\neg\mathcal{K}_n\forall\varphi(\varphi\vee\neg\varphi)}{\scriptstyle(1)}}
$$

The argument rests on three main premises: that both the Law of Excluded Middle and the Knowability Principle are presently known, on the one hand, and that we do not presently know, of every statement, that either it or its negation is knowable, on the other. This is a perspicuous presentation of the basic structure of the Basic Revisionary Argument, and it is to Salerno's credit to have been the first to point it out in sufficient detail.

A question immediately arises, however. Why would the classicist adopt the thesis of the Missing Guarantee? According to Salerno, the classicist

> would endorse this new form of modesty [...] just because it is so modest. The principle simply amounts to the humble recognition that

[31]See Wright (2001, p. 65).

[32]Salerno (2000) offers a criticism of Wright's revisionary argument that is essentially based on an interpretational mistake. He interprets Wright's reference to the "a priori unwarrantability of the claim that the scales of in principle evidence must tilt, sooner or later" as a commitment to the thesis that it is epistemically possible that "the decidability of the discourse could be false". In symbols:

$$(\neg(\mathrm{DEC}_s))\ \Diamond\neg(\Diamond\mathcal{K}A\vee\Diamond\mathcal{K}\neg A),$$

where '$\Diamond$' expresses epistemic possibility (notice that '$\Diamond$' is not to be confused with '$\Diamond$'). But, he says, "the anti-realist cannot endorse the epistemic possibility that decidability is false [...] because his epistemic constraint on truth is inconsistent with that possibility. More importantly, [KP] is *intuitionistically* inconsistent with the negation of the decidability thesis" (Salerno, 2000, p. 217). This is problematic for at least two reasons. First, the negation of the decidability thesis is only intuitionistically inconsistent when formulated as a schema. That is, $\neg\forall\varphi(\Diamond\mathcal{K}\varphi\vee\Diamond\mathcal{K}\neg\varphi)$ is *not* intuitionistically inconsistent. Second, Wright has never claimed that, for all we know, the decidability thesis could be false. Rather, he deems the decidability thesis to be inconsistent with "the admission that not every issue can be guaranteed to be decidable" (Wright, 1992, p. 41), i.e. with the thesis of the Missing Guarantee.

> we have not yet confirmed that each understood mathematical claim or its negation is humanly provable in the long run. (Salerno, 2000, p. 219)

Then again:

> my claim is that epistemic modesty is modest enough to warrant its endorsement by the relevant parties, and it is strong enough to play the logical role that the anti-realist intends for it. (Salerno, 2000, p. 223)

Salerno's idea is that the thesis of the Missing guarantee, that we do not presently know that every statement or its negation is knowable ($\neg\mathcal{K}_n(\mathsf{DEC})$), is strong enough to grant the desired output, and weak enough to be accepted by both parties. But this is not a very convincing argument! For one thing, one does not in general endorse φ "just because" φ is a weak claim. For another, it is unclear whether the classicist can appreciate the *weakness* of $\neg\mathcal{K}_n(\mathsf{DEC})$. For $\neg\mathcal{K}_n(\mathsf{DEC})$ is weaker than a claim to the effect that we presently do not know that for every statement it is not the case that it and its negation are unknowable

$$(\neg\mathcal{K}_n(\mathsf{DEC}^*)) \quad \neg\mathcal{K}_n\forall\varphi\neg(\neg\Diamond\mathcal{K}\varphi \wedge \neg\Diamond\mathcal{K}\neg\varphi)$$

only in logics that are themselves weaker than classical logic! And why would the classicist want to weaken her logic in order to appreciate that $\neg\mathcal{K}_n(\mathsf{DEC})$ is weaker than its classically equivalent counterpart $\neg\mathcal{K}_n(\mathsf{DEC}^*)$? We will turn this worry into a full-fledged objection in § 2.4 below.[33]

[33]There is a second, though minor, problem with Salerno's own comments to his own (partial) regimentation of the Basic Revisionary Argument. Salerno asks:

> How is it that this new formulation of epistemic modesty succeeds where our original formulation of Wrightian modesty fails? Notice that it is the *extra expressive power of quantified propositional logic* that blocks the anti-realistically unwelcome contradiction between the modesty principle and anti-realism. (Salerno, 2000, p. 219; italics added)

Salerno is here saying that the reason why the schema

$$(\neg(\mathsf{DEC}_s))\Diamond\neg(\Diamond\mathcal{K}A \vee \Diamond\mathcal{K}\neg A)$$

is inconsistent with anti-realism, while

$$(\neg\mathcal{K}_n(\mathsf{DEC}))\ \neg\mathcal{K}_n\forall\varphi(\Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\neg\varphi)$$

is not, is that the latter, but not the former, allows quantification over propositional variables. This is incorrect, however. It is certainly true that $\neg(\mathsf{DEC}_s)$ is inconsistent *qua* schematic: as we have pointed out, its quantified counterpart $\neg(\mathsf{DEC})$ *is* intuitionistically consistent. But it does not follow from this that quantification into sentence position is essential to the Basic Revisionary Argument. Indeed, the argument could be easily formulated with schemata. The thesis of the Missing Guarantee $\neg\mathcal{K}_n(\mathsf{DEC})$ would become $\neg\mathcal{K}_n(\Diamond\mathcal{K}A \vee \Diamond\mathcal{K}A)$, but the argument would go

2.3.3 Introducing the Basic Revisionary Argument

It is now time to present the Basic Revisionary Argument in detail, and to consider some first potential concerns.[34]

Consider the three following claims: the anti-realist thesis that all truths are knowable

$$\text{(WVER)}\ \ \forall\varphi(\varphi \rightarrow \Diamond\mathcal{K}\varphi);^{35}$$

the Law of Excluded Middle

$$\text{(LEM)}\ \ \forall\varphi(\varphi \vee \neg\varphi);$$

and the seemingly innocuous claim that we do not now know that every statement or its negation is knowable

$$(\neg\mathcal{K}_n(\text{DEC}))\ \ \neg\mathcal{K}_n\forall\varphi(\Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\neg\varphi),$$

where '$\Diamond$' denotes some notion of possibility, '$\mathcal{K}_n$' is to be interpreted as 'it is now known that', and '$\mathcal{K}$' is to be read as 'it is known at some time that'. The Basic Revisionary Argument, as presented by Wright and Salerno, proceeds from the assumption that $\neg\mathcal{K}_n(\text{DEC})$ holds and that WVER and LEM are known. The argument further requires that present knowledge is closed under known material implication:[36]

through just as well:

$$\dfrac{\dfrac{\underbrace{\dfrac{}{\mathcal{K}_n(A \vee \neg A)}^{\,(1)} \quad \mathcal{K}_n(A \rightarrow \Diamond\mathcal{K}A)}_{\displaystyle \vdots} \atop \mathcal{K}_n(\Diamond\mathcal{K}A \vee \Diamond\mathcal{K}\neg A) \qquad\qquad \neg\mathcal{K}_n(\Diamond\mathcal{K}A \vee \Diamond\mathcal{K}\neg A)}{\bot}}{\neg\mathcal{K}_n(A \vee \neg A)}\,(1)$$

Pace Salerno, quantification over sentence position is by no means essential to the Basic Revisionary Argument.

[34]The first formal presentation of the Basic Revisionary Argument was given by the present author and Luca Incurvati in the paper *How Basic is the Basic Revisionary Argument?* (see Incurvati and Murzi, 2008).

[35]Some restrictions might be called for. In particular, semantic anti-realism is usually taken as applying only to propositions expressed by sentences we do understand, and further restrictions have been suggested in order to solve the Paradox of Knowability. In keeping with the debate on logical revision, I set aside these complications for present purposes.

[36]To prevent this version of closure from being trivially false, I treat 'now' as referring to a time interval including the moment of utterance.

$$\text{(Closure)}\ \frac{\mathcal{K}_n A \qquad \mathcal{K}_n(A \rightarrow B)}{\mathcal{K}_n B}$$

and that if we have proved that A from no assumptions, then we can infer that A is now known:

$$(\mathcal{K}\text{-Introduction}), 1\ \frac{\overline{A}^{\,(1)}}{\mathcal{K}_n A}\,.$$

The argument is in three steps. First, it is proved that

$$(\mathsf{DEC})\ \forall\varphi(\Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\neg\varphi)$$

follows from WVER and LEM. By $\mathcal{K}$-Introduction, we thereby know that $(\mathsf{LEM} \wedge \mathsf{WVER}) \rightarrow \mathsf{DEC}$. On the assumption that LEM and WVER are also known, it follows by closure[37] that $\mathcal{K}_n(\mathsf{DEC})$.[38] However, this latter claim is inconsistent with $\neg\mathcal{K}_n(\mathsf{DEC})$. Something must go. Suppose that WVER has been conclusively established. If $\neg\mathcal{K}_n(\mathsf{DEC})$ is not up for grabs, we are left with no choice but to discharge the assumption that LEM is known. At a glance, the argument may be represented as follows:

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\overline{\mathsf{LEM}}^{\,(1)} \qquad \overline{\mathsf{WVER}}^{\,(2)}}{\mathsf{DEC}}}{(\mathsf{LEM} \wedge \mathsf{WVER}) \rightarrow \mathsf{DEC}}\,{}^{(1,2)}}{\mathcal{K}_n((\mathsf{LEM} \wedge \mathsf{WVER}) \rightarrow \mathsf{DEC})}\,{}^{(\mathcal{K}\text{-I})} \qquad \overline{\mathcal{K}_n(\mathsf{LEM})}^{\,(3)} \qquad \mathcal{K}_n(\mathsf{WVER})}{\mathcal{K}_n(\mathsf{DEC})}\,{}^{\text{(Closure)}} \qquad \neg\mathcal{K}_n(\mathsf{DEC})}{\cfrac{\bot}{\neg\mathcal{K}_n(\mathsf{LEM})}\,{}^{(3)}}$$

If LEM is unknown, Wright contends, its status as a logical law is jeopardized. As he puts it:

> Since logic has no business containing logical principles that are uncertain, classical logic is not acceptable in our present state of information. (Wright, 2001, p. 66)

[37] Strictly speaking, it does not follow by Closure, which only allows single-premise closure. However, it does follow by Closure and $\mathcal{K}$-Introduction, which jointly yield normality for $\mathcal{K}$. For ease of exposition, I simply talk of closure, here and throughout.

[38] *Proof*: Assume LEM and WVER. Now show, by disjunction introduction and arrow introduction that $\Diamond\mathcal{K}P \rightarrow (\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P)$. It follows, by transitivity of '$\rightarrow$' and WVER, that $P \rightarrow (\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P)$. By similar reasoning, we can show that $\neg P \rightarrow (\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P)$. But LEM licenses us to infer $\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P$ from $P \rightarrow (\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P)$ and $\neg P \rightarrow (\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P)$ by disjunction elimination and arrow elimination. Therefore, by arrow introduction, $\forall$ introduction and pushing of $\forall$ from conditional with closed antecedent to consequent, $(\mathsf{LEM} \wedge \mathsf{WVER}) \rightarrow \forall\varphi(\Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\neg\varphi)$. Hence, by $\mathcal{K}$-Introduction, $\mathcal{K}_n((\mathsf{LEM} \wedge \mathsf{WVER}) \rightarrow \forall\varphi(\Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\neg\varphi))$. From this, $\mathcal{K}_n(\mathsf{LEM})$ and $\mathcal{K}_n(\mathsf{WVER})$, it follows, by closure, that $\mathcal{K}_n\forall\varphi(\Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\neg\varphi)$. ■

Some *prima facie* potential concerns

Some potential concerns with the Basic Revisionary Argument are worth mentioning. Firstly, one might argue that the argument equivocates on the relevant notion of knowledge involved. Thus, Jon Cogburn writes:

> the more reasonable it is to claim that it is currently known that Dummettian anti-realism is true, the less reasonable it is to claim that epistemic modesty [i.e. $\neg\mathcal{K}_n(\mathsf{DEC})$] is true. (Cogburn, 2002, p. 242)

Cogburn's worry seems to be this: the weaker one's epistemic attitude towards WVER is, the less plausible is the claim that we do not have that epistemic attitude towards DEC. For instance, whereas it may after all be plausible to say that it is rational to believe WVER, it seems less plausible to say that it is not rational to believe that every statement or its negation is knowable.

Secondly, even anti-realists might be reluctant to affirm that they *know* WVER. Thus Cogburn again:

> though I consider myself an anti-realist, I would never claim to know that [WVER] is true. I think, feel, and hope that it's true, but (warrants for philosophical views being what they are) I would never claim to know that it's true. (Cogburn, 2002, pp. 241–242)

Thirdly, Wright provides no argument for his claim that logical laws must be known *a priori*, so that there might be room to claim that the conclusion of the Basic Revisionary Argument does not lead us to give up our acceptance of LEM.

These worries, however, disappear on reflection. For consider the following simplified version of the Basic Revisionary Argument:[39]

$$\dfrac{(\mathcal{K}\text{-I})\;\dfrac{\dfrac{\dfrac{}{\mathsf{LEM}}{}^{(1)} \quad \dfrac{}{\mathsf{WVER}}{}^{(1)}}{\mathsf{DEC}}}{\mathcal{K}_n(\mathsf{DEC})}{}^{(1)} \qquad \neg\mathcal{K}_n(\mathsf{DEC})}{\bot}$$

This simplified argument derives a contradiction from LEM, WVER, and $\mathcal{K}_n(\mathsf{DEC})$. By itself, though, this does not force us to negate and discharge one of our assumptions. We may simply take this *reductio* as a reason not to include LEM among our axioms, in presence of WVER and $\mathcal{K}_n(\mathsf{DEC})$. The modified argument, therefore, does not establish the negation of LEM, but only that we should not include LEM in our system, on pain of contradiction. Moreover, the argument does not assume

[39]See, for example, Williamson (1992, p. 65). See also Cozzo (1989) and Cozzo (1998).

$\mathcal{K}_n$(WVER), but only that WVER is an axiom of the system. As a result, the two worries raised by Cogburn disappear: one does not need to assume that WVER is *known*, but only that it is an axiom of our system. The fact that both versions of the argument rely on $\mathcal{K}$-I, on the other hand, seems to show that some principle of this kind is indeed required in order to carry it out.

Wright on $\neg\mathcal{K}_n$(DEC)

The Basic Revisionary Argument presents us with a trilemma whose horns are our right to assert that anti-realism holds, our right to apply classical logic across the board, and the plausible claim that we presently lack a guarantee that every statement is decidable. Most of the weight is arguably on the first premise: that anti-realism is known. The third premise, however, that DEC is presently unknown, has recently come under attack. Wright writes:

> There is a problem [...] with the Basic Revisionary Argument. It is: what justifies $\neg\mathcal{K}_n$(DEC)? It may seem just obvious that we do not know that is feasible to decide any significant question (what about vagueness, backwards lights cones, Quantum Mechanics, Goldbach, the Continuum Hypothesis, etc.?). But for the anti-realist, though not for the realist, this modesty needs to be able to stand alongside our putative knowledge of WVER. And there is a doubt about the stability of that combination. (Wright, 2001, p. 67; Wright's terminology has been adapted to ours)

Wright is here suggesting that the anti-realist's reasons for adopting the thesis of the Missing Guarantee, that we do not presently know that every statement, or its negation, is knowable, may turn out to be inconsistent with semantic anti-realism. His argument is as follows. Let us ask ourselves: "what does it take *in general* to justify the claim that a certain statement is not known?" (*Ibid.*). Wright suggests the following *principle of agnosticism*:

> (AG) P should be regarded as unknown just in case there is some possibility Q such that if it obtained, it would ensure not-P, and such that we are (warranted in thinking that we are) in no position to exclude Q. (Wright, 2001, pp. 67-8)

The principle says that we do not know P if there is a Q such that (i), we are warranted in thinking that, for all we know, Q holds, and (ii) Q entails $\neg P$. For instance, I do not know that my bike is still parked where I left it, given that (i)

I am warranted in thinking that, for all I know, it has been stolen, and (ii) that it has been stolen entails that it is *not* still parked where I left it. Now to the Basic Revisionary Argument. If we accept AG, Wright says, the case at hand will demand us to find a Q whose obtaining would entail the falsity of DEC. But can there be such a Q, Wright asks? Apparently not. For if there were such a Q, DEC would be false, which is however classically inconsistent with WVER.[40] Wright concludes that

> given WVER, there can be no such appropriate Q. So given WVER and AG there can be no way of justifying $\neg\mathcal{K}_n(\mathsf{DEC})$. Thus the intuitive justification for $\neg\mathcal{K}_n(\mathsf{DEC})$ is, seemingly, not available to the anti-realist. (Wright, 2001, p. 68)

It may be objected that the foregoing problem only arises if the background logic is classical. However, Wright convincingly argues, this is a context in which logic has not *yet* been revised. Therefore, it would be question begging to appeal to intuitionistic restrictions in a context in which we are trying to establish their validity. As Wright puts it:

> Obviously we cannot just help ourselves to distinctively intuitionistic restrictions in the attempt to stabilise the argument if the argument is exactly intended to motivate such restrictions. (*Ibid.*)

Response to Wright

It appears on reflection that Wright is creating an unnecessary difficulty for the anti-realist, for at least two reasons.

To begin with, the principle AG seems just circular, on the plausible assumption that 'We are in no position to exclude P' is to be glossed as 'We do not know $\neg P$'. On this assumption, the principle tells us that P is not known just in case there is a Q such that (i) if Q were the case, $\neg P$ would be the case, and (ii) it is *not known* that $\neg Q$. Wright might object that 'We are in no position to exclude P' is to be rather glossed as 'We are not warranted in believing P'. But there are problems with this too. On this reading, AG now entails that, for some P, P is unknown only if we are (warranted in thinking that we are) not warranted in believing that P.

> *Proof*: Let Q be $\neg P$. Then, AG gives us that P is unknown just in case (i) if $\neg P$ were the case, then $\neg P$ would be the case (which is a trivial

[40] *Proof*: Assume WVER. Now assume that not every statement is decidable. By classical reasoning, it follows that, for some φ, φ is absolutely undecidable, i.e. $\neg\Diamond\mathcal{K}\varphi \wedge \neg\mathcal{K}\neg\varphi$ holds. Given WVER, this entails $\neg P \wedge \neg\neg P$. ■

> logical truth) and (ii) we are (warranted in thinking that we are) not warranted in believing $\neg\neg P$. Classically, it follows that P is unknown only if we are (warranted in thinking that we are) not warranted in believing P. ■

But this is clearly false. We may have a warrant for P, even though we do not *know* P.

Secondly, Wright seems wrong in thinking that, in the case at hand, "there can be no such appropriate Q", i.e. a Q whose obtaining would ensure the falsity of DEC. For let Q be $\neg$(DEC). Then, according to Wright's principle of agnosticism, we get that DEC is unknown only if (i) $\neg$(DEC) entails itself and (ii) we are not warranted in thinking that $\neg\neg$(DEC).

Wright might object that, intuitionistically, we *are* warranted in thinking $\neg\neg$(DEC), though classically we are not, since this would entail that we are warranted in thinking DEC. However, recall, we are assuming, with Wright, that this is a context in which we may not already assume that the logic is intuitionistic—we are here trying to establish one of the premises of an argument for intuitionistic logic.

So how to justify the thesis of the Missing Guarantee? Anti-realists, I take it, have a standard answer to this question: statements of the form $\neg A$, such as the thesis of the Missing Guarantee, can be correctly asserted if A is inconsistent with what we presently know. This standard answer, I think, can help us finding a justification for the thesis of the Missing Guarantee. Let us assume that we presently know that every statement is decidable. If this is true, then we presently know that Goldbach's Conjecture is decidable too, on the further, and plausible, assumption that present knowledge is closed under presently known logical consequence. That is, we now know that there is either a proof, or a disproof, of Goldbach's Conjecture. But, of course, we know that we do *not* know that! It follows that the assumption that $\mathcal{K}_n$(DEC) holds is inconsistent with what we presently know. By one step of negation introduction, we may legitimately infer $\neg\mathcal{K}_n$(DEC).

The most pressing issue, I suggest, is not whether $\neg\mathcal{K}_n$(DEC) can itself be justified. Rather, the problem is whether our reasons for accepting it are weak enough not to be reasons for accepting a different, and more dangerous, formulation of our epistemic modesty. It is to this problem that we now turn.

2.4 How basic is the Basic Revisionary Argument?

In commenting Salerno's own comments to the Basic Revisionary Argument, we observed that it is unclear why one should accept, as a formulation of one's epistemic modesty, $\neg\mathcal{K}_n(\mathsf{DEC})$, the claim that we do not presently know that every statement or its negation is knowable, instead of $\neg\mathcal{K}_n(\mathsf{DEC})^*$, the claim that we do not presently know that, for every statement, it is not the case that both it and its negation are unknowable. Salerno suggests that she should do so "just because $[\neg\mathcal{K}_n(\mathsf{DEC})]$ is so weak". However, we observed, its weakness cannot be appreciated by the classical logician! This is a serious problem. A revisionary argument starting from $\neg\mathcal{K}_n(\mathsf{DEC})^*$, which is classically just as weak as $\neg\mathcal{K}_n(\mathsf{DEC})$, does not lead to the adoption of intuitionistic logic. Rather, the reasoning involved in the Basic Revisionary Argument, if correct, validates a parallel argument that leads to conclusions that are unacceptable to classicists and intuitionists alike, namely that the Law of Non-Contradiction is presently unknown. As I show, the point generalizes to our emended version of the Single Premise Argument.[41]

2.4.1 How Basic is the Basic Revisionary Argument?

I focus on the third premise of the argument: $\neg\mathcal{K}_n(\mathsf{DEC})$. Recall that Wright's argument is meant to convince the classicist that, if WVER is known, classical logic is to be given up, since we do not know that for every statement it or its negation is knowable. However, our epistemic condition also seems to be such that we do not know that for every statement it is not the case that it and its negation are unknowable. That is, our reasons for thinking that $\neg\mathcal{K}_n(\mathsf{DEC})$ holds also seem to be reasons for thinking that

$$(\neg\mathcal{K}_n(\mathsf{DEC}^*)) \quad \neg\mathcal{K}_n\forall\varphi\neg(\neg\Diamond\mathcal{K}\varphi \wedge \neg\Diamond\mathcal{K}\neg\varphi)$$

holds. But now, armed with $\neg\mathcal{K}_n(\mathsf{DEC}^*)$, the classicist can run, in perfect analogy with the Basic Revisionary Argument, an argument to the effect that the Law of Non-Contradiction,

$$(\mathsf{LNC}) \quad \forall\varphi\neg(\varphi \wedge \neg\varphi),$$

is not known and should not thereby be taken as a logical law, at least according to Wright's own standards of logical lawhood. Like the Basic Revisionary Argument,

[41] Some of the contents of this section constitute an elaboration of materials presented in Incurvati and Murzi (2008).

the argument is in three steps. First, it is proved that

$$(\mathsf{DEC}^*)\ \forall\varphi\neg(\neg\Diamond\mathcal{K}\varphi \wedge \neg\Diamond\mathcal{K}\neg\varphi)$$

follows from WVER and LNC. By $\mathcal{K}$-Introduction, we thereby know that $(\mathsf{LNC} \wedge \mathsf{WVER}) \rightarrow \mathsf{DEC}^*$. On the assumption that LNC and WVER are also known, it follows by closure that $\mathcal{K}_n(\mathsf{DEC}^*)$.[42] But $\mathcal{K}_n(\mathsf{DEC}^*)$ is inconsistent with $\neg\mathcal{K}_n(\mathsf{DEC}^*)$. As a result, we have to discharge the assumption that LNC is known.[43] Again, we can give a perspicuous formalization of the argument as follows:

$$
\begin{prooftree}
\AxiomC{}
\RightLabel{(1)}
\UnaryInfC{LNC}
\AxiomC{}
\RightLabel{(2)}
\UnaryInfC{WVER}
\BinaryInfC{DEC^*}
\RightLabel{(1, 2)}
\UnaryInfC{$(\mathsf{LNC} \wedge \mathsf{WVER}) \rightarrow \mathsf{DEC}^*$}
\RightLabel{($\mathcal{K}$-I)}
\UnaryInfC{$\mathcal{K}_n((\mathsf{LNC} \wedge \mathsf{WVER}) \rightarrow \mathsf{DEC}^*)$}
\AxiomC{}
\RightLabel{(3)}
\UnaryInfC{$\mathcal{K}_n(\mathsf{LNC})$}
\AxiomC{$\mathcal{K}_n(\mathsf{WVER})$}
\RightLabel{(Closure)}
\TrinaryInfC{$\mathcal{K}_n(\mathsf{DEC}^*)$}
\AxiomC{$\neg\mathcal{K}_n(\mathsf{DEC}^*)$}
\BinaryInfC{$\bot$}
\RightLabel{(3)}
\UnaryInfC{$\neg\mathcal{K}_n(\mathsf{LNC})$}
\end{prooftree}
$$

The argument is intuitionistically valid. Hence, we cannot solve the problem just by discharging LEM. Since both the classicist and the intuitionist are agreed that LNC *is* a logical law, the argument leads to a conclusion that is unacceptable to both parties.[44]

[42]*Proof*: Assume LNC and WVER. Now assume $\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P$. By conjunction elimination, $\neg\Diamond\mathcal{K}P$ and $\neg\Diamond\mathcal{K}\neg P$. By contraposition of WVER, $\neg\Diamond\mathcal{K}P \rightarrow \neg P$. It thus follows, by arrow elimination, that $\neg P$. On the other hand, from WVER and substitution of P with $\neg P$, $\neg P \rightarrow \Diamond\mathcal{K}\neg P$. It follows, by arrow elimination, that $\Diamond\mathcal{K}\neg P$. But this contradicts $\neg\Diamond\mathcal{K}\neg P$. Hence, LNC licenses us to infer, by negation introduction, $\neg(\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P)$. Therefore, by arrow introduction, $\forall$ introduction and pushing of $\forall$ from conditional with closed antecedent to consequent, $(\mathsf{LNC} \wedge \mathsf{WVER}) \rightarrow \forall\varphi\neg(\neg\Diamond\mathcal{K}\varphi \wedge \neg\Diamond\mathcal{K}\neg\varphi)$. Hence, by $\mathcal{K}$-Introduction, $\mathcal{K}_n((\mathsf{LNC} \wedge \mathsf{WVER}) \rightarrow \forall\varphi\neg(\neg\Diamond\mathcal{K}\varphi \wedge \neg\Diamond\mathcal{K}\neg\varphi))$. From this, $\mathcal{K}_n(\mathsf{LNC})$ and $\mathcal{K}_n(\mathsf{WVER})$, it follows, by closure, that $\mathcal{K}_n\forall\varphi\neg(\neg\Diamond\mathcal{K}\varphi \wedge \neg\Diamond\mathcal{K}\neg\varphi)$. ■

[43]I consider another possible outcome of the argument in Section 2.4.4 below.

[44]Bob Hale and Crispin Wright have independently pointed out the following alleged disanalogy between the two arguments. In the Basic Revisionary Argument LEM is only used as an assumption, whereas in the parallel argument LNC is used as as an assumption as well as a rule of inference, in order to conclude that LNC itself ought to be abandoned. This, one might object, makes the parallel argument viciously circular. While I agree that there is a disanalogy between the two arguments, as they have been presented here, I think that more needs to be done in order to show that we cannot use a rule of inference to show its own invalidity. Moreover, it might be argued, the disanalogy between the two arguments disappears once we regiment the proofs in a Hilbert-style system whose sole rule of inference is modus ponens. For clearly in such a regimentation neither LEM nor LNC are used as inference rules (thanks to Marcus Rossberg for this suggestion). For reasons of space, I cannot discuss the problem further in this thesis, though I intend to do so in my future work.

2.4.2 Objections and replies

How could the anti-realist respond to the problem? To begin with, she might insist that discharging $\mathcal{K}_n$(LNC) is not an option and that the classicist should rather negate and discharge $\neg\mathcal{K}_n$(DEC*). This would be a bad move, however, since it would also block the Basic Revisionary Argument, unless an argument is provided that explains why one can discharge $\neg\mathcal{K}_n$(DEC*) but not $\neg\mathcal{K}_n$(DEC) in Wright's original proof. One such argument runs as follows. The idea is that the classicist cannot legitimately assume $\neg\mathcal{K}_n$(DEC*) since the classicist who knows WVER already knows DEC*, given that the latter intuitionistically follows from the former. This would show that, if $\mathcal{K}_n$(WVER) holds, it is a fact of the matter that we know that for every statement it is not the case that it and its negation are unknowable. This argument does not work, however, since a parallel argument shows that, on the same assumption, it is a fact of the matter that for every statement either it or its negation is knowable, given that $\mathcal{K}_n$(DEC) classically follows from $\mathcal{K}_n$(WVER).

To be sure, if the background logic were intuitionistic, the classicist would be in a position to appreciate that DEC* follows from WVER while being unable to see that DEC follows from it. That the background logic of logical revision should be intuitionistic has been claimed, for example, by Joe Salerno. In support of this claim, he writes:

> A classicist sincerely and meaningfully disagreeing with the anti-realist about anti-realism cannot invoke logical norms that the anti-realist finds unfavourable. (Salerno, 2000, p. 221)

The idea seems to be that the background logic of logical revision should be neutral between the classicist and her opponent. In reply to this, one might argue that the background logic should be classical, since the revisionist cannot ask the classicist to weaken her logic *before* the revisionary argument is run. Wright himself seems to endorse this view when discussing the issue of the background logic in another context:

> The trouble with this, of course, is that we precisely may not take it that the background logic *is* (already) intuitionistic; rather the context is one in which we are seeking to capture an argument to the effect that it *ought* to be. (Wright, 2001, p. 68)

But even if we grant that the background logic of logical revision should be neutral between the classicist and her opponent, the difficulty remains. For in order to solve the problem, the background logic of the parallel argument would have to be at least as strong as to contain LNC.[45] And this simply does not follow from the requirement that the background logic should be neutral between the classicist and her opponent. Consider the debate between the classicist and someone who wants to convince her to abandon LNC. Here, there seems to be no reason why the derivation of DEC* via LNC should be taken as showing that the premise $\neg\mathcal{K}_n(\text{DEC}^*)$ is lacking, unless it is accepted that—in the context of the debate between the classicist and the intuitionist—the classical derivation of DEC from WVER shows that $\neg\mathcal{K}_n(\text{DEC})$ is also lacking.

Salerno elsewhere suggests that 'both the anti-realist and the classicist would endorse $[\neg\mathcal{K}_n(\text{DEC})]$ ...just because it is so modest' (Salerno, 2000, p. 219). This seems to be off the point, however. What needs to be shown is that our reasons for endorsing $\neg\mathcal{K}_n(\text{DEC})$ are not as strong as to be reasons for endorsing $\neg\mathcal{K}_n(\text{DEC}^*)$, or that our reasons for accepting $\mathcal{K}_n(\text{DEC}^*)$, if any, are not as strong as to be reasons for accepting $\mathcal{K}_n(\text{DEC})$.[46]

2.4.3 Wright on epistemic modesty

How could anti-realists respond to the foregoing challenge? Wright (2001) offers two arguments for discriminating between $\neg\mathcal{K}_n(\text{DEC})$ and $\neg\mathcal{K}_n(\text{DEC}^*)$. The first

[45] This is needed in order to derive DEC*, as we have seen.

[46] Cesare Cozzo has suggested the following argument for accepting $\mathcal{K}_n(\text{DEC}^*)$. Assume that we have established that it is not possible to know, for any particular φ, that it is not possible to know that φ and it is not possible to know that $\neg\varphi$ (because, say, for any φ, we can never rule out that there is a possible way of coming to know φ). Without appealing to the intuitionistic meaning of negation, we therefore have, for an arbitrary P:

(i) $\neg\Diamond\mathcal{K}(\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P)$.

By an instance of WVER, however:

(ii) $(\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P) \rightarrow \Diamond\mathcal{K}(\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P)$.

Hence, by *modus tollens* from (i) and (ii):

(iii) $\neg(\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P)$.

By $\forall$-I and $\mathcal{K}$-I,

(iv) $\mathcal{K}_n\forall\varphi\neg(\neg\Diamond\mathcal{K}\varphi \wedge \neg\Diamond\mathcal{K}\neg\varphi)$.

That is, we have just *proved* $\mathcal{K}_n(\text{DEC}^*)$. Wright's original argument, it would seem, validates no parallel argument to the effect that LNC is presently unknown, because the premise $\neg\mathcal{K}_n(\text{DEC}^*)$ can be independently shown to be lacking.

But this will not do. The step from (ii) to (iii) requires that *modus tollens*, and hence LNC, be already in place. But, if it were legitimate in the present dialectic to prove $\mathcal{K}_n(\text{DEC}^*)$ by means of LNC, then it should be equally legitimate to prove $\mathcal{K}_n(\text{DEC})$ by means of LEM. The foregoing argument offers no reasons to discriminate between $\neg\mathcal{K}_n(\text{DEC})$ and $\neg\mathcal{K}_n(\text{DEC}^*)$ as formulations of our epistemic modesty.

aims at establishing that we know DEC*, independently of the question whether we also have knowledge of DEC. The second is meant to provide a compelling reason for adopting $\neg\mathcal{K}_n$(DEC) while rejecting $\neg\mathcal{K}_n$(DEC*). Neither argument ultimately succeeds, or at least so I will argue. I begin with the first.

Wright's first argument

In presence of WVER, Wright argues, affirming that P is unknowable is tantamount to affirming that P is false. Yet, Wright points out, there are cases of epistemic indeterminacy, say, Goldbach's conjecture or a borderline case of 'x is red', such that we do not seem in a position to rule out P's truth. As he puts it:

> if we could know that we couldn't know, then we would know that someone who took a view, however tentative—say that x was red—was wrong to do so. But we do *not* know that they are wrong to do so—the indeterminacy precisely leaves it open. (Wright, 2001, p. 73)

If WVER holds, P's unknowability is tantamount to affirming that P is false. Yet, Wright suggests, this is in contrast with our intuition that an assertion of P, however tentative, should not be ruled out *a priori*. But then, if the very thought that $\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P$ is *a priori* mistaken, we cannot but conclude that $\neg(\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P)$, i.e. DEC*, is known.

Following Wright, let us call cases of epistemic indeterminacy such as Goldbach's conjecture or, say, borderline cases of 'x is red' *quandaries*. Wright's definition of a quandary is as follows: φ is a quandary just if

> we do not know, do not know how we might come to know, and can produce no reason for thinking that there is any way of coming to know what to say or think. (Wright, 2001, p. 71)

With this definition on board, we may rewrite Wright's argument as follows:

> Let P be a quandary. Now assume that it is possible to know that neither P nor $\neg P$ are knowable. Then, we would know that an utterance of either P or its negation would be mistaken. Since we do *not* know that an utterance of either P or its negation would be mistaken, our initial assumption must be discharged. That is, we must conclude that it is impossible to know that neither P nor $\neg P$ are knowable. By WVER, this gives us DEC*, i.e. the claim that it is not the case that both P are $\neg P$ are unknowable. ■

This argument warrants a couple of remarks.

To begin with, its first main step needs further clarification. For how can we legitimately infer from the possibility of knowing that both P and $\neg P$ are unknowable knowledge that an utterance of either P or its negation would be mistaken? Presumably, Wright is here assuming that knowability is factive:

$$(\mathsf{FAC}_{\Diamond})\ \forall\varphi(\Diamond\mathcal{K}\varphi \to \varphi).$$[47]

More importantly, the argument makes use of *modus tollens*, and hence of LNC. Thus, if we were to accept it, we would also have to accept a parallel argument involving LEM, and possibly $\mathsf{FAC}_{\Diamond}$, to the effect that DEC is also presently known. One such argument goes as follows. Assume that DEC is false. In symbols: $\neg\forall\varphi(\Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\neg\varphi)$. By $\mathsf{FAC}_{\Diamond}$, it follows that $P \vee \neg P$ is itself false, for some P. But this contradicts the Law of Excluded Middle. Hence, $\neg\neg\forall\varphi(\Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\neg\varphi)$ holds. By $\mathcal{K}$-I, and by one step of Double Negation Elimination, we can conclude that DEC is known.

Again, the moral is familiar: it is of no use to the anti-realist to prove DEC* by means of LNC, since a parallel argument resting on LEM to the effect that DEC is known would also be available.

Wright's second argument

Wright's second argument rests on (i) the following necessary condition for knowledge:

> (AG^+) $\forall\varphi$(φ is known only if there is an assurance that a suitably matching distribution of evidence for (or against) its (relevant) constituents may be feasibly acquired) (Wright, 2001, p. 76)

and (ii) on the assumption that there are *quandaries*. Consider now $\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P$, where P is a quandary. By AG^+, the disjunction $\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P$ is known only if we have an assurance that a suitably matching distribution of evidence for its relevant constituents may be acquired. However, Wright argues, if P is a quandary, we do *not* have such an assurance. Therefore, $\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P$, and hence DEC, is presently unknown. On the other hand, if WVER and LNC are in place, we have an assurance that, on pain of contradiction, no suitably matching distribution of evidence may be feasibly acquired for the conjunction $\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P$. Hence,

[47] *Proof*: Assume that $\Diamond\mathcal{K}(\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P)$. By $\mathsf{FAC}_{\Diamond}$, $\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P$. By conjunction elimination, $\neg\Diamond\mathcal{K}P$. By contraposition of WVER, $\neg P$. Similarly for the other conjunct. ∎

courtesy of WVER, $\neg(\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P)$ follows. Thus, DEC* is known. Is this argument correct?

Let us begin with the key epistemic principle here at work: AG$^+$. As Wright himself acknowledges, his formulation of the principle does not apply to compounds statements in which negation is the principal operator (see Wright, 2001, fn. 36, p. 77). Hence, AG$^+$ does not apply, as stated, to either $\neg\mathcal{K}_n$(DEC) or $\neg\mathcal{K}_n$(DEC)*. If Wright's argument is to make headway, therefore, AG$^+$ needs to be adapted to negated statements. Here is one natural option:

(AG$^+_\neg$) $\forall\varphi(\neg\varphi$ is known only if there is an assurance that no suitably matching distribution of evidence for φ may be feasibly acquired).

The revised principle allows us to prove $\mathcal{K}_n$(DEC*): if WVER and LNC hold, we know that no distribution of evidence can make the conjunction $\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P$ true.

There are two problems with this argument. The first is that it relies, once more, on LNC. Yet, we have already seen that, if the anti-realist were allowed to give a proof of DEC* resting on the Law of Non-Contradiction, nothing would prevent us from giving a proof of DEC resting on the Law of Excluded Middle. The second problem is that it is not at all clear that AG$^+$ actually provides a justification for $\neg\mathcal{K}_n$(DEC). For how to interpret AG$^+$, when applied to disjunctions? If we gloss 'there is an assurance' and 'evidence may be feasibly acquired for' as, respectively, 'it is known that' and 'it is possible to know that', there are two possibilities:

$$(\mathsf{AG_a})\ \forall\varphi\forall\psi(\mathcal{K}_n(\varphi \vee \psi) \rightarrow \mathcal{K}_n\Diamond\mathcal{K}\varphi \vee \mathcal{K}_n\Diamond\mathcal{K}\psi);$$

$$(\mathsf{AG_b})\ \forall\varphi\forall\psi(\mathcal{K}_n(\varphi \vee \psi) \rightarrow \mathcal{K}_n(\Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\psi).$$

Let us first consider AG$_b$, as applied to $\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P$. The idea would then be that $\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P$ is known only if $\Diamond\mathcal{K}\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\Diamond\mathcal{K}\neg P$ is. Then, on the plausible assumption that, if A is a quandary, so is $\Diamond\mathcal{K}A$, the quandary view tells us that neither disjunct in $\Diamond\mathcal{K}\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\Diamond\mathcal{K}\neg P$ is presently known. We can thus infer that $\neg\mathcal{K}_n(\Diamond\mathcal{K}\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\Diamond\mathcal{K}\neg P)$. By *modus tollens*, we can finally conclude that $\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P$, and hence DEC, is not presently known either.

This argument cannot be accepted by Wright, however. On a factive interpretation of $\Diamond\mathcal{K}\varphi$, which he explicitly endorses,[48] what the quandary view entitles us to say, namely $\neg\mathcal{K}_n(\Diamond\mathcal{K}\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\Diamond\mathcal{K}\neg P)$, collapses on $\neg\mathcal{K}_n(\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P)$.

[48] See e.g. Wright (2001, pp. 60-61).

Wright's purported justification of $\neg\mathcal{K}_n(\mathsf{DEC})$ would thus be viciously circular: it would precisely assume what it is meant to show!

Wright is therefore left with AG_a as the only available interpretation of AG^+ as applied to disjunctions. But this principle is also problematic: it requires that a disjunction is known only if we already know which is the knowable disjunct. However, in presence of a factive interpretation of $\Diamond\mathcal{K}$, this is tantamount to requiring that a disjunction is known only if at least one of its disjuncts is—something that not even hardened constructivists are ready to accept.

How could the revisionist react? A natural move would be to reject, *pace* Wright, that knowability is factive. But this would not do. To begin with, dropping $\mathsf{FAC}_\Diamond$ might be difficult to swallow for an anti-realist. As Wright himself puts it:

> The obvious question is how abstention from [$\mathsf{FAC}_\Diamond$] might be motivated: is it after all to be allowed that propositions [known] in epistemically ideal circumstances might yet be false? In that case, it would seem, an ideal theory could be false—and how could that admission possibly be reconciled with anything in keeping with the spirit of [anti-realism]? (Wright, 2000, p. 342)

Even more importantly, $\mathsf{FAC}_\Diamond$ surely holds good for *mathematical* statements: if it is possible to prove φ, where φ is some mathematical statement, then φ *is* true. But surely, intuitionists are not willing to lose the opportunity to revise logic within mathematical discourse.

Neither argument, I conclude, offers us independent reasons for accepting $\mathcal{K}_n(\mathsf{DEC}^*)$ that are not as strong as to validate reasons for accepting $\mathcal{K}_n(\mathsf{DEC})$.

2.4.4 Intuitionism and Dialetheism

Let us take stock. The Basic Revisionary Argument is intended to lead the classicist who knows WVER to abandon LEM. However, there is a parallel argument that would lead her to abandon LNC instead. The challenge for the anti-realist is to offer a non-circular way of discriminating between the two arguments, and in particular between $\neg\mathcal{K}_n(\mathsf{DEC})$ and $\neg\mathcal{K}_n(\mathsf{DEC}^*)$.

In the absence of such a way, anti-realists should either reject the Basic Revisionary Argument and the Single Premise Argument, thereby losing two arguments for logical revision, or accept the parallel argument and the truth of $\neg\mathcal{K}_n(\mathsf{DEC}^*)$. The latter option opens up two possible outcomes for anti-realists. They might accept that the Law of Non-Contradiction is presently unknown and, in keeping with Wright's attitude, adopt a logic in which LNC is not a logical law.[49] The upshot of the argument would therefore be a situation which is completely analogous to the intuitionistic case. This would bring to light the logical fact that underlies the analogy between the two arguments, *viz.* the duality of LEM and LNC on the one hand and $\neg\mathcal{K}_n(\mathsf{DEC})$ and $\neg\mathcal{K}_n(\mathsf{DEC}^*)$ on the other.[50] Alternatively, they might refrain from carrying out the last step in the parallel argument and retreat to a paraconsistent logic in which LNC *is* a logical law. That is, they might keep $\mathcal{K}_n(\mathsf{LNC})$ while accepting their commitment to the epistemic possibility of a contradiction, namely $\mathcal{K}_n(\mathsf{DEC}^*) \wedge \neg\mathcal{K}_n(\mathsf{DEC}^*)$. Anti-realists would then become dialetheists.[51]

2.4.5 How Basic is the Single Premise Argument?

If the foregoing considerations are correct, the Basic Revisionary Argument may lead to conclusions that are unwelcome to realists and anti-realists alike. The point carries over to our emendation of what Tennant calls the Single Premise Argument. Recall, the argument is a *reductio* of the claim that the Principle of Bivalence holds, on the assumption that the Manifestation Requirement also holds. Schematically:

$$\cfrac{\begin{array}{c}\text{Bivalence for } D,\ \exists p(p\text{ is undecidable})\\ \Pi_{\Diamond VT}\\ \exists\varphi\in D\ \varphi\text{ is }\Diamond VT\end{array}\qquad \begin{array}{c}\underbrace{\overline{\psi\text{ is }\Diamond VT,}^{\,(1)}\quad \begin{array}{c}\text{Manifestation}\\ \text{Requirement}\end{array}}\\ \Sigma_{\Diamond VT}\\ \bot\end{array}}{\bot}\,(1),$$

where 'φ is $\Diamond VT$' reads 'φ is a potentially verification-transcendent truth'. Just as in the case Basic Revisionary Argument one can substitute LEM and $(\neg\mathcal{K}_n(\mathsf{DEC}))$ with, respectively, LNC and $(\neg\mathcal{K}_n(\mathsf{DEC}^*))$, one can here provide a parallel argument resting on the following three principles: the Manifestation Requirement, the Principle of Contravalence, that no statement is both true and false,

$$(\mathsf{CONTR})\ \neg\exists\varphi(\mathcal{T}\varphi \wedge \mathcal{T}\neg\varphi),$$

[49]Potential candidates include David Nelson's N3, in which LNC fails to be a theorem and negation introduction is not accepted. See Nelson (1949) and Nelson (1959).

[50]For a clear and exhaustive account of the duality of the Law of Non-Contradiction and the Law of Excluded Middle, see Restall (2004).

[51]Kallestrup (2007) offers an argument to the effect that anti-realists are dialetheists. His argument, however, makes use of a self-referential sentence, assumes the factivity of knowability, and requires that if it is possible that somebody knows that p, then somebody possibly knows that p. This latter claim in particular implies, in the presence of WVER, that if there are no knowers every proposition is false, a form of idealism which standard anti-realists are unlikely to endorse. The parallel argument we have presented, on the other hand, appeals to no controversial moves beyond $\neg\mathcal{K}_n(\mathsf{DEC}^*)$.

and the assumption that there are *undecidable** statements, where φ is undecidable* if it satisfies

$$\neg\mathcal{K}_n(\neg(\neg\Diamond\mathcal{K}\varphi \wedge \neg\Diamond\mathcal{K}\neg\varphi)).$$

At a glance, the new parallel argument may be represented as follows:

$$\dfrac{\begin{array}{c}\text{CONTR}, \exists\varphi(\varphi \text{ is undecidable*})\\ \Pi_{\Diamond VT}\\ \exists\varphi \in D\ \varphi \text{ is } \Diamond VT\end{array} \qquad \begin{array}{c}\underbrace{{}^{(1)}\dfrac{}{\psi \text{ is } \Diamond VT} \quad \begin{array}{c}\text{Manifestation}\\ \text{Requirement}\end{array}}\\ \Sigma_{\Diamond VT}\\ \bot\end{array}\ {}_{(1)}}{\bot}.$$

We now need to show that, if contravalence holds, and if there are undecidable* statements, it is epistemically possible that there are verification-transcendent truths. Here is the proof:

> *Proof*: Assume Contravalence. By the Equivalence Thesis, the Law of Non-Contradiction, $\forall\varphi\neg(\varphi \wedge \neg\varphi)$ follows. Now assume that there are undecidable* statements, and let P be one of them. In symbols: $\neg\mathcal{K}_n(\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P)$. Then assume that it is presently known that there are no verifiication-transcendent truths. In symbols: $\mathcal{K}_n\neg\exists\varphi(\varphi \wedge \neg\Diamond\mathcal{K}\varphi)$. By the factivity of knowledge, $\forall\varphi\neg(\varphi \wedge \neg\Diamond\mathcal{K}\varphi)$. By valid intuitionistic reasoning, $\forall\varphi(\varphi \to \neg\neg\mathcal{K}\varphi)$. However, this entails that, for some P, both P and its negation are not verification-transcendent; that is, $\neg(P \wedge \neg\Diamond\mathcal{K}P)$ and $\neg(\neg P \wedge \neg\Diamond\mathcal{K}\neg P)$ hold. These conjunctions intuitionistically entail, respectively, $\neg\Diamond\mathcal{K}P \to \neg P$ and $\neg\Diamond\mathcal{K}\neg P \to \neg\neg P$. Now assume $\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P$. Then, $\neg P \wedge \neg\neg P$ intuitionistically follows. Hence, $\neg(\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P)$. If knowledge is closed under presently known entailment, $\mathcal{K}_n\neg(\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P)$. By arrow introduction, $\mathcal{K}_n\neg\exists\varphi(\varphi \wedge \neg\Diamond\mathcal{K}\varphi) \to \mathcal{K}_n\neg(\neg\Diamond\mathcal{K}P \wedge \neg\Diamond\mathcal{K}\neg P)$. Contradiction. We must therefore negate and discharge our assumption that it is presently known that there are no verification-transcendent truths. In symbols: $\neg\mathcal{K}_n\neg\exists\varphi(\varphi \wedge \neg\Diamond\mathcal{K}\varphi)$. ■

The Principle of Contravalence and the existence of undecidable* statements entail the existence of verification-transcendent truths, which is however incompatible with the Manifestation Requirement. If the latter holds, contravalence must go. In short: the Manifestation Requirement and the assumption that there are undecidable* statements require, *modulo* the Equivalence Thesis, that the Law of Non-Contradiction be given up. But this is of course unacceptable, from an anti-realist standpoint.

Once more, the problem could be solved if the realist's reasons for accepting the existence of undecidable statements, if she has any, were not also reasons for accepting the existence of undecidable* statements. However, it is difficult to see how the realist's reasons for accepting the former claim should not be also reasons for accepting the latter, if the two claims are classically equivalent, and if the realist is not to be asked to weaken her logic *before* as a precondition to engage with her revisionary opponent. On the other hand, if it is thought that the background logic should be the result of intersecting the revisionist's logic with the logic of the non-revisionist, we are faced with the problem that one can only accept $\forall\varphi\neg(\neg\Diamond\mathcal{K}\varphi \wedge \neg\Diamond\mathcal{K}\neg\varphi)$, and reject the existence of undecidable* statements, provided the background logic is strong enough as to contain CONTR. Yet, the validity of CONTR can coherently be questioned, if the validity of BIV can, just as, in the context of the original Basic Revisionary Argument, the validity of both LEM *and* LNC can be questioned. We are thus back to square one: the ensuing dialectic will just repeat the same moves we have examined in the case of the Basic Revisionary Argument. Our emended version of the Single Premise Argument is no more basic than the Basic Revisionary Argument.

2.5 Conclusions

The Basic Revisionary Argument has three main premises: semantic anti-realism, epistemic modesty, and the Law of Excluded Middle. I have argued that the second premise proves problematic. It is ambiguous between at least two readings, $\neg\mathcal{K}_n(\mathsf{DEC})$ and $\neg\mathcal{K}_n(\mathsf{DEC}^*)$. However, the choice between these two readings depends upon prior logical choices about, respectively, LEM and LNC. If we hold on the the latter, we may keep both $\mathcal{K}_n(\mathsf{DEC}^*)$ and $\neg\mathcal{K}_n(\mathsf{DEC})$. Then, our acceptance of WVER would indeed enjoin a rejection of LEM. If, on the other hand, we hold on to LEM, then our acceptance of WVER will enjoin a rejection of LNC instead, or perhaps the acceptance of both $\mathcal{K}_n(\mathsf{DEC}^*)$ and $\neg\mathcal{K}_n(\mathsf{DEC}^*)$. Ditto, *mutatis mutandis*, for our emended version of the Single Premise Argument.

How are we to make these logical choices? How can we rationally decide between dual logical principles, such as LEM and LNC? More generally: how are we to revise some logical principles at the expenses of others? Pending a satisfactory answer to these questions, we cannot but conclude that the Basic Revisionary Argument has an uncertain output: for all we know, it may turn realists into either intuitionists, or paraconsistentists, or Dialetheists. The point carries over to the Single Premise Argument.

Anti-realists might object that, on an *inferentialist* approach to logic, our parallel arguments are blocked, but the original ones are not. On the inferentialist view, the meaning of a logical constant $ is fully determined by its operational rules (in a natural deduction system, its introduction and its elimination rules), provided that these rules satisfy some independently motivated proof-theoretic constraints. A familiar inferentialist complaint about classical logic, as we shall see in Chapter 5 and Chapter 6, is that the inference rules that are need for proving LEM do *not* respect these proof-theoretic constraints, contrary to those that are needed for proving LNC. As a result, the anti-realist might conclude that we *do* have reasons for accepting $\neg\mathcal{K}_n(\mathsf{DEC}^*)$ while rejecting $\neg\mathcal{K}_n(\mathsf{DEC})$: $\mathcal{K}_n(\mathsf{DEC})$ is only provable if WVER and LEM are known, whereas $\mathcal{K}_n(\mathsf{DEC}^*)$ merely requires knowledge of WVER and LNC. Hence, if LNC is known, but LEM is not, DEC* is after all known, while DEC is not (similar considerations apply to the Single Premise Argument). This line of response, however, requires that an inferentialist approach to logic validates intuitionistic logic, but not classical logic—in particular, that it validates LNC but not LEM. We shall attempt to shed some light on this vexed question in Chapter 7.

For the time being, we will turn in the next chapter to a second possible argument from the knowability of truth to the adoption of intuitionistic logic: the Church-Fitch Paradox of Knowability.

Chapter 3

The Paradox of Knowability

There are many things nobody will ever know. Nobody will ever know how many leaves there are on the tree in front of my window. Nobody will ever know the number of occurrences of the letter 'a' in the books that I have in my library. Nobody will ever know how many hairs there are now on the top of my head. Nobody bothers to count, and, it seems reasonable to assume, nobody ever will. Our topic in this chapter is a very simple argument to the effect that for every such point of contingent ignorance, there is a point of *necessary* ignorance: for every truth φ that nobody will ever know, there is, as a matter of logic, a truth that nobody *can* know, namely the truth that φ is true and nobody will ever know it. The argument was first published by Frederic Fitch in 1963 as Theorem 5 of his paper *A Logical Analysis of Some Value Concepts*. The bulk of the proof, however, was first discovered by Alonzo Church in 1945.[1] The contrapositive of Theorem 5 is known as the *Paradox of Knowability*: if all truths are knowable, then all truths will be known by someone at some time. Since its rediscovery by William Hart and Colin McGinn in 1976,[2] the Paradox has plagued metaphysical doctrines committed to the knowability of truth, such as semantic anti-realism. For of course, it would seem, *there are* truths nobody will ever know. My main focus will be on so-called *intuitionistic* treatments of the paradox—treatments that have been influentially recommended by, among others, Michael Dummett and Timothy Williamson. Dummett writes:

> What is wrong with [Fitch's reasoning]? The fundamental mistake is that the justificationist does not accept classical logic. (Dummett, 2009, p. 2)

[1] Fitch credits the bulk of the proof to an anonymous referee of an unpublished paper of 1945. We now know that the referee in question was Alonzo Church. See Church (2009).

[2] See Hart and McGinn (1976).

In a similar spirit, Williamson has argued that the availability of an intuitionistic treatment of the Church-Fitch proof transforms the proof—a would be refutation of semantic anti-realism—in an argument for the adoption of intuitionistic logic:

> If anti-realism is defined as the principle that all truths are knowable, then anti-realists have a reason to revise logic. For an argument first published by Fitch seems to reduce anti-realism to absurdity within classical but not constructivist logic. (Williamson, 1988, p. 422)

Crispin Wright also seems sympathetic to a revisionary approach to the Church-Fitch puzzle. In his *Realism, Pure and Simple?*, he observes that "classical logic must be casualty in any region of discourse where truth is held to be epistemically constrained but it is acknowledged that not all issues a guaranteed to be (weakly) decidable". He then points out that, if these two conditions apply to the Church-Fitch conjunction $P \wedge \neg \mathcal{K}P$, then "the final step of the [Church-Fitch] argument would be undermined" (Wright, 2003a, p. 69).[3]

In this chapter, I will argue that intuitionistic treatments of the Church-Fitch problem are not very promising. Hence, I will suggest, the Church-Fitch argument is not by itself a reason for adopting intuitionistic logic.

3.1 The Church-Fitch argument

The Church-Fitch argument purports to show that semantic anti-realism, the view that all truths are knowable, collapses into a naïve form of idealism, according to which all truths will be known by someone at some time. The original proof published by Fitch establishes the following theorem:

> (T5) If there is a proposition which nobody knows (or has known or will know) to be true, then there is a proposition which nobody can know to be true. (Fitch, 1963, p. 139)

Formally:

$$\text{(T5) } \exists\varphi(\varphi \wedge \neg\mathcal{K}\varphi) \rightarrow \exists\varphi(\varphi \wedge \neg\Diamond\mathcal{K}\varphi).$$

The *Paradox of Knowability* is the contrapositive of T5:

[3] Wright's final word on the matter is that anti-realists should give up "the vague idea that truth is somehow intimately connected with justification—that it cannot 'totally outrun' it" (Wright, 2003a, p. 304), and adopt instead a weaker epistemic account of truth—one that is consistent with the existence of 'blindspots'. I don't have time to discuss here Wright's own approach—I do so in Murzi (2008). It is worth pointing out, though, that, if correct, Wright's approach indirectly confirms the main thesis advanced in the present chapter, viz. that weakening the logic does not get to the heart of the problem.

(CT5) If all truths are knowable, then all truths are known.

Since the consequent of CT5 is clearly false, the anti-realist claim that

(KP) All truths are knowable

is under threat. Now to the details of the proof.

3.1.1 The proof

Call the most straightforward formalization of the CT5 antecedent of CT5 *weak verificationism*:

$$\text{(WVER)}\ \ \forall\varphi(\varphi \rightarrow \Diamond\mathcal{K}\varphi).$$

And call the formalization of its consequent *strong verificationism*:

$$\text{(SVER)}\ \ \forall\varphi(\varphi \rightarrow \mathcal{K}\varphi).$$

The Paradox of Knowability assumes that knowledge is necessarily factive and closed under conjunction elimination. In symbols:

$$\text{(FACT)}\ \ \Box\forall\varphi(\mathcal{K}\varphi \rightarrow \varphi);$$

$$\text{(DIST)}\ \ \Box\forall\varphi\forall\psi\, \mathcal{K}((\varphi \wedge \psi) \rightarrow (\mathcal{K}\varphi \wedge \mathcal{K}\psi)).^4$$

The argument further requires that provable formulas are necessary, and that necessary falsehoods are impossible. In symbols, respectively:

(NEC) If $\vdash A$, then $\Box A$;

(ME) From $\Box\neg A$, infer $\neg\Diamond A$.

It may be presented in two steps. One first proves that sentences of the form $P \wedge \neg\mathcal{K}P$ are unknowable, for any particular P:

(1)	$\mathcal{K}(P \wedge \neg\mathcal{K}P)$	Assumption for $\neg$-I
(2)	$\mathcal{K}P \wedge \mathcal{K}\neg\mathcal{K}P$	1, DIST
(3)	$\mathcal{K}P \wedge \neg\mathcal{K}P$	2, FACT
(4)	$\neg\mathcal{K}(P \wedge \neg\mathcal{K}P)$	1-3, $\neg$-I
(5)	$\Box\neg\mathcal{K}(P \wedge \neg\mathcal{K}P)$	4, NEC
(6)	$\neg\Diamond\mathcal{K}(P \wedge \neg\mathcal{K}P)$	5, ME

[4] Alternatively, we may frame these two principles as inference rules as follows:

$$\text{(FACT)}\ \frac{\mathcal{K}A}{A}; \qquad \text{(DIST)}\ \frac{\mathcal{K}(A \wedge B)}{\mathcal{K}A \wedge \mathcal{K}B}.$$

One then proceeds to show that, given (6), WVER collapses into SVER (PC below abbreviates 'Propositional Calculus'):

(WVER)	$\forall\varphi(\varphi \rightarrow \Diamond\mathcal{K}\varphi)$	Assumption for $\rightarrow$-I
(8)	$(P \wedge \neg\mathcal{K}P) \rightarrow \Diamond\mathcal{K}(P \wedge \neg\mathcal{K}P)$	WVER, $\forall$ E
(9)	$P \wedge \neg\mathcal{K}P$	Assumption for $\neg$-I
(10)	$\Diamond\mathcal{K}(P \wedge \neg\mathcal{K}P)$	8, 9, $\rightarrow$-E
(11)	$\Diamond\mathcal{K}(P \wedge \neg\mathcal{K}P) \wedge \neg\Diamond\mathcal{K}(P \wedge \neg\mathcal{K}P)$	10, 6, $\wedge$-I
(12)	$\neg(P \wedge \neg\mathcal{K}P)$	9-11, $\neg$-I
(13)	$\neg\mathcal{K}P \rightarrow \neg P$	12, PC (Intuitionistic)
(14)	$P \rightarrow \neg\neg\mathcal{K}P$	13, PC (Intuitionistic)
(15)	$P \rightarrow \mathcal{K}P$	14, PC (Classical)
(SVER)	$\forall\varphi(\varphi \rightarrow \mathcal{K}\varphi)$	13, $\forall$-I
(CT5)	$\forall(\varphi \rightarrow \Diamond\mathcal{K}\varphi) \rightarrow \forall\varphi(\varphi \rightarrow \mathcal{K}\varphi)$	WVER-SVER $\rightarrow$-I

From WVER and the seemingly innocuous assumption that some truths are forever unknown, a contradiction follows: we cannot know, for any given truth, that it is forever unknown, on pain on contradiction. Something must go. The anti-realist will discharge the second assumption, thereby committing herself, by an exclusively classical step, to SVER. Yet, it would seem, SVER is plainly false. Hence, WVER should be regarded as false as well. As Colin McGinn and William Hart put it:

> In presence of obvious truths, [SVER] is deducible from [WVER]. [But] [SVER] is obviously false and is an objectionably strong thesis of idealism [...]. Therefore [WVER] is false: there are truths which absolutely cannot be known. (Hart and McGinn, 1976, p. 139)

The proof appears to be valid in classical modal logics as strong as K, for any factive operator closed under conjunction elimination.[5] The minimal modal principles it requires cannot be reasonably questioned. Likewise, it would be terribly hard to deny that knowledge is factive and distributes under conjunction.

[5] Where K is the logic obtained by adding to classical logic NEC and the axiom: (K) $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$.

3.1.2 Possible ways out: a shopper's guide

The Church-Fitch argument does not *obviously* undermine semantic anti-realism. In Chapter 2, we defined semantic anti-realism as the thesis that truth is epistemically constrained. Formally:

$$\text{(EC)}\ \forall\varphi(\varphi \rightarrow \mathcal{E}\varphi),$$

where '$\mathcal{E}$' is an epistemic predicate of some sort. Following the majority of anti-realists philosophers, we interpreted '$\mathcal{E}\varphi$' as 'φ is possibly known by someone at some time', and we identified semantic anti-realism with what we called the Knowability Principle, the thesis that truth must be knowable:

(KP) For all φ, if φ, then it is possible to knowφ.

The principle is explicitly endorsed by leading anti-realists. Here is, for instance, Dummett:

> [KP] is a regulative principle governing the notion of truth: if a statement is true, it must be in principle possible to know that it is true. (Dummett, 1976, p. 98)

And here is Wright:

> The distinctive anti-realist thesis [...] is that truth must be epistemically constrained, cannot be evidence-transcendent. So some principle will be endorsed of the form 'Evidence is available for my proposition which is true'; more formally, $\forall\varphi(\varphi \rightarrow \Diamond\mathcal{K}\varphi)$. (Wright, 1987, p. 310)

In turn, we logically interpreted KP as WVER: the claim that, if φ is true at world w, then there is a world x accessible from w where somebody knows at some time that φ is true at x. If classical logic holds, and if we grant the epistemic principles FACT and DIST, together with the modal rules NEC and ME, WVER effectively collapses on SVER.

Notice, though, that a number of steps were required in order to reach this conclusion. We had to assume that:

(i) Semantic anti-realism entails the knowability of truth;

(ii) The logical form of the Knowability Principle is correctly captured by WVER;

(iii) WVER holds unrestrictedly;

(iv) The Church-Fitch proof is classically valid;

(v) Anti-realists are willing to apply exclusively classical rules in the derivation of the Church-Fitch proof;

(vi) The anti-realist is willing to accept the epistemic and modal principles required by the proof.

Perhaps unsurprisingly, each of these steps has been questioned. Cesare Cozzo, Dag Prawitz, Michael Hand, and Carrie Jenkins, among others, reject (i). In their view, semantic anti-realism does not require that truth be knowable. Hence, they conclude, the Church-Fitch proof should be no cause of concern for the anti-realist philosopher.[6] Dorothy Edgington has influentially questioned (ii). According to her, KP is ambiguous between a paradoxical reading, the one exploited in the Church-Fitch proof, and a second reading, which is not obviously paradoxical.[7] Neil Tennant and Michael Dummett have, among others, questioned (iii).[8] They suggest that WVER should be appropriately restricted to a class of non-problematic truths. Jon Kvanvig rejects (iv) and (vi). He claims that the Church-Fitch proof is *classically* invalid, because of the failure of some of the modal and epistemic principles it involves.[9] He presses us to adopt a theory of propositional content that does not validate NEC, and hence blocks the derivation of Church-Fitch's key theorem, that propositions of the form $P \wedge \neg \mathcal{K}P$ are unknowable. Timothy Williamson and, most recently, Michael Dummett, have both questioned (v). Their suggestion is that the proof leaves the *intuitionist* anti-realist unscathed.[10]

For reasons of space, I will focus my attention on the intuitionistic approaches favoured by Williamson and Dummett. We will first consider Dummett's intuitionistic restriction of WVER to *basic* statements, as proposed in Dummett (2001). We will then turn to the hypothesis that the problem raised by Church-Fitch is exclusively owed to the *logic* in which the paradox is run, rather than to the falsity of its main metaphysical premise (WVER). More specifically, we shall examine the possibility that semantic anti-realism can be salvaged from the Church-Fitch threat by either (a) giving up the Law of Excluded Middle or (b) reflecting upon the intuitionistic understanding of the logical constants. Either strategy, if successful, would transform the Church-Fitch argument into an argument from the knowability of truth to the adoption of intuitionistic logic. My main claim will be that there is a tension between the intuitionist's claim that the Paradox of Knowability, and

[6] See Cozzo (1994), Prawitz (1998b), Hand (2003) and Hand (2009), Jenkins (2007).

[7] See Edgington (1985) and Edgington (2010).

[8] See e.g. Tennant (1997) and Dummett (2001).

[9] See Kvanvig (1995) and Kvanvig (2006).

[10] See e.g. Williamson (1982) and Dummett (2009). But see also De Vidi and Solomon (2001) and De Vidi and Solomon (2006).

related knowability proofs, are blocked in intuitionistic logic, and her contention that intuitionistic negation applies to mathematical and empirical statements alike. I examine in Appendix C a proposal by David De Vidi and Graham Solomon's for solving the problem by introducing an empirical negation in the language of intuitionistic logic.

3.2 Victor's anti-realism

The key thought behind the so-called restriction strategies is to treat sentences such as $A \wedge \neg \mathcal{K}A$ as exceptional cases, to which the Knowability Principle need not apply. The approach, already foreshadowed in Church (2009), has been recently recommended by Dummett, in a short piece entitled *Victor's Error.*

3.2.1 Dummett on anti-realism and basic statements

According to Dummett (2001), the anti-realist's mistake was "to give a blanket characterization of truth, rather than an inductive one" (Dummett, 2001, p. 1). Let 'Tr' be a truth predicate. Then, Dummett suggests, the anti-realist, he calls him Victor, could offer the following inductive characterization of truth:

(i) $\mathrm{Tr}(A)$ iff $\Diamond\mathcal{K}(A)$, if A is a basic statement;

(ii) $\mathrm{Tr}(A$ and $B)$ iff $\mathrm{Tr}(A) \wedge \mathrm{Tr}(B)$;

(iii) $\mathrm{Tr}(A$ or $B)$ iff $\mathrm{Tr}(A) \vee \mathrm{Tr}(B)$;

(iv) $\mathrm{Tr}($if A, then $B)$ iff $(\mathrm{Tr}(A) \rightarrow \mathrm{Tr}(B))$;

(v) $\mathrm{Tr}($it is not the case that $A)$ iff $\neg\mathrm{Tr}(\mathrm{A})$;

(vi) $\mathrm{Tr}(A(\text{something}))$ iff $\exists x \mathrm{Tr}(A(x))$;

(vii) $\mathrm{Tr}(A(\text{everything}))$ iff $\forall x \mathrm{Tr}(A(x))$,

where, Dummett writes, "the logical constants on the right hand-side of each clause is understood as being subject to the laws of intuitionistic logic" (Dummett, 2001, p. 2). Dummett does not offer, nor attempts to offer, a definition of a basic statement. However, he says, "the principle is clear". On these assumptions, the Church-Fitch conjunction $P \wedge \neg \mathcal{K}P$ does not pose any problem for Victor. It can nevertheless be true, provided that its basic components are knowable. If Victor is more careful, Dummett writes, "he can easily avoid the appearence of putting forward an incoherent conception of truth" (*Ibid.*).

3.2.2 A weaker Manifestation Requirement

Dummett's proposal must pass a threefold test: first, it has to be motivated in a non *ad hoc* way; second, it has to block further potential Fitch-like paradoxes; third, it must deliver an anti-realistically acceptable notion of truth. I will set the second requirement aside for present purposes.[11] Concerning the first, Dummett does not even attempt to motivate his own restriction,[12] However, I shall assume for the sake of argument that the restriction *can* be motivated. For instance, one might argue that it is validated by the following weaker version of the Manifestation Requirement:

(WM) A speaker S manifests understanding of a sentence A if,

(i) if A is basic, she is disposed to recognize a proof (disproof) of A when presented with one,

(ii) if A is not basic, she is *either* disposed to recognize a proof (disproof) of A if presented with one, *or* (a) she is disposed to recognize proofs (disproofs) of A's basic components if presented with them, and (b) she can manifest her understanding of whatever logical constant \$ may occur in A by correctly using \$ in other compound sentences.

Dummett's may be taken to be validated by the foregoing formulation of the Manifestation Requirement. Consider, for instance, the Church-Fitch sentence. On Dummett's account, supplemented by the foregoing version of the Manifestation Requirement, although the sentence may turn out to be unknowable, anti-realists may nevertheless have the resources to account for the manifestability of its meaning. They could argue that knowledge of $P \wedge \neg \mathcal{K}P$'s meaning can be manifested by manifesting knowledge of the meanings of its component expressions—namely, 'P', '$\mathcal{K}$', '$\wedge$' and and '$\neg$'. What remains to be seen is whether Dummett's proposed restriction is a viable response to the Church-Fitch problem.

3.2.3 Williamson on basic statements

Timothy Williamson has recently argued that Dummett's restriction is at odds with semantic anti-realism. He writes, in footnote n. 5 of his *Tennant's Troubles*:

[11]For some objections along these lines, see Brogaard and Salerno (2002).

[12]But see Bermudez (2009) for an attempt to motivate it via an argument from the assumption that the concept of proposition is *indefinitely extensible*.

> This restriction is hard to reconcile with Dummett's original motivation for the Knowability Principle, a motivation that applies to complex sentences just as much as to atomic ones. It will not do to say that the use of complex sentences is indirectly epistemically grounded because their atomic constituents are. For connectives such as conjunction and negation are used as constituents of complex sentences, not by themselves. Thus any epistemic grounding of the use of connectives must derive from an epistemic grounding of complex sentences in which they occur, not *vice versa*: yet Dummett's strategy against Fitch is just to avoid any such direct epistemic grounding of the use of complex sentences. Thus his anti-realism unravels. (Williamson, 2009b, p. 187, fn. 5)

The anti-realist, Williamson argues, cannot insist that Dummett's restriction delivers an epistemic notion of truth on the grounds that (i) atoms are knowable, and that (ii) truth for compound statements is inductively defined via the recursive clauses for the intuitionist logical constants. For, Williamson alleges, any reason one might have for adopting an epistemically grounded account of the logical constants would have to derive from considerations applying to the *statements* in which the constants may themselves occur. But, if this is correct, the anti-realist who wishes to assign a constructive meaning to the logical constants needs to assume that true statements, *of whatever logical complexity*, must be knowable, contradicting Dummett's concession that non-basic statements may be unknowable.

Tennant (2002) makes a related point. If knowability is restricted to atomic arithmetical statements, he argues, the anti-realist would not be in a position to convince the classical mathematician to adopt intuitionistic logic. Tennant writes:

> By having confined the knowability principle to atomic statements, it would appear that Dummett has foregone the most important principled way for the anti-realist to argue against the illicit application of strictly classical rules of inference. No longer is he requiring of *every* proposition of arithmetic that, if it is true, then it is knowable. The suggestion that Victor restrict the knowability requirement to just the atomic truths of arithmetic happens to fall on very attentive ears on the part of his classically-inclined interlocutor. There is no longer any principled ground on which the latter can be enjoined not to treat the logical operators $\neg$, $\vee$ and $\exists$ in the non-constructive way that he does. (Tennant, 2002, p. 141)

On Dummett's restriction, Tennant surmises, it is not immediately clear that classicists would still have reasons not to interpret the logical constants the way they do.

Both objections are problematic, however. As we shall see in the second part of this thesis, it is simply not true that one can claim that the logical constants are epistemically grounded *only* on the grounds that so are the statements in which they may occur. As Williamson and Tennant well know, there are *other* arguments for adopting an epistemically grounded interpretation of the logical constants—arguments that do not rely on the assumption that truth is in general epistemically constrained. To mention but two: the inferentialist arguments we will examine in Chapter 5 and 6, and Crispin Wright's argument from *vagueness* to the rejection of Double Negation Elimination.[13]

To be sure, it may well be that these arguments do not succeed. But this is something that needs to be shown—it may not be merely assumed. What is more, if anti-realists interpret the Manifestation Requirement in the weak way I have suggested earlier (roughly, that in order to manifest knowledge of the meaning of a complex sentence A, one merely needs to manifest knowledge of A's components), Dummett's restriction *would* be validated by "Dummett's original motivation for the Knowability Principle", contrary to what Williamson alleges.

3.2.4 Basic statements and logical revision

Both Williamson and Tennant may be wrong in letter, but they are correct in spirit. They correctly diagnose a tension between Dummett's restriction of the Knowability Principle and *one* argument for intuitionistic logic. At a closer look, it is indeed easy to see that *Dummett's proposed restriction invalidates the Basic Revisionary Argument*. For consider Dummett's restricted version of the Knowability Principle. In order for it to figure as the main metaphysical premise of the Basic Revisionary Argument, the second key premise of the argument,

$$(\neg\mathcal{K}_n(\mathsf{DEC}))\ \neg\mathcal{K}_n\forall\varphi(\Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\neg\varphi),$$

would have to be restricted to basic statements too, as follows:

$$(\neg\mathcal{K}_n(\mathsf{DEC}_{\mathbf{B}}))\ \forall\varphi(\mathbf{B}\varphi \rightarrow \neg\mathcal{K}_n(\Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\neg\varphi)),$$

[13] See e.g. Wright (2003b) and Wright (2007b).

where '$\mathbf{B}\varphi$' reads 'φ is basic'. However, it is unclear whether this holds. Dummett's examples of undecidable statements all involve *non-basic* statements. Williamson nicely makes the point:

> Note [...] that his original (1959) examples of sentences that the realist contentiously treated as verification-transcendent involved complex constructions such as universal quantification and the counterfactual conditional: 'A city will never be built on this spot' and 'If Jones had encountered danger, he would have acted bravely' are not atomic sentences. (Williamson, 2009b, p. 187, fn. 5)

Dummett's original examples of sentences that are not guaranteed to be decidable are arguably *not* basic. Hence, pending further argument, Dummett's restriction requires that intuitionists give up one of their main arguments for revising logic.

3.2.5 Summing up

Dummett's restriction appears to face serious philosophical problems. For one thing, it seems unmotivated: Dummett does not offer any argument for imposing his proposed restriction. For another, it violates Dummett's original Manifestation Challenge. Knowledge of $P \wedge \neg \mathcal{K}P$'s meaning cannot be manifested by being disposed to recognize a correct argument for it: after all, there cannot be any such argument in the first place. Granted, anti-realists might adopt a weaker requirement of manifestability, according to which, in order to manifest knowledge of the meaning of a complex sentence A, one merely needs to manifest knowledge of A's components. Then, only 'basic' statements need to be knowable. However, this yields an epistemic constraint on truth that is too weak to serve as a premise of the Basic Revisionary Argument. Anti-realists, I suggest, should look for different solutions to the Church-Fitch problem.

3.3 A seemingly trivial way out

Let us now turn to a different, and little explored, intuitionistic solution to the Paradox of Knowability.[14] Classicists and intuitionists assign different meanings to the logical constants. Classicists take them to express truth-functions. Intuitionists identify their meaning with their contribution to the *proof-conditions* of the complex sentences in which they may occur. In the case of the conditional, we are told that

[14]Section 3.3 and 3.4 are extracted from Murzi (2009).

a proof of $P \to Q$ is a method which evidently transforms any proof of P into a proof of Q. On this reading, strong verificationism says that any proof of any arbitrary sentence P can be turned into a proof of $\mathcal{K}P$. But is not this acceptable, after all? If one proves P, then one can also *know*, on mere reflection, that P has been proved. As William Hart puts it:[15]

> Suppose we are given a sentence [...] and a proof that it is true. Read the proof; thereby you come to know that the sentence is true. Reflecting on your recent learning, you recognize that the sentence is now known by you; this shows that the truth is known. (Hart, 1979, p. 165)

Enrico Martino and Gabriele Usberti (1994) suggest that this provides a "trivial" solution to the Church-Fitch problem:

> strong verificationism [...] can be interpreted only according to the meaning of implication, so that it expresses the trivial observation that, as soon as a proof of P is given, P becomes known. (Martino and Usberti, 1994, p. 90; their terminology is adapted to ours.)

Drew Khlentzos makes essentially the same point:

> the puzzle with Fitch's argument for the antirealist is this: [...] "$(P \to \mathcal{K}P)$" [...] is perfectly acceptable *if interpreted in the intuitionistic way*. [...] How then can Fitch's argument be thought to "refute" [this principle]? (Khlentzos, 2004, p. 180)

3.3.1 Proof-types and-proof tokens

Much depends on what the intuitionist means here by proof, however. Some intuitionists, like Dag Prawitz, identify proofs with some kind of Platonic objects, outside of space and time. Prawitz writes:[16]

> that a sentence is provable is here to mean simply that there is proof of it. It is not required that we have actually constructed the proof or that we have a method for constructing it. (Prawitz, 1998b, p. 287)

> A sentence is true if and only if [...] there is a proof of it [...] in an abstract, tenseless sense of exists. (Prawitz, 1998a, p. 297)

[15]See also Williamson (1988, p. 429).
[16]See also Hand (2003) and Hand (2009).

If proofs are abstract objects, $P \to \mathcal{K}P$ is *not* validated by the intuitionistic semantics for the conditional: from the fact that there is, in a abstract and tenseless sense of 'exists', a proof π of P, nothing follows about the *actual construction* of π. Hence, not every Platonic proof of P can be transformed into a Platonic proof of $\mathcal{K}P$.

Of course, an intuitionist might object that Platonism about proofs is not available to an *anti-realist*. For is not the notion of a Platonistic proof an inherently realist one?[17] And surely, the objector might insist, once proofs are identified with *tokens*, instead of Platonic *types*, $P \to \mathcal{K}P$ is indeed validated by the intuitionistic interpretation of the conditional. This line of reply faces problems from at least two different scores. First, a conception of proofs as types need not enjoin commitment to a Platonic realm of proofs. Second, there are well-known difficulties in identifying proofs with actually existing tokens. Let us have a closer look.

3.3.2 Proofs as Aristotelian types

In his (1988), Williamson urges intuitionists to adopt a broadly *Aristotelian* conception of proofs. According to him, they should identify proofs with types, and define proof-types in terms of structural identity of proof-tokens, "where two proof-tokens of the same type are required to have identical conclusions and structure, but need not occur at the same time" (Williamson, 1988, p. 430). On this view, talk of proof-types can always be reduced to talk of proof-tokens: it carries no commitment to Platonic objects. Proofs-tokens of the Pythagorean theorem carried out at different times would count as the same proof-type, provided only that they have the same structure. But there would be no such thing as a proof-type of the Pythagorean theorem, if there were no proof-tokens of it. On the foregoing assumptions, Williamson suggests that intuitionists interpret the conditional as a function f from proof-tokens to proof-tokens of a special kind, "one that is unitype in the sense that if π and ρ are proof-tokens of the same type then so are $f(\pi)$ and $f(\rho)$" (1988, p. 430). Then,

> a proof of $P \to \mathcal{K}P$ is a unitype function that evidently takes any proof-token of P to a proof-token, for some time t, of the proposition that P is proved at t. (*Ibid.*)

[17] See e.g. Dummett (1987a, p. 285). I for one do not think this a very serious problem. As Cesare Cozzo (1994, p. 77) observes, the standard intuitionistic argument for rejecting Bivalence holds even if proofs are conceived of as Platonic objects—after all, we have no guarantee that there is either a Platonic proof, or a Platonic disproof, of Goldbach's Conjecture. If Bivalence is necessary for semantic realism, then a conception of truth as the existence of a Platonistic proof counts as an anti-realist one. See also Prawitz (1998b, p. 289) for a response to an argument by Dummett (1987, 1998) to the effect that Platonism about proofs enjoins commitment to the Principle of Bivalence.

Williamson observes that intuitionists are *not* committed to the existence of such a function. That is, the Aristotelian conception of proofs does not validate strong verificationism. His argument runs thus. If P has already been decided, then every proof-token of P can be transformed into a proof-token that P is proved at some time t, and $P \rightarrow \mathcal{K}P$ indeed holds. If P has *not* yet been decided, however, we can only consider *hypothetical* proof-tokens of P. Now let π be a hypothetical proof-token of P carried out on Monday, and let ρ be a hypothetical proof-token of P carried out on Tuesday. Then, Williamson argues, the function f that transforms π and ρ into proof-tokens of $\mathcal{K}P$ is not guaranteed to be of a *unitype* kind. For $f(\pi)$ and $f(\rho)$ are now proofs of, respectively, the proposition that P is proved *on Monday* and of the proposition that P is proved *on Tuesday*. They cannot be of the same type, since their conclusion differ.

3.3.3 Truth and provability

Martino and Usberti have advanced the following objection to Williamson's argument. If a proof of a conditional is to be regarded as a function at all, then it should map *real* proof-tokens of the antecedent into a proof-token of the consequent, not merely *hypothetical* ones. As they put it:

> the required function f is not expected to operate on the hypothetical proof-token: such an object does not exist! Its arguments cannot be but *given* proof-tokens; as long as no proof of P is known, f has nothing to map. So we can still define f as the constant function which, once a proof π of P is known, maps every proof ρ of P into the proof that $\mathcal{K}P$ is known at time $t(\pi)$. (Martino and Usberti, 1994, p. 91)

This objection is mistaken, however. To begin with, on Martino and Usberti's interpretation of the intuitionistic conditional, one could assert $P \rightarrow Q$ only if one had a proof of P. But this seems odd: in many circumstances, we assert conditionals without knowing whether their antecedent is true. Second, if functions could only map *given* proof-tokens, intuitionists could not interpret $\neg P$ the way they do, i.e. as $P \rightarrow \bot$, where '$\bot$' expresses a constantly false proposition. For a proof of $P \rightarrow \bot$ is a function g which evidently maps any proof of P into a proof of $\bot$. Yet there cannot be a *proof*-token of $\bot$! Nor can there be a *proof*-token of P, if it is false. It follows that g can only map *hypothetical* proof-tokens, contrary to what Martino and Usberti assume.

Intuitionists who are willing to trivially solve the Paradox of Knowability along Martino and Usberti's lines must thus reject Williamson's proposed interpretation of the conditional as a *unitype* function. They are forced to identify proofs with proof-*tokens*, and insist that every proof-token of P can be transformed into a proof-token of $\mathcal{K}P$. There are, however, reasons for thinking that intuitionists may not plausibly conceive of proofs this way.

The main problem is that they equate truth with the existence of a proof. If proofs are temporal objects, therefore, so is truth. Dummett has himself pointed out some rather counterintuitive consequences of the view.[18] Suppose that P intuitionistically follows from Γ, and that all the sentences in Γ have a proof-token. Furthermore, suppose that there is no proof-token of P. We may then have the following situation: all the sentences in Γ are true, but P is not. If validity requires preservation of truth, it follows that the inference from Γ to P is not valid after all, contrary to what we had assumed. Another difficulty concerns disjunctions. Suppose truth is identified with the existence of a proof-token, or with the actual possession of a means of constructing one. Then, any disjunction with unknown disjuncts will have *untrue* disjuncts. But how can a *true* disjunction have untrue disjuncts? There is finally a well-known problem with past-tensed statements. If truth is equated with the existence of a proof-token, past-tensed statements for which all the evidence has been lost may become *untrue*. Dummett has recently come to regard the view as "repugnant".[19] Truth, he now thinks, is something that cannot be gained, or lost. He writes:

> I now believe that a proposition, whether about the past, the future or the present, is true, *timelessly*, just in case someone optimally placed in time and space could have, or could have had, compelling grounds for recognizing it as true—that is to say, if such compelling evidence would be or have been available to him. (Dummett, 2006, p. x; italics added)

If we are to give credit to Dummett's own arguments, truth cannot be identified with the existence of actual proof-tokens. The consequences of doing so are no less paradoxical than the claim that all truths will be known at some time. Yet, it appears that the Paradox of Knowability can only be 'trivially' solved if proofs are conceived of as proof-tokens. If there is an intuitionistic solution to the Paradox of Knowability, it must be found elsewhere.

[18]See Dummett (1973b, pp. 239-43).
[19]See Dummett (2004) and Dummett (2006).

3.4 The standard intuitionistic response

The following alternative strategy suggests itself. As Williamson first pointed out, the Church-Fitch proof is intuitionistically *invalid*. All that intuitionistically follows from Weak Verificationism, is what we may label *Intuitionistic Verificationism*:

$$(\mathsf{IVER})\ \forall\varphi(\varphi \rightarrow \neg\neg\mathcal{K}\varphi).$$

Unlike the claim that all truths will be known, however, IVER is not obviously problematic:

> it forbids intuitionists to produce claimed *instances* of truths that will never be known: but why should they attempt something so foolish? (Williamson, 1982, p. 206)

3.4.1 Dummett's new line

Dummett himself has recently suggested that IVER, as opposed to WVER, expresses the correct formalisation of the conceptual connection between truth and knowledge. He writes, in a recent reply to Wolfgang Künne:

> 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, 2007c, p. 348)

Rather, Dummett now claims, what is wrong with the realist's use of the Paradox of Knowability as a counterexample to anti-realism is that *intuitionist* anti-realists need not accept the conclusion of the argument. He writes:

> what is wrong with [the Paradox of Knowability]? The fundamental mistake is that the justficationist does not accept classical logic. He is happy to accept [IVER], provided that the logical constants are understood in accordance with intuitionistic rather than classical logic. In fact [...] he will prefer IVER to WVER as a formalisation of his view concerning the relation of truth to knowledge. (Dummett, 2009, Dummett's terminology is adapted to ours)

It is however unclear whether the adoption of intuitionistic logic may itself solve the problem raised by Church-Fitch. For although IVER might be acceptable, *other* intuitionistic consequences of weak verificationism already seem worrisome. Intuitionists, for instance, are still committed to

$$(16)\ \forall\varphi\neg(\varphi \wedge \neg\mathcal{K}\varphi),$$

which is tantamount to denying the highly plausible claim that there exist forever-unknown truths.[20]

Dummett has recently objected that intuitionists cannot even *hear* the problem.[21] When intuitionists assert (16), he writes,

> it is not being asserted that there cannot be a true statement which will not *in fact* ever be known to be true: this "in fact" expresses a realist understanding of universal quantification as infinite conjunction and is therefore constructively unintelligible. (Dummett, 2007a, p. 348)

Rather, in Dummett's view, intuitionists can legitimately assert $\neg\mathcal{K}\varphi$ only if there is an obstacle in principle to our coming to know φ:

> intuitionistically interpreted, "$\forall t\neg\mathcal{K}(\varphi, t)$" holds good only if there is a general reason why it cannot be known at each time t that φ. (Dummett, 2009, p. 52)

But this can only mean that intuitionists can assert $\neg\mathcal{K}\varphi$ only if they are in a position to assert that φ is *false*. No wonder, then, that Dummett is prepared to embrace the "supposedly absurd consequences" of semantic anti-realism (Dummett, 2007c, p. 348). If $\neg\mathcal{K}\varphi \rightarrow \neg\varphi$ holds, φ and $\neg\mathcal{K}\varphi$ are indeed incompatible, which is enough to grant the intuitionist's commitment to (16). Williamson himself concludes:

> that a little logic should short circuit an intensely difficult and obscure issue was perhaps too much to hope, or fear. (Williamson, 1982, p. 207)

3.4.2 The Standard Argument

Williamson's and Dummett's defence of Weak Verificationism falters on closer inspection. The problem is that intuitionists themselves seem *forced* to accept the existence of forever-unknown truths. Consider some decidable statement P such that all the evidence for or against it has been lost—say "The number of hairs now

[20]Besides (16), Philip Percival points out two more untoward intuitionistic consequences of weak verificationism: $\forall\varphi(\neg\mathcal{K}\varphi \leftrightarrow \neg\varphi)$ and $\forall\varphi\neg(\neg\mathcal{K}\varphi \wedge \neg\mathcal{K}\neg\varphi)$. See Percival (1990).

[21]Crispin Wright has made essentially the same point in conversation.

on Dummett's head is even", as uttered just before some of Dummett's hairs are burned.[22] Given that P is decidable—we could have counted Dummett's hairs—even intuitionists should be willing to assert that either it or its negation is true. But since both P and its negation are *ex hypothesi* forever-unknown, the disjunction $(P \wedge \neg\mathcal{K}P) \vee (\neg P \wedge \neg\mathcal{K}\neg P)$ also holds, from which $\exists\varphi(\varphi \wedge \neg\mathcal{K}\varphi)$ trivially follows.

Call this the Standard Argument. It essentially rests on two assumptions:

(i) that the evidence for settling whether P has been lost,

and

(ii) that P is decidable, i.e. that there is a method whose application would settle in a finite amount of time whether P or $\neg P$.

These assumptions respectively yield, in turn,

$$\text{(i*) } \neg\mathcal{K}P \wedge \neg\mathcal{K}\neg P$$

and

$$\text{(ii*) } P \vee \neg P.$$

Dummett sometimes questions the step from (i) to (i*): from the fact that all the evidence for P has been lost, he says, we cannot infer that nobody will *ever* know whether P or its negation is true. He writes:

> I indeed believe that it can never be wholly ruled out, of any statement that has not been shown to be false, that it may eventually be shown to be true. (Dummett, 2007c, p. 348)

He also acknowledges, though, that in the example above, "that it will be never known whether the number of hairs on [Dummett's] head at a certain time was even or odd would seem to be the safest of predictions" (*Ibid.*). And while, on the one hand, he still claims that "when the point is pressed as hard as possible, we cannot absolutely rule out that some means of deciding the question, now wildly unexpected, may come to light: say some physiological condition proves to be correlated with the parity of the number of hairs on the head, and it can be determined whether [Dummett] was in that condition at the time" (*Ibid.*); on the other, he deems similar scenarios as "bizarre" (p. 348) and "implausible" (p. 350). One wonders whether the case for anti-realism should rest on "bizarre speculations" (p. 348). Can anti-realists do better?

[22]The example is Wolfgang Künne's. See Künne (2007).

3.4.3 Knowability and bivalence

Similar difficulties have recently led Dummett to defend anti-realism "without invoking implausible scenarios" (p. 350). The general idea is that anti-realists may evade the paradox by showing, on independent grounds, that one of its premisses is not assertible. As Dummett puts it in another context:

> a genuine solution [to the paradox] ought to show [...] that one of the premisses is false, or at least not assertible. (Dummett, 2007d, p. 452)

Again, the target is the Standard Argument for $\exists\varphi(\varphi \wedge \neg\mathcal{K}\varphi)$. But instead of quibbling with premise (i), Dummett now questions the step from (ii) to (ii*), i.e. the inference from the existence of a decision procedure for P to $P \vee \neg P$. The problem with this inference, he claims, is that anti-realists can assert $P \vee \neg P$ only if the decision procedure for P can *always* be applied. And whereas the decision procedure for a mathematical statement is always applicable, that for *empirical* statements may cease to be so. I could decide now, say, whether there is a dog behind the wall, but I may not be able to do so in one year time. Similarly, we could have decided whether Dummett's hairs were even in number at t before they were burned at a later time t^*, but not after that time.

However, why should not the applicability at *some* time of P's decision procedure be sufficient for asserting $P \vee \neg P$? From an anti-realist perspective, Dummett says, the truth of empirical statements such as "The number of Dummett's hairs was even at t" and its negation amount, respectively, to the truth of the counterfactual conditionals "If we had counted Dummett's hairs at t, they would have proved to be even in number" and "If we had counted Dummett's hairs at t, they would have proved to be odd in number".[23] But that one of these two counterfactuals is true, he writes,

> cannot be inferred from the unquestionable truth that, if [Dummett's] hairs were counted, they would be found to be either even or odd in number. (Dummett, 2007c, p. 350)

This would be an instance of the problematic step from $\phi \mathrel{\Box\!\!\rightarrow} (\psi \vee \chi)$ to $(\phi \mathrel{\Box\!\!\rightarrow} \psi) \vee (\phi \mathrel{\Box\!\!\rightarrow} \chi)$. And, according to Dummett, this inference

> does follow in the mathematical case, but not in the empirical case [...], the reason [being] that the outcome of the mathematical procedure is

[23] See Dummett (2007c, p. 349).

> determined entirely internally, but that of the empirical procedure is not. (Dummett, 2007c, p. 349).[24]

Anti-realists have no right to assert the disjunction:

(Hairs) Either if we had counted Dummett's hairs at *t*, they would have proved to be even in number, or, if we had counted Dummett's hairs at *t*, they would have proved to be odd in number.

Neither disjunct is presently assertible. Although *there was* a time at which we could have applied a decision procedure and find out which one is true, this possibility has now elapsed: the two disjuncts are now no more decidable than, say, Goldbach's Conjecture.

In order to apply classical logic to empirical statements that *could have been known*, but whose knowledge is now beyond our ken, the *unrestricted* Principle of Bivalence is needed, or so Dummett argues:

> [the realist] relies on assuming bivalence in order to provide an example of a true statement that will never be known to be true—more exactly, of a pair of statements one of which is true. He has to. If he could instance a specific true statement, he would know that it was true. This illustrates how important the principle of bivalence is in the controversy between supporters and opponents of realism. (Dummett, 2007c, p. 350)

Now recall the derivation of the Paradox of Knowability: the central core of the argument shows that weak verificationism is inconsistent with the existence of forever-unknown truths. If the latter claim is intuitionistically unacceptable, though, intuitionists may face no inconsistency after all. The argument shows, they might argue, that one cannot consistently maintain that all truths are knowable and that some truths are forever unknown. However, if anti-realists are only committed to the first claim, and if, as Dummett suggests, they should refrain from asserting the second, the Paradox dissolves. Or so Dummett argues.

3.5 The Paradox of Idealisation

Let us grant, for the sake of argument, the upshot of Dummett's reasoning: bivalence fails for empirical statements that could have been known, but no longer

[24]See also Dummett (2007a, pp. 303–4)

can. (We shall return to this assumption in § 3.5.3 below.) One might still wonder, though, why one could not run a version of the Knowability Paradox starting from a pair of *mathematical* statements, say Q and $\neg Q$, one of which is true but forever-unknown. Then, $Q \vee \neg Q$ would hold, even by Dummett's standards, and the Standard Argument would go through. Such an obvious response faces an obvious problem, however: namely, it would be very hard to motivate (i), i.e. the claim that we have lost the evidence for some *mathematical* statement. After all, as Dummett himself writes, "in mathematics, if an effective procedure is available, it always remains available" (Dummett, 2001, p. 1). On the face of it, I wish to argue, a result by Salvatore Florio and the present author—the Paradox of Idealisation—suggests that the obvious response may still be a good one, if properly motivated.[25] The result also raises a *prima facie* difficulty for the so-called *hierarchical* treatments of the original Paradox of Knowability. I begin by briefly introducing the hierarchical strategy. I then present the Paradox of Idealisation, and I finally discuss some potential objections.[26]

3.5.1 Hierarchical treatments

A quite natural way to block the paradox had already been suggested by Alonzo Church in 1945:

> Of course the foregoing refutation [...] is strongly suggestive of the paradox of the liar and other epistemological paradoxes. It may therefore be that Fitch can meet this particular objection by incorporating into the system of his paper one of the standard devices for avoiding the epistemological paradoxes. (Church, 2009, p. 17)

Bernard Linsky and Alexander Paseau have recently developed this thought.[27] Though the Church-Fitch proof makes no use of self-referential sentences, they observe, it is nevertheless invalid on a logical account of knowledge reminiscent of Russell's theory of types. The intuitive idea is that each formula is assigned a *logical type*, which reflects the nesting of occurrences of $\mathcal{K}$ within that formula. Formally, one introduces infinitely many knowledge operators $\mathcal{K}_n$, one for each natural number n. The type of any formula φ is defined by the greatest index of the knowledge operators occurring in φ. A formula of the form $\mathcal{K}_n\varphi$ is well-formed just in case n is strictly greater than the type of φ. In this framework, only

[25]See Florio and Murzi (2009).
[26]The materials from §§ 3.5.1-3 are drawn from Florio and Murzi (2009).
[27]See Linsky (2009) and Paseau (2008).

$\Diamond(\mathcal{K}_{n+2}\varphi \wedge \neg\mathcal{K}_{n+1}\varphi)$ follows from WVER. But unless it is assumed that $\mathcal{K}_{n+1}\varphi$ entails $\mathcal{K}_n\varphi$ for every index n and formula φ, that is not a formal contradiction.

Does the hierarchical treatment represent a viable answer to the Church-Fitch Paradox? And can a simple appeal to intuitionistic logic salvage semantic anti-realism from its paradoxical consequences?

3.5.2 Strict Finitism and the Paradox of Idealisation

There is a dispute among anti-realists over whether or not knowability requires idealisation. Strict Finitists think that idealisation is not required. The word 'knowable', for them, is to be interpreted as 'possibly known by agents just like us':

> the meaning of all terms, including logical constants, appearing in mathematical statements must be given in relation to constructions which we are capable of effecting, and of our capacity to recognise such constructions as providing proofs of those statements. (Dummett, 1975, p. 301)

Strict Finitism, though, has highly revisionary consequences. On that view, any decidable proposition that cannot be known for mere 'medical' limitations, e.g. some arithmetical propositions involving very large numbers, turns out to be false. Yet, it would seem, this result is hardly acceptable. As Dummett puts it:

> The intuitionist sanctions the assertion, for any natural number, however large, that it is either prime or composite, since we have a method that will, at least in principle, decide the question. But suppose that we do not, and perhaps in practice cannot apply that method: is there nevertheless a fact of the matter concerning whether the number is prime or not? There is a strong impulse that there must be. (Dummett, 1994, pp. 296-7)

Is Strict Finitism coherent? Dummett has famously argued that it is not. Call a number *apodictic* "if it is possible for a proof (which we are capable of taking in, i.e. of recognizing as such) to contain as many as n steps" (Dummett, 1975, p. 306). Then, Dummett tells us in his 'Wang's Paradox', the Strict Finitist should accept both of the following claims:

(A) For any n, if n is apodictic, so is $n + 1$;

(B) There is a number m "sufficiently large that it is plainly not a member of the totality [of apodictic numbers]" (Dummett, 1975, p. 306).

But, Dummett claims, (A) and (B) are jointly inconsistent: they jointly entail that the totality of apodictic numbers is both infinite and finite.[28]

Following Dummett, most anti-realists reject Strict Finitism and concede that 'knowable' in WVER is to be read as 'knowable in principle', i.e. knowable by agents endowed with cognitive capacities like ours or that finitely exceed ours.[29] Here is Tennant:

> The truth does not have to be knowable by all and sundry, regardless of their competence to judge. [...] This would be to hostage too much of what is true to individual misfortune. At the very least, we have to abstract or idealize away from the limitations of actual individuals. [...] At the very least, then, we have to imagine that we can appeal to an ideal cognitive representative of our species. (Tennant, 1997, p. 144)

Call such anti-realists *moderate*. Crucially for our present purposes, moderate anti-realists are committed to (B), at least in Dummett's view. That is, consider some decidable mathematical proposition P whose proof has at least m steps (if m exists, then P exists too). Then, according to Dummett, moderate anti-realists can legitimately say that either P or its negation is true: although neither P nor its negation is feasibly knowable, at least (and at most) one of them is nevertheless knowable in an idealised sense. In spite of its initial plausibility, I shall now argue, this move runs the risk of becoming a Trojan horse.

The argument starts from the moderate anti-realist's concession that there are feasibly unknowable truths, i.e. truths that, because of their complexity or of the complexity of their proofs, can only be known by agents whose cognitive capacities finitely exceed ours. In symbols:

$$(17)\ \exists\varphi(\varphi \land \Box\forall x(\mathcal{K}_x\varphi \rightarrow Ix)).^{30}$$

[28]This argument has been heavily criticized—see e.g. Wright (1982) and Magidor (2007). Considerations of space prevent me from examining the issue more closely, however.

[29]See especially, Tennant (1997, Chapter 5).

[30]Some readers may object that (17) should rather read $\exists\varphi(\varphi \land \forall x(\mathcal{K}_x\varphi \rightarrow Ix))$, perhaps on the grounds that, in some very remote world, non-ideal agents may be able to know propositions that are actually feasibly unknowable. I ask those readers to be kind enough to set aside this objection until the very beginning of § 3.5.3, where I will introduce a suitably modified formulation of (17) which, I will suggest, circumvents this objection. Thanks to Crispin Wright for raising this potential concern.

Let Q be one such truth and let 'Ix' read 'x is an idealised agent', where an agent counts as idealised if and only if her cognitive capacities—perceptual discrimination, memory, working memory etc.—finitely exceed ours.[31] Now let us assume that there are no idealised agents:

$$(18)\ \neg\exists x Ix.$$

It can be proved that the conjunction

$$(19)\ Q \wedge \neg\exists x Ix$$

is unknowable:

> *Proof*: Assume that $Q \wedge \neg\exists x Ix$ is knowable. Then there is a world w where some agent knows $Q \wedge \neg\exists x Ix$. Call this agent a. By (17), every agent who knows Q in w is idealised. Therefore, a is idealised. However, since a knows $Q \wedge \neg\exists x Ix$, by distributivity and factivity, $\neg\exists x Ix$ is true at w. Hence, a cannot be an idealised agent. Contradiction. Therefore, $Q \wedge \neg\exists x Ix$ is unknowable. ■

Let us call this the Paradox of Idealisation.

The argument generalizes. Similar proofs can be constructed for every formula φ and $\mathcal{P}(x, \varphi)$ such that the following holds:

$$(20)\ \exists\varphi(\varphi \wedge \Box\forall x(\mathcal{K}_x\varphi \rightarrow \mathcal{P}(x, \varphi)) \wedge \neg\exists x\mathcal{P}(x, \varphi)).$$

Relevant instances of $\mathcal{P}(x, \varphi)$ may include traditional necessary conditions for knowledge, such as justification or belief. The Paradox of Knowability itself may be thought of as a trivial instance of (20), with $\mathcal{P}(x, \varphi) \equiv \mathcal{K}_x\varphi$:

$$(20')\ \exists\varphi(\varphi \wedge \Box\forall x(\mathcal{K}_x\varphi \rightarrow \mathcal{K}_x\varphi) \wedge \neg\exists x\mathcal{K}_x\varphi).$$

The argument poses a problem for anti-realists who appeal to intuitionistic logic to block the Church-Fitch Paradox. If it is not to be regarded as a *reductio* of WVER, anti-realists have no choice but to deny either (17) or (18). I argue below that neither option seems viable, regardless of whether intuitionistic logic is adopted. However, if (17) and (18) hold, the proof outright contradicts both Dummett's IVER and WVER, thereby threatening to collapse the anti-realist's rejection of Strict

[31] I shall consider an alternative definition of an idealised agent in Section 3.5.3 below.

Finitism into a rejection of anti-realism itself.[32] The new paradox equally threatens to undermine hierarchical approaches to the Paradox of Knowability.[33] Although the definition of 'Ix' involves reference to cognitive capacities, it does not involve reference to knowledge of any particular proposition. Hence, typing '$\mathcal{K}$' would be uneffective here.[34]

The foregoing considerations suggest two claims. First, on the further assumption that there are no ideal agents, Q is feasibly unknowable only if Q is forever-unknown (more on the existence of ideal agents below).[35] Hence, (17) straightforwardly implies the existence of forever-unknown truths. Second, since Q is *ex hypothesi* a decidable *mathematical* statement, the above proof is intuitionistically unexceptionable—even by Dummett's enforced intuitionistic standards. I now turn to some potential concerns about the soundness of the idealisation proof.

3.5.3 Objections and replies

Let us begin with (17), i.e. the claim that there are feasibly unknowable truths. In light of the Paradox of Idealisation, anti-realists might reconsider their moderation and argue that for any true proposition φ, it is possible that φ be known by a non-idealised agent:

$$(21)\ \forall\varphi(\varphi \rightarrow \Diamond\exists x(\mathcal{K}_x\varphi \wedge \neg Ix)).$$

Since (21) intuitionistically entails the falsity of (17), the Paradox of Idealisation would be blocked. This thought might be motivated in different ways. For instance, anti-realists might claim that, if there is a method to verify φ, then there is a possible world whose space-time structure is such that agents with cognitive capacities just like ours know that φ. Alternatively, they might claim that for any

[32]*Proof*: Assume that $Q \wedge \neg\exists xIx$. Then, $\Diamond\mathcal{K}(Q \wedge \neg\exists xIx)$ follows by weak verificationism. By the Paradox of Idealisation, however, $\neg\Diamond\mathcal{K}(Q \wedge \neg\exists xIx)$ holds too. We thus have a contradiction resting on (17), (18) and Weak Verificationism. A parallel reasoning shows that the Paradox of Idealisation and Dummett's IVER give us the intuitionistically inconsistent $\neg\mathcal{K}(Q \wedge \neg\exists xIx)$ and $\neg\neg\mathcal{K}(Q \wedge \neg\exists xIx)$. ■

[33]Thanks to Tim Williamson for pointing this out.

[34]It might be objected that anti-realists could still block the Paradox of Idealisation by typing the predicate 'Ix'. It is however unclear whether they would have any independent reason for doing so. As Paseau (2008) points out, the main motivation for typing $\mathcal{K}$ is to avoid other paradoxes, such as the Paradox of the Knower. Yet, no analogous motivation seems to be available in the case of 'Ix'. Moreover, it is worth reminding that merely typing 'Ix' will not do: anti-realists would also need to type any other predicate one could substitute in (20).

[35]*Proof*: Assume that some agent knows Q. Call this agent a. By (17), a is an ideal agent, which contradicts our assumption that there are no ideal agents. Hence, nobody knows Q. ■

true φ, there is a possible world at which φ itself, or a proof of it, are expressed in a language that renders them cognitively accessible.[36]

This objection does not work. Let S be a description of the space-time structure of the actual world or a description of which languages are actually used. Now consider the modified premise:

$$(17^*)\ \exists\varphi((\varphi \wedge S) \wedge \Box\forall x(\mathcal{K}_x(\varphi \wedge S) \to Ix)).$$

In perfect analogy with the Paradox of Idealisation, we can argue as follows:

> *Proof*: Assume that $(Q \wedge S) \wedge \neg\exists xIx$ is knowable. Then there is a world w where some agent a knows $(Q \wedge S) \wedge \neg\exists xIx$. This forces w to have the space-time structure described by S, or a to speak an actual language. It also follows that $\neg\exists xIx$ is true in w. Therefore, a is a non-idealised knower of Q in a world whose space-time structure is S or where no non-actual language is used. Contradiction, since we are assuming that, necessarily, $\forall x(\mathcal{K}_x(Q \wedge S) \to Ix))$. Thus, $(Q \wedge S) \wedge \neg\exists xIx$ is unknowable. ■

Anti-realists might reply by exploiting the characteristic weakness of intuitionistic logic. They may deny (17), on the one hand, and express their moderation by claiming that not every truth is feasibly knowable, on the other:

$$(22)\ \neg\forall\varphi(\varphi \to \Diamond\exists x(\mathcal{K}_x\varphi \wedge \neg Ix)).$$

Classically, (22) is inconsistent with the denial of (17), but not intuitionistically. The problem with this move, though, is that intuitionists seem to be in a position to *prove* the existence of feasibly unknowable truths. Let Q be some decidable yet undecided mathematical statement whose decision procedure is feasibly unperformable. Then, Q satisfies both of the following:

$$(23)\ \Box\forall x(\mathcal{K}_xQ \to Ix);$$
$$(24)\ \Box\forall x(\mathcal{K}_x\neg Q \to Ix).$$

Since Q is *ex hypothesi* decidable, even the intuitionist should be willing to assert that either Q or its negation is true. The existence of a feasibly unknowable truth can then be easily derived from $Q \vee \neg Q$, (23), and (24).

Intuitionists might object that one can never rule out that a sentence that is now feasibly unknowable will turn out to be feasibly knowable. However, on the

[36] I wish to thank Cesare Cozzo and Luca Incurvati for pressing this point.

same grounds, one would be prevented from asserting empirical generalisations, as Dummett himself observes:

> there may be some point in saying that, for any statement not known to be false, we can never absolutely rule out the possibility that some indirect evidence for its truth may turn up; but if we are ever to be credited with knowing the truth of a universal empirical statement other than one that follows from scientific laws, this possibility may be so remote that we are sometimes entitled to say—as we often do—that it will be never be known whether P. (Dummett, 2001, p. 1)

Moderate anti-realists might bite the bullet and, instead, deny (18), i.e. the claim that there are no idealised agents. But would this be advisable? There are two possibilities, depending on how anti-realists define the notion of an idealised agent. If an agent counts as idealised just in case her cognitive capacities finitely exceed those of *any actual epistemic agent*, then (18) is indeed an a priori truth. It would say that there are no (actual) epistemic agents whose cognitive capacities finitely exceed those of any (actual) epistemic agent, which is of course a truism. One might object that, on this reading, the claim that there is a decidable proposition satisfying (23) and (24) would be hardly acceptable. For how do we know that in the actual world there will never be agents so clever that they will be able to decide Q? However, the existence of a decidable proposition satisfying (23) and (24) is only problematic if one assumes that there is no bound to the cognitive capacities of *actual* epistemic agents. If, as I think plausible, there is a bound, then it would seem difficult to maintain that there is no decidable and yet feasibly unknowable proposition. On the other hand, anti-realists might take (18) to be an empirical claim, for example following Tennant in defining 'Ix' in terms of human cognitive capacities. The worry would then be that a principle such as WVER, thought to be necessary and a priori, would carry a commitment, $\neg\neg\exists x Ix$, that is open to empirical refutation.

Be that as it may, if anti-realists went as far as denying $\neg\exists x Ix$, this would not help them with another variant of the Paradox of Idealisation, that rests on the following weaker assumption:

$$(25)\ \exists\varphi(\Diamond(\varphi \wedge \neg\exists x Ix) \wedge \Box\forall x(\mathcal{K}_x\varphi \rightarrow Ix)).$$

Presumably, even for an anti-realist there is some feasibly unknowable proposition φ, such that φ and $\neg\exists x Ix$ are compossible. Provided that the relation of accessibility is transitive, we can now run a version of the Paradox of Idealisation via (25) and the necessitated formulation of WVER:

$$(\text{WVER}^*)\ \Box\forall\varphi(\varphi \to \Diamond\mathcal{K}\varphi).$$

Anti-realists could reply by rejecting WVER*, thereby sticking to WVER. This, however, would be a desperate move: it would leave them with a contingent version of their core metaphysical tenet. They might still maintain that WVER is a priori, though contingent. But this does not seem to square with the modal profile of WVER as supported by the standard anti-realist arguments: semantic anti-realists like Dummett would find it problematic to give up the thought that, as a matter of *conceptual necessity*, truth cannot outstrip our capacity to know. Then, provided that the logic of conceptual necessity obeys the minimal modal principles required for our proof, the problem would still remain. Anti-realists would thus seem to have only one option left: giving up transitivity. But this would be a surprising consequence of accepting WVER.

3.5.4 Church-Fitch and empirical negation

The Paradox of Idealisation threatens the viability of intuitionist and hierarchical defences of semantic anti-realism. Hierarchical approaches might block the original Paradox of Knowability, but fail to block the cognate Paradox of Idealisation. As for the appeal to intuitionistic logic, it does not help the anti-realist avoid the inconsistency among the three assumptions on which the Paradox of Idealisation depends. Denying

$$(18)\ \neg\exists x I x$$

does not seem an option, independently of whether classical logic is admitted. Rejecting

$$(17)\ \exists\varphi(\varphi \wedge \Box\forall x(\mathcal{K}_x\varphi \to Ix)),$$

on the other hand, is tantamount to abandoning moderate anti-realism.

To be sure, there are some options left. As we have seen, setting aside his 2001 piece *Victor's Error*, Dummett hesitates between at least two different ways of dealing with the Paradox. On the one hand, he is tempted to embrace the intuitionistically unexceptionable claims that (i) we cannot legitimately say that nobody will ever know P, even if, for all we know, all the evidence for P has been lost, and that (ii) we cannot legitimately deny that there are ideal agents, if we cannot disprove their existence. The problem with this, as we have seen, is that Dummett himself acknowledges that, on this horn of the dilemma, anti-realism requires that we take seriously "bizarre" and "implausible" scenarios—scenarios where the evidence for P somehow comes to light, even though we had very good reasons for thinking that it had all been lost.

On the other hand, Dummett wishes to argue that we can only assert that there are forever-unknown truths if we are willing to apply Bivalence to statements that could have been known, but, for all we know, no longer can. If we are willing to drop Bivalence for these statements, then one of the premises of the Paradox of Knowability is no longer assertible. But even conceding this assumption, it should be noted, one can still derive a contradiction from WVER and $\exists\varphi(\varphi \wedge \neg\mathcal{K}\varphi)$. Hence, Dummett is still committed to $\neg(A \wedge \neg\mathcal{K}A)$; not much of a vindication of the thought that we can be justified in asserting that all the evidence for a given statement has been lost. Moreover, as I have argued, it seems possible show that there are true *mathematical* statements that we will never know, because their proofs, or the statements themselves, cannot be 'taken in', given our cognitive limitations. Granted, even *this* claim can be resisted. Intuitionists may insist that, for all we know, there *are* ideal agents, and that, for this reason, feasibly unknowable statements may after all be known. This line of argument, however, appears to bring us back to the "bizarre" and "implausible" speculations that Dummett himself is sometimes willing to dismiss.

How to resolve this tension? Intuitionists may be able to solve the problem by introducing an *empirical negation* $\sim$ in their language, alongside the negation they already have. They would have to ensure that $\sim$ can be applied to contingent, empirical statements, and that no contradiction follows from $\mathcal{K}(P \wedge \sim\mathcal{K}P)$ and $\mathcal{K}(Q \wedge \sim\exists xIx)$ (where Q is, of course, some feasibly unknowable statement). If they could do so, they would be in a position to consistently assert that nobody will ever know P, even if P may be true, or that there are no ideal agents, without thereby being landed in contradiction. I argue in Appendix C that the prospects for coherently introducing an empirical negation in the intuitionistic language look grim. For one thing, an empirical negation would seem to require knowledge in non-actual situations of what is actually the case—a very problematic assumption, as Williamson first showed.[37] For another, it appears to force intuitionists to give up assertibility-conditional semantics, as intuitionists who are willing to adopt an empirical negation are themselves willing to acknowledge.[38] But we will not take

[37]See Williamson (1987) and Williamson (2000). See also Murzi (2008) and *infra*, Appendix C, § C.6.

[38]See *infra*, Appendix C, §§ C1-C5. It remains to be seen whether an empirical negation can be defined in the framework of Crispin Wright's *Truth and Objectivity* (see Wright, 1992), where truth—at least in some discourses—is identified with *superassertibility*. I hope to be able to explore

matters further here.

3.6 Conclusions

I have argued that intuitionistic treatments of the Church-Fitch problem are problematic, for a number of reasons. While it is certainly true that the intuitionistic consequences of WVER are somewhat less unintuitive that the classical ones (the intuitionist's point here is well-taken), this does not seem to be a very strong, let alone sufficient, reason for adopting intuitionistic logic. What is more, the Paradox of Idealisation makes even more acute the relatively well-known problem of how to apply intuitionistic logic *outside* of intuitionist mathematics. Pending a viable account of empirical negation, intuitionists face the dilemma of being confronted with the paradoxical consequences of the knowability paradoxes, on the one hand, and the adoption of a very strong negation, on the other—one that can only apply in mathematical contexts.

To be sure, this is not to say that there may be non-intuitionst revisionary treatments of the Church-Fitch proof that are comparatively more palatable than the intuitionist one. Indeed, Heinrich Wansing (2002), Beall (2003), and Priest (2009a) have, among others, recently motivated some broadly paraconsistent treatments of the Church-Fitch problem. I do not have space here to evaluate these approaches. However, it is worth asking ourselves *how* they are to be evaluated. The problem is analogous to the one that was raised in connection of the Basic Revisionary Argument: the derivation of an untoward consequence from a set of assumptions is evidence that there is something wrong with our assumptions together with the principles of reasoning we relied on in our derivation. But we are not told *which* assumptions, if any, are at fault, nor are we told *which* logical principles, if any, should be deemed as invalid. My suggestion is that metaphysical principles *alone* cannot help us finding the right logic. What is needed is a general conception of logic: one that can help us selecting among competing revisionist options. It is to this more general project, and to its connections with intuitionistic logic, that we now turn.

this issue in my future research.

Part II

Intuitionism and logical inferentialism

Chapter 4

Logical inferentialism

So far, we have examined two main arguments for the adoption of intuitionistic logic: the Basic Revisionary Argument, and the Paradox of Knowability. Both arguments are *metaphysical*, in the sense that their main assumption, the Knowability Principle, is a metaphysical one. We saw, however, that both arguments are problematic, albeit for different reasons. The Basic Revisionary Argument validates a parallel argument for a conclusion that is unwelcome to classicists and intuitionists alike. As for the Paradox of Knowability, I have suggested that, *pace* Dummett and Williamson, the key for solving the paradox, if there is one, may after all not lie in the adoption of intuitionistic logic.

Our central topic in the reminder of this thesis will be yet another family of arguments for the adoption of intuitionistic logic. Unlike the Basic Revisionary Argument and the Paradox of Knowability, these arguments do not rely on any explicit metaphysical claim. Rather, their driving assumption is a semantic one: the *inferentialist* idea that the meanings of the logical constants are fully determined by the rules for their correct use—an idea that many philosophers, realist and anti-realist alike, find compelling. The intended upshot of the argument is that this idea is in tension with classical logic, or, at the very least, with its standard formalisations.

In a nutshell, the thought is that rules can determine meanings only if rules satisfy some proof-theoretic requirements. As it turns out, standard formalizations of *intuitionistic logic* by and large satisfy these requirements, but standard formalizations of classical logic do not. If the inferentialist approach is a viable one, we are indeed confronted with a very strong argument against classical logic. The reminder of this thesis is divided into three main parts. Chapter 4 sets itself the threefold task of introducing, motivating, and defending from two major objections, the inferentialist approach to logic. Chapter 5 and Chapter 6 investigate in detail the arguments for the inferentialist's proof-theoretic requirements. I will argue that, although not all of these requirements can in general be justified, the ones that can are strong enough to effectively undermine the standard formalizations of classical logic. Chapter 7 explores some possible classicist ways out of this bind. I will argue that classical logic *can* be made consistent with an inferentialist approach to logic, although some extra—possibly controversial—assumptions are needed.

Our plan in the present chapter will be as follows. Section 4.1 introduces logical inferentialism, in very broad strokes. Section 4.2 presents three possible arguments for it. Sections 4.3-5 consider, and address, two objections to the inferentialist approach to logic: Timothy Williamson's contention that logical inferentialism delivers an inadequate account of understanding, and Arthur Prior's attempted *reductio* of the idea that rules can determine meanings. Section 4.6 offers some concluding remarks.

4.1 Logical inferentialism

It is sometimes held that the meaning of a logical constant is fully determined by the rules for its correct use. There is nothing more to the meaning of conjunction, it is suggested, than the fact it is governed by its operational rules—in a natural deduction system, its introduction and elimination rules:

$$\wedge\text{-}I\ \frac{A \quad B}{A \wedge B} \qquad \wedge\text{-}E\ \frac{A \wedge B}{A}\ \frac{A \wedge B}{B}\,.$$

Similarly for the other sentential connectives, and for the quantifiers: their meaning is fully determined by their introduction and elimination rules, or so the thought goes. And it is a tempting thought. A speaker who did not master $\wedge$-I could hardly be credited with an understanding of conjunction. Conversely, it would seem to be a mistake not to attribute an understanding of conjunction to a speaker who did master $\wedge$-I and $\wedge$-E. Call this view *logical inferentialism.*

Inferentialists typically individuate *two* central aspects of the correct use of a sentence: the conditions under which it may correctly asserted, and the consequences that may be correctly derived from (an assertion of) it. Here is a often quoted remark by Dummett:

> crudely expressed, there are always two aspects of the use of a given form of sentence: the conditions under which an utterance of that sentence is appropriate, which include, in the case of an assertoric

> sentence, what counts as an acceptable ground for asserting it; and the consequences of an utterance of it, which comprise both what the speaker commits himself to by the utterance and the appropriate response on the part of the hearer, including, in the case of assertion, what he is entitled to infer from it if he accepts it. (Dummett, 1973a, p. 396)

On their most common interpretation, introduction rules (henceforth, I-rules) state the sufficient, and perhaps necessary, conditions for asserting complex statements; elimination rules (henceforth, E-rules) tell us what we may legitimately infer from any such statement.

To the best of my knowledge, the inferentialist approach to logic was first formulated in some detail by Gerhard Gentzen, the founder of proof-theory. In a justly celebrated passage, Gentzen writes:

> To every logical symbol &, $\vee$, $\forall$, $\exists$, $\rightarrow$, $\neg$, belongs precisely one inference figure which 'introduces' the symbol—as the terminal symbol of a formula—and which 'eliminates' it. The fact that the inference figures &-E and $\vee$-I each have two forms constitutes a trivial, purely external deviation and is of no interest. 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 may 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'. (Gentzen, 1934, p. 80)

Gentzen argues that the I-rules of his newly invented calculus of natural deduction 'fix', or 'define', the meanings of the expressions they introduce. He also observes that, on this assumption, E-rules cannot be chosen randomly. They must be justified by the corresponding I-rules: they are, in some sense, their 'consequences'. This is a key thought. It expresses *in nuce* the idea that I- and E-rules must be, in Dummett's phrase, in *harmony* between each other. Conversely, if it is thought that E-rules are meaning-constitutive, I-rules cannot be chosen arbitrarily either. Dummett writes:

> The two complementary features [verifications and consequences] of any [...] linguistic practice ought to be in harmony with each other: and there is no automatic mechanism to ensure that they will

> be. The notion of harmony is difficult to make precise but intuitively compelling: it is obviously not possible for the two features of the use of any expression to be determined quite independently. Given what is conventionally accepted as serving to establish the truth of a given statement, the consequences of accepting it as true cannot be fixed arbitrarily; conversely, given what accepting a statement as true is taken to invlove, it cannot be arbitrarily determinated what is to count as establishing it as true. (Dummett, 1991b, p. 215)

I shall attempt to make the notion of harmony precise in Chapter 5. For the time being, let us elaborate, in some more detail, on Gentzen's suggestion that I-rules 'fix' the meanings of the logical operators.

4.1.1 The Determination Thesis

Logical inferentialists maintain that I- and E-rules *fully determine* the meanings of the expressions they respectively introduce and eliminate. Thus Popper and Kneale:

> the meaning of [the logical constants] can be *exhaustively* determined by the rules of inference in which these signs occur; this fact is established by defining our definitions of these formative signs explicitly in terms of rules of inference. (italics added Popper, 1947, p. 220)
>
> Formal (or logical) signs are those whose *full* sense can be given by laying down rules of development for the propositions expressed by their help. (Kneale, 1956, pp. 254-5; italics added)

Dummett also embraces the view:

> The meaning of [a] logical constant can be completely determined by laying down the fundamental laws governing it. (Dummett, 1991b, p. 247)

We shall call this the *Determination Thesis*:

> (DT) The meaning of a logical constant is fully determined by (possibly a non-empty subset of) its operational rules (i.e. rules containing occurrences of some logical operator in the their schematic form).

But what does it mean to say that rules determine meanings? And what do inferentialists mean by 'meaning'?

It is natural to take inferentialists as saying that operational rules determine the meanings of the logical vocabulary in the sense that they fully determine the correct use of logical expressions. It will be useful, however, to distinguish between a weak and a strong interpretation of this claim. On the *weak* interpretation, to which we will return in § 4.1.2 and § 5.1 below, *I-rules* determine meanings in virtue of determining a complete set of instructions for introducing complex statements. On the *strong* interpretation, operational rules determine meanings in the sense that they allow us to derive *all* the correct uses of the logical operators. As Peter Milne puts it:

> [all correct] use[s] of the constant in question [are], in some sense to be specified, derivable and/or justified on the basis of the putatively meaning-conferring rule or rules. (Milne, 1994, pp. 49-50)

In short: meaning-constitutive rules must be complete with respect to *all* the intuitively correct uses. This is a very strong completeness assumption—one that we will expound in Chapter 6, and that, as we shall see in due course, is obviously in tension with the *incompleteness* of higher-order logics (i.e. logics where we are not allowed to quantify over objects and individuals, but also over sets, properties, sets of sets, properties of properties, etc.).

4.1.2 The Fundamental Assumption

Let us now have a closer look at the weak interpretation of the thesis that basic inference rules fully determine the correct use of the logical operators. In the inferentialist's jargon, I-rules are interpreted as determining the *canonical* or *direct grounds* for asserting complex statements. As Dummett puts it:

> what the introduction rules for a constant \$ are required collectively to do is to display all the canonical ways in which a sentence with principal operator \$ can be inferred. (Dummett, 1991b, p. 257)

Thus, the I-rule for conjunction tells us that there is one canonical way of introducing $A \wedge B$: from A and B, infer $A \wedge B$. The (standard) I-rules for disjunction tell us that there are two canonical ways of introducing $A \vee B$: from A, infer $A \vee B$, and from B, infer $A \vee B$. And so on. I-rules specify *canonical* grounds in the sense that they are assumed to account for *all* the possible uses of the complex sentences they introduce. As Dummett puts it, I-rules are "collectively in a certain sense *complete*" (Dummett, 1991b, p. 252). Likewise, Stephen Read writes:

> what is implicit in the totality of cases of the introduction-rule for a connective is that they *exhaust the grounds* for assertion of that specific conclusion. (Read, 2008, p. 6; italics added)

To be sure, this claim should not be taken literally: it is strictly speaking false that we can *only* introduce complex statements by means of one of their I-rules. For instance, one may legitimately introduce $A \wedge B$ from C and $C \rightarrow (A \wedge B)$. The idea is rather that, on the inferentialist view, I-rules specify sufficient and *in principle* necessary conditions for assertion. Dummett makes the point:

> A statement may frequently be established by *indirect* means, but to label certain means 'canonical' is to claim that, whenever we are justified in a asserting the statement, we *could have* arrived at our entitlement to do so by those restricted means. [...] If a statement whose principal operator is one of the logical constants in question can be established at all, it can be established by an argument ending with one of the stipulated I-rules. (Dummett, 1991b, p. 252)

In short: whenever we can introduce a complex statement, we *could have* introduced it by means of an argument ending with an application of one the the introduction rules for its main logical operator.

Dummett and Prawitz call this the *Fundamental Assumption.*[1] It amounts to assuming that I-rules are, in Dummett's own words, "collectively in a certain sense *complete*" (Dummett, 1991b, p. 252): "in a certain sense", they cover *all* the uses of the logical operators they introduce. The assumption really deserves its name. For one thing, as we shall see in § 5.1, it directly justifies the inferentialist's requirement of proof-theoretic harmony—one of the key inferentialists requirements on admissible meaning-constitutive rules. For another, it underpins the standard inferentialist account of validity.[2]

Two observations are in order. First, notice that Dummett's qualification (*"in a certain sense* complete") is crucial. The assumption requires that I-rules be complete

[1]See Dummett (1991b, pp. 252-254) and Prawitz (2006, p. 522). The assumption only applies to *closed arguments*, i.e. on arguments that have no undischarged premises and no unbound variables. Thus, the fact that there is a (non-canonical) derivation of $\perp$ from A and $\neg A$ does not mean that there must be a canonical argument for $\perp$: because the argument from A and $\neg A$ to $\perp$ is not closed, the assumption does not apply to it in the first place. See also Schroeder-Heister (2007) for a defense of the claim that I-rule define meanings in virtue of collectively specifying necessary and sufficient conditions for assertion.

[2]As Prawitz first showed, inferentialists can define (first-order) validity in proof-theoretic terms, where, roughly, an argument is valid if and only if it can be converted into an argument which only consists of applications of I-rules. See e.g. Prawitz (1985) and Prawitz (2006). Prawitz's account of validity is briefly presented, and discussed, in Appendix E.

only in a rather *weak* sense, viz. that every assertible complex statement A must be provable by means of an argument whose *last step* is taken into accordance with one of the I-rules for its main logical operator. That is, the assumption does not say anything about what else can be used in order to introduce A canonically. For all it tells us, I-rules may be complete, but, as we shall see in Chapter 6, we may have to *enrich the language* in order for this to be the case.

Second, when inferentialists say that I-rules exhaust the grounds for asserting complex statements, they do not claim that I-rules cover, in principle, all the uses of *English words* such as 'and', 'or', 'every', and the like. They more modestly claim that I-rules are complete with respect to the correct uses of *logical operators*, which are in turn assumed to be complete with respect to *certain* key uses of 'and', 'or', 'every', and their ilk: what we may call their *logical uses*, e.g. their uses in mathematical proofs.[3]

It may be objected that the Fundamental Assumption is clearly incompatible with classical logic: after all, in standard formalizations of logic we cannot prove the Law of Excluded Middle by means of an argument ending with an application of disjunction introduction:

$$\frac{A}{A \vee B} \qquad \frac{B}{A \vee B}\ \vee\text{-I}$$

For this rule only allows us to infer $A \vee \neg A$ from either A or $\neg A$. And yet, as we have seen in Chapter 2, our epistemic situation is such that we are not in a position to assert, for every A, either it or its negation. Moreover, one might think, the assumption is at odds with a number of key uses of 'or'; e.g. cases in which we seem to be in a position to assert a disjunction without being in a position to know which of the disjuncts is true. These worries are legitimate, but, I will

[3]Thus, for instance, Dag Prawitz writes:

> One must distinguish [...] between two different questions: what is the meaning of this or that expression in an historically given language, and what meaning do I choose to confer on a certain expression in the language that I will use? Not that one can always keep these questions strictly apart—they may influence each other. But, as is well known, we would hardly find any logical principles if we just relied on the meaning of logical constants as they are used in a natural language. Even such a simple principle as the commutative law for logical conjunction does not hold in general for the English particle "and"; for instance, the principle fails when "and" has a temporal connotation, as it often has. So the validity or legitimacy of an inference usually depends on our conferring a particular meaning to logical constants, *which may agree only partially with some usage in a natural language.* (Prawitz, 2010, p. 9; italics added)

An interesting question, which Prawitz does not addresss, is whether the distinction between ordinary uses of a logical constant and its 'logical' uses can be made good without presupposing an understanding of the logical operators.

suggest, they ultimately depend on one's choice of the meaning-constitutive rules for disjunction: the Fundamental Assumption need not be the culprit. I will return to these potential issues in § 5.1, and I will present my preferred solution to the problem, in the form of an alternative set of rules for disjunction, in § 7.4.1.

4.1.3 From rules to truth-conditions

Inferentialists may hold that the foregoing reading of the Determination Thesis, in either its weak or in its strong interpretation, is all there is to the claim that operational rules determine the meanings of the logical operators. They may then *identify* the meaning of a logical expression with its inferential role. But although inferentialists like Robert Brandom[4] are willing to identify meanings with rules themselves, the identification of meaning with correct use is a very radical view. If sentence meanings in general at least individuate *truth-conditions*, meanings cannot be identified with inferential roles. As Dummett himself writes:

> the meanings of the logical constants cannot *consist* in their role in deductive inference: they must have meanings of a more general kind, whereby they contribute to the meanings of sentences containing them just as other words do. (Dummett, 1991b, p. 205)

Dummett's point seems correct. Lest inferentialists are willing to give an inferentialist account of *all* the expressions of the language, the meanings of the logical expressions cannot be conceived in purely inferential terms. Otherwise, it is difficult to see how the meaning of the logical vocabulary could contribute to the truth-conditions of the complex statements in which logical expressions may occur.

Logical inferentialist who are willing to reconcile an inferentialist approach to logic with a truth-conditional semantic framework may adopt the following broadly Fregean account of the meaning of the logical vocabulary—see e.g. Wagner (1981), Hodes (2004), and MacFarlane (2005, § 6). They may claim that, on the one hand, the meaning-constitutive rules for a logical operator $ determine its *sense*, i.e. what is sufficient for understanding $; and that, on the other, they also determine its *reference*; for instance, the truth-function it expresses, if $ is truth-functional.

What is the connection between the sense and the referent of a logical expression? One might think, following Frege, that sense must determine reference. In the case of the logical operators, however, one cannot simply *assume* that this is the case. Dummett makes the point:

[4]See e.g. Brandom (1994) and Brandom (2000).

> it may [...] be that the meanings of the logical constants are *determined* by the logical laws that govern their use in deductive arguments [...] this cannot be assumed—it needs to be *shown*. (Dummett, 1991b, p. 205)

Yet, nowhere in the *Logical Basis of Metaphysics* does he attempt to show how to derive a logical constant's meaning from its basic inference rules. Here is how one such story might go. Let L be language of classical propositional logic, where V is the set of admissible valuations v mapping the well-formed formulae of L to the set of Boolean values $\{1, 0\}$. Now consider conjunction. Our task is to derive its standard satisfaction clause, or its truth-conditions,

$$(\wedge)\quad v(A \wedge B) = 1 \text{ iff } v(A) = 1 \text{ and } v(B) = 1,$$

from its introduction and elimination rules. Since rules themselves do not say anything about truth, though, a semantic assumption is needed at this point. A natural candidate is the relatively uncontroversial claim that valid inference rules are truth-preserving, i.e. that they preserve truth on every valuation. Thus, Ian Hacking writes that "[only] given the underlying notions of *truth* and *logical consequence*, the [...] operational rules "fix the meanings of the logical connectives" in the sense of giving a semantics" (Hacking, 1979, p. 300; italics added). Likewise, Dummett stresses that "a theory of meaning [...] needs a notion of truth, as that which is guaranteed to be transmitted from premises to conclusion of a deductively valid argument" (Dummett, 2004, p. 32).[5]

With this assumption on board, the introduction and elimination rules for conjunction tell us that, for any valuation v, if A and B are true on v, so must be $A \wedge B$, and that if $A \wedge B$ is true on v, so must be A and B. Putting the pieces together, the rules for conjunction determine its truth-table, and its standard satisfaction clause. On the further assumption that competent speakers know, perhaps implicitly, that the rules for conjunction are valid, and that valid rules are truth-preserving, one might even say, as most inferentialists do, that our grasp of the rules for conjunction is constitutive of our understanding of 'and'. But what about the remaining logical operators? Matters become more complicated for disjunction and negation, as Rudolf Carnap (1943) first showed, and as we shall see in § 6.5 and § 7.4.4.

[5]See also Dummett (2005, p. 674). Notice, too, that truth-preservation is entailed by the proof-theoretic definition of validity as preservation of closed valid canonical arguments (see e.g. Prawitz, 2006), provided that there is a closed argument for A only if A is true.

4.1.4 The Stipulation Thesis

On the inferentialist view, the I- and E-rules of a logical operator \$ play a double role: they determine \$'s meaning, and they are constitutive of our understanding of \$. These semantic and epistemological assumptions have surprising semantic and epistemological consequences: that basic inference rules are *analytically valid*, and that we are *entitled* to infer according to the basic rules in virtue of our understanding of the logical vocabulary. We shall say a bit more about the inferentialist's contention that we are entitled to the validity of certain basic logical laws in § 4.2.1 below. For the time being, let us briefly focus on the inferentialist's own interpretation of the slogan that logical laws are *analytically valid*—valid in the virtue of the meaning of the logical vocabulary.

Dummett writes:

> Although it is not true of logical laws generally that we are entitled simply to stipulate that they shall be treated as valid, there must be certain laws or systems of laws of which this holds good. Such laws will be 'self-justifying': we are entitled simply to stipulate that they shall be regarded as holding, because by so doing we fix, wholly or partly, the meanings of the logical constants that they govern.[6] (Dummett, 1991b, p. 246)

Dummett's thought seems to this. In the overwhelming majority of cases, the question whether we may or may not accept a certain logical law is already settled: it depends on whether the given law can be justified with respect to the laws we already accept. However, certain basic laws *cannot* be justified in this way, on pain of an infinite regress. These basic laws, Dummett suggests, are *self-justifying*: they are constitutive of the meaning of the expressions whose logical behaviour they govern. For instance, it is constitutive of the meaning of conjunction that it obeys the rules of $\wedge$-I and $\wedge$-E. But, were it to obey *different* laws, $\wedge$ would cease to mean what it actually means. Dummett's thought boils down to what I shall call the *stipulation thesis*:

(ST) Meaning-determining inferential rules are valid by stipulation.

The resulting view is an *analytic* approach to logic. The basic rules are stipulated to be valid, in that they determine the meanings of the logical expressions they

[6]Dummett continues: "without thereby risking any conflict with the already given meanings of other expressions". We will investigate the question whether 'self-justifying' laws must be conservative in Chapter 6.

introduce or eliminate. The remaining rules can be justified with respect to them. Thus Tennant declares that

> logic is analytic: its rules are to be justified by appeal to the meanings of the logical operators. Indeed, certain of these rules are so basic as to be meaning-constituting; they afford a complete analysis of the meanings of the logical operators. They show that immediate moves in reasoning may be taken as irreducibly justified on grounds of [...] logical form and meaning alone. The remaining rules can then be justified by appeal to those meaning-constituting rules. (Tennant, 1997, p. 313)

Notice that the order of explanation is very important here. Meaning-constitutive rules are not made valid by some pre-existing meanings. For instance, the rule of $\wedge$-I is not valid because of the fact that $\wedge$ denotes a certain truth-function. It is rather the other way round: the logical constants have the meanings that they have in virtue of the use we make of them—a use that, inferentialists conjecture, by and large conforms to the meaning-constituting rules. Alberto Coffa famously makes the point—see also Carnap (1934, p. XV):

> The semantic explanatory route does not go from [...] "objects" or meanings to the laws concerning them and then to our reasonable linguistic behaviour, but the other way around, from our behaviour to meanings. The ultimate explanatory level in semantics is not given by reference to [...] objects or meanings, but by reference to the meaning-giving activity of human beings, an activity embodied in their endorsement of rules. (Coffa, 1991, p. 267).

This completes our brief introduction to the inferentialist approach to logic.

4.2 Three arguments for logical inferentialism

Is the view we have just sketched worth taking seriously? I will argue that there are good reasons for adopting a broadly inferentialist account of logic, and that some of the problems this account is alleged to face falter on closer inspection. Section 4.2.1 considers an *epistemological* argument, to the effect that inferentialism offers a plausible account of deductive knowledge—possibly the only one. Sections 4.2.2-3 introduce and develop two broadly Dummettian arguments for logical inferentialism—arguments that, I will suggest, are available to proof- and model-theorists alike.

4.2.1 Inferentialism and deductive knowledge

Logical inferentialists equate our understanding of the logical constants with our grasp of the basic rules for their correct use. They further contend that we are *entitled* to the validity of basic logical rules, and that this feature of their view enables them to respond to an epistemological puzzle, made famous by Lewis Carroll in his famous note *What the Tortoise said to Achilles* (see Carroll, 1895).

Suppose I'm in bed. It's 7 am, and I hear someone knocking at my door. Since I know that only the postman knocks the door at 7 am, I thereby come to know that the postman is knocking at my door. Or do I? Shouldn't I also know that the argument from

(1) Someone is knocking at my door, and it is 7 am;

(2) If someone knocks at the door at 7 am, then it is the postman;

to

(3) The postman is knocking at my door

is valid, in order to come to know (3)? Lewis Carrol's well-known regress suggests that this cannot in general be required. If, in order to infer (3) from (1) and (2), I also need to know

(4) The inference from (1) and (2) to (3) is valid,

it would seem that I would also need to know

(5) The inference from (1), (2), and (4) to (3) is valid.

And so on. What has gone wrong?

One part of the problem is that there is a difference between the *premises* of an argument and the *rules* that are used in that argument. To use a rule in an argument is *not* tantamount to implicitly using a premise in that argument. However, even granting this point, the difficulty still remains, since not every valid inference rule transmits knowledge. In order for a rule R to be *knowledge-transmitting*, speakers must be either aware of R's validity, or they must be somehow *entitled* to use R. For suppose B is a very remote consequence of $A_1, \ldots, A_n$. Then, inferring B from $A_1, \ldots, A_n$ will not give us knowledge of B, even if each $A_{i,\ 1<i<n}$ is known. How to characterize the class of knowledge-transmitting inferences, without initiating an infinite regress?

So-called *externalists* about knowledge purport to solve the problem by dropping the requirement that validity be known. They require instead that the subject be logically reliable, or logically capable, in the sense that

> [s]he is disposed to deduce a conclusion from some premisses only when the conclusion really does follow from them, and to recognize at least some of the more obvious cases of one statement's following from others. (Rumfitt, 2008, pp. 62-3)

There are well-known objections to externalism, however (see e.g. Boghossian, 2003, pp. 227-8). To mention but two: it is unclear how to characterize the class of the most obvious or simple inferences, and it is difficult to explain a subject's reliability. It might be thought that logically reliable subjects can discriminate good simple inferences from the bad ones, just as there are subjects who can reliably discriminate male from female chickens. But there appears to be a dissimilarity between the two cases. In the chicken case, there is a fact of the matter as to what subjects are sensitive to, viz. chicken sex. By contrast, one wonders what are logically capable subjects sensitive to, when they infer reliably (see Philie, 2007, pp. 191-2).

Logical inferentialists offer a different way out of the problem—one that is more in line with a broadly *internalist* account of knowledge. In their view, subjects are *entitled* to the validity of the operational rules for the logical constants, since, they argue, these rules fix the meanings of the expressions they introduce and eliminate, and grasp of these rules constitutes our understanding of these expressions. For instance, if we know A and $A \to B$, and we thereby infer B, we do not need to explicitly know that the inference from A and $A \to B$ is valid, in order to come to know B. Our knowledge of $\to$'s meaning, inferentialists claim, suffices to give us knowledge of B because, inferentialists argue, it is constitutive of our understanding of $\to$ that the grounds for asserting A and $A \to B$ are also grounds for asserting B.[7] As Paul Boghossian puts it: "it's constitutive of [our understanding of 'if'] that one take P and $P \to Q$ as a reason for believing Q" (Boghossian, 2003, p. 240).[8]

[7] An important observation: it does not follow from the inferentialist's assumption that some rules are meaning-constitutive that these rules are also knowledge-transmitting. The latter is a further assumption inferentialists must make. Thanks to Dominic Gregory and Crispin Wright for helpful discussion on this point

[8] The foregoing *semantic* route to entitlement is by no means the option available to the inferentialist. A prominent alternative can be found in Crispin Wright's notion of *entitlement of cognitive project*—see e.g. Wright (2004b) and, more closely connected to our present concerns, Wright (2004a). Wright's admittedly rough and tentative definition of entitlement is as follows:

> P is a presupposition of a particular cognitive project if
>
> (a) to doubt P (in advance) would rationally commit one to doubting the significance or competence of the project;
>
> (b) We have no sufficient reason to believe that P is untrue;

It may be argued that, pending an account of what it takes for a rule to be meaning-constitutive, the problem of characterizing the 'entitling' rules has now just been moved to the next level. But inferentialists *have* resources at their disposal to solve this problem, as I shall argue in Chapter 5 and Chapter 6. What remains to be seen is whether, irrespective of whether inferentialists can select meaning-constitutive rules in a principled way, understanding logical expressions can really be a matter of being willing to infer according to their basic inference rules, as inferentialists maintain. I will argue for a positive answer to this question in § 4.3.5 below.

4.2.2 Modesty and full-bloodedness

Let us now turn to a second argument for logical inferentialism. The argument relies on the admittedly controversial assumption that, in Dummett's terminology, a theory of meaning should be *full-blooded*: it should give an account of what it takes to understand the meanings of the expressions whose meanings it accounts for.

Consider the standard, Tarskian account of *logical consequence* and *logical truth*. A sentence A is a logical consequence of a set of sentences Γ if and only if every model of Γ is a model of A, where a model is an ordered pair $\langle D, I\rangle$ consisting in a domain of object D and an interpretation multi-function I assigning the expressions of the language appropriate extensions in D. In symbols:

$$(\mathsf{LC}_{\mathsf{Tarksi}})\ \Gamma \models A \Leftrightarrow_{df} \forall \mathcal{M}\ ((\forall B \in \Gamma)(\mathcal{M} \models B) \Rightarrow \mathcal{M} \models A).$$

This yields a definition of logical truth, as a limiting case where Γ is empty. A sentence is logically true if and only if it is true in every model:

$$(\mathsf{LT}_{\mathsf{Tarski}})\ \models A \Leftrightarrow_{df} \forall \mathcal{M}(\mathcal{M} \models A).$$

These definitions tell us that logical consequence is preservation of truth in *all* models, and that logical truth is truth in *all* models. But this is slightly mislead-

(c) The attempt to justify P would involve further presuppositions in turn of no more secure a prior standing ... and so on without limit; so that someone pursuing the relevant enquiry who accepted that there is nevertheless an onus to justify P would implicitly undertake a commitment to an infinite regress of justificatory projects, each concerned to vindicate the presuppositions of its predecessor. (Wright, 2004b, pp. 191-2)

Unfortunately, I do not have space here to investigate the question whether Wright's notion of entitlement may be better suited than the meaning-theoretic notion sketched in the main text.

ing. The quantification is implicitly restricted to the so-called *admissible* models: interpretations of the non-logical vocabulary that hold fix the interpretation of the logical constants. It is then legitimate to ask: what do the logical constants mean, for the model-theorist?

This question is easily answered: the meanings of the logical constants are to be identified with their contribution to the truth-conditions of the logically complex statements in which they may occur. Here is John McDowell:

> A [truth-conditional theory of meaning] would deal with the sentential logical connectives by saying things to this effect: '*A* and *B*' is true just in case '*A*' is true and '*B*' is true, '*A* or *B*' is true just in case '*A*' is true or '*B*' is true, 'If *A*, then *B*' is true just in case, if '*A*' is true, then '*B*' is true. (McDowell, 1997, p. 122)

The problem with this answer, though, is that, in itself, it is only informative if one already understands a meta-language rich enough to express the very concepts we are trying to elucidate. Jean-Yves Girard makes the point, not without sarcasm:

> In fact there is a complete absence of explanation. This is obvious if we look at the Tarskian "definition" of truth "*A* is true iff *A* holds". The question is not to know whether mathematics accepts such a definition but if there is any content in it [...]. What is disjunction? Disjunction is disjunction [...]. The distinction between $\vee$ and a hypothetical meta-$\vee$ is just a way to avoid the problem: you ask for real money but you are paid with meta-money. (Girard, 2003, p. 133)
>
> *"The rules of logic have been given to us by Tarski, who in turn got them from Mr. Metatarski"*, something like *"Physical particles act in this way because they must obey the laws of physics.* (Girard, 1999, p. 6)

The issue is not so much that the standard semantic clauses for the logical constants are incorrect: certainly $A \wedge B$ is true if and only if A is true and B is true. Rather, it is that truth-conditions, thus specified, do not say anything about what is to possess a certain concept: they can only be informative if one understands a metalanguage rich enough to express those concepts.[9]

This is a familiar point. On the one hand, there are, in Dummett's terminology, *modest* theories of meaning: theories which merely "show or state which concepts

[9] Girard also argues that they can easily lead us astray. He invites the reader to define a *broccoli logic* as follows. One introduces "new connectives, new rules, the worse you can imagine, and [then] define everything *à la Tarski*". Then, he argues, the standard meta-logical results will still be provable: "miracles of miracles, completeness and soundness still hold" (Girard, 2003, p. 133).

are expressed by which words" (Dummett, 1978c, p. 5). On the other, there is the thought that modesty is not enough: it does not give us, as Girard puts it, "real money". Any adequate theory of meaning should rather be *full-blooded*, i.e. it "must embody an explanation of all the concepts expressible in that language" (*Ibid.*). In Dummett's words:

> A more robust conception of what is to be expected from a meaning-theory is that it should, in all cases, make explicit in what a grasp of those concepts consists—the grasp that a speaker of the language must have of the concepts expressed by the words belonging to it. (Dummett, 1991b, p. 108)

The inferentialist approach to the meaning of the logical constants is full-blooded in Dummett's sense: it aims at explaining the meaning of the logical constants by offering an account of what a speaker must *do* in order to manifest a grasp of the relevant concepts, in agreement with Dummett's requirement that meaning be manifestable in use. For instance, the account tells us that, in order to understand $\wedge$ one needs to be willing to to infer according to its meaning-constitutive rules—in a natural deduction system, the rules of conjunction introduction and conjunction elimination.

It may be objected that, if it is good, this argument surely generalizes. After all, it is not only in the case of the logical constants that a modest theory merely shows or states which concepts are expressed by which words, assuming prior possession of those concepts. Quite the contrary: the point, if sound, would seem to apply to all the primitve predicates of the object-language. Does it follow, then, that the remedy is to give an inferentialist account of their meanings too?[10]

This objection is only partially correct. What follows from Dummett's requirement of full-bloodedness is that, for any primitive predicate F of the language, one must be able to state how an understanding of F can be manifested in our linguistic use. But this does not by itself implies that F's meaning be given by its inferential role: it only implies that knowledge of F's meaning must be manifestable in its linguistic use.

4.2.3 Admissible models

The foregoing considerations suggest a related argument for logical inferentialism. According to the standard Tarskian definition of logical consequence and logical

[10] Thanks to Bob Hale for raising this potential concern.

truth, the choice of admissible models requires a prior knowledge of the meaning of the logical constants. Thus, admissible models do not assign the value 1 to both A and its negation, they assign the value 1 to $A \wedge B$ if and only if A and B also have value 1, and so on. But how to justify the choice of the admissible models, on this view? Model-theorists might be able to motivate their choice of the truth-conditions for *some* logical constants,[11] but it is far from clear whether they can do so in a general way for all of them. By contrast, as we shall see in detail in Chapter 5 and 6, logical inferentialists can avail themselves of a number of sophisticated proof-theoretic constraints, in their quest for the correct meaning-constitutive rules for the logical constants—constraints that apply to *any* purported set of meaning-constitutive rules.

Model-theorists might insist that any way of fixing truth-conditions for logically complex sentences is admissible, so long as it does not result in inconsistencies. Which ways are useful or interesting, they might add, is another (perhaps pragmatic) matter. I am partly sympathetic with this line of thought: ultimately, the choice of the correct logic may in part be dictated by pragmatic reasons. The problem, however, is that it is unclear which notion of consistency the model-theorist can legitimately appeal to.

Suppose it is a *syntactic* notion, viz. either *Post-consistency*, that there is a A such that $\nvdash A$, or *Aristotle-consistency*, viz. that $\nvdash A \wedge \neg A$. The latter notion assumes a prior understanding of conjunction and negation, which appears to be question-begging in the present context. The model-theorist may not appeal to the meaning of negation and conjunction in order to define consistency, if this notion is to be used as a means of selecting admissible meanings for the logical constants. As for the first option, it presupposes a prior knowledge of what is derivable and what is not, i.e. of *which* rules are valid. But again, valid rules are the ones that hold in all admissible models. Pending a justification of the basic inference rules defining the relation of derivability, the model-theorist's choice of the admissible rules, and of the admissible models, seems once more arbitrary.

What if the model-theorist relied on a *model-theoretic* notion of consistency instead? Then, consistency may be model-theoretically defined the standard way:

> A is logically consistent if and only if there is a model of A.

But, of course, this is not much of an improvement either. For by 'model' we are here certainly intending *admissible* model, i.e. a model which respects the meaning of the logical vocabulary. Just as syntactic consistency, the notion of

[11] See e.g. Priest (2009b).

model-theoretic consistency relies too on a previous understanding of the logical vocabulary.

The model-theorist may wish to resort to a *primitive* notion of consistency. Thus, Hartry Field writes:

> When I say that we should regard the notion of consistency as primitive, I don't mean that there is nothing we can do to help clarify its meaning. The claim that consistency should be regarded as a primitive notion does involve the claim that we can't clarify its meaning by giving a definition of it in more basic terms. Similarly, logical notions like 'and', 'not', and 'there is' are primitive. We don't learn these notions by defining them in more basic terms. Rather, we learn them by learning to use them in accordance with certain rules; and we clarify their meaning by unearthing the rules that govern them. The same holds for consistency and implication, I claim: there are "procedural rules" governing the use of these terms, and it is these rules that give the terms the meaning they have, not some alleged definitions, whether in terms of models or of proofs or of substitution instances. (Field, 1991, p. 5)

The procedural rules Field alludes to are the *model-theoretic* principle:

(MTP) If there is a model in which, for every $A \in \Gamma$, A is true, then Γ is consistent;

and what he calls *modal-soundness*:

(MS) If Γ is consistent, then Γ is formally irrefutable.

He then adds that

> on this analysis *consistency is neither a proof-theoretic notion nor a model-theoretic notion*. The analysis puts proof theory and model theory on a par: *neither* are built into a definition of consistency; and [...] *both* are needed in order to formulate the intuitive principles that govern the notion. (Field, 1991, p. 6)

I find this puzzling, however. On the one hand, Field claims that certain notions are primitive, and that this does not prevent them to be elucidated by giving some rules for the use of the expressions that are meant to express them. On the other, he also says that "these rules give the terms the meaning they have", i.e. the rules are implicitly *defining*, in Field's view, the very notions he takes to be primitive. But even more importantly, Field's proposal is no less problematic than the model-theoretic definition of consistency. For without a notion of an admissible model at hand, the mere existence of a model is obviously not a sufficient condition for consistency.[12]

A primitivist about consistency should rather take logical notions such as consistency and logical consequence as a primitive in the sense that it is just a *brute fact* that, say, A and $\neg A$ are not consistent, and $A \vee A$ follows from A. This view, however, has the same limitations faced by the modest approach to meaning we have considered in the previous section. If consistency and validity are primitive notions, the logician's main guide for defining the class of the admissible models, and, with it, the extension of the relation of logical consequence, are her own *intuitions* about consistency. But this would be to give to intuitions—which are, after all, unjustified judgements—too prominent a role in debates concerning the choice of the correct logic. I am not denying that, in making these choices, we will have to ultimately rely on assumptions that we are not able to justify. I do think, though, that there is more philosophical work to be done before we can truly claim to have hit the bedrock.

If the foregoing considerations are correct, model-theorists have a *prima facie* difficulty in justifying their choice of the admissible models. By contrast, as we shall see in Chapter 5 and 6, inferentialists *have* means for selecting admissible rules, and hence admissible meanings, for the logical constants.

4.2.4 Inferentialism and model-theory

It should be noted that the difficulty only arises for model-theorists who refuse to avail herself of proof-theoretic tools in her attempt to characterize the class of the admissible models. Yet, it is unclear why model-theorists should refuse to do so. We have seen that, on a natural understanding of the Determination Thesis, inference rules determine meanings in the sense that one can derive the truth-conditions for the logical constants from the assumption that their meaning-constitutive rules are truth-preserving. For instance, one can easily derive the standard valuation-clause for conjunction, that $A \wedge B$ is true on a valuation if and only if both A and B are true on that valuation, from the assumption that the I- and E-rules for $\wedge$ preserve truth in all valuations. But then, if truth-conditions can

[12]Consider, for instance, a model $\mathcal{M}$ that makes A and $\neg A$ true—say that A is of the form Fa, and that $[a]_{\mathcal{M}}$ is both in the extension and in the anti-extension of F. Then, $\mathcal{M}$ exists, but it obviously does not guarantee consistency.

be so derived, an inferentialist account of the meaning of the logical constants may enable model-theorists themselves to better justify the choice of the admissible models, and hence the extension of the logical consequence relation. Thus Vann McGee:

> The rules of inference determine truth-conditions. The truth-conditions together [...] determine the logical consequence relation. (McGee, 2000, p. 72)

For instance, admissible models will be the ones that satisfy, among other things, the standard clause for conjunction. It is a mistake, therefore, to think that model-theoretic accounts of validity are necessarily incompatible with an inferentialist account of the meanings of the logical operators. It is open to argue that rules determine meanings, which in turn determine the extension of the relation of logical consequence, standardly defined as preservation in all admissible models.

Let us now turn to two standard objections to logical inferentialism: Timothy Williamson's recent contention that the inferentialist account of understanding falters on closer inspection (§ 4.3), and Arthur Prior's celebrated attack to the very idea that inference rules can determine meanings (§ 4.4).

4.3 Williamson's case against logical inferentialism

Logical inferentialists equate our understanding of logical expressions with our grasp of their meaning-constitutive rules. For instance, Paul Boghossian writes:

> inferring according to [a deductive pattern of inference] P is a precondition for having one of the concepts ingredient in it. (Boghossian, 2003, p. 239)
>
> it's constitutive of [our understanding of 'if'] that one take A and $A \rightarrow B$ as a reason for believing B. (Boghossian, 2003, p. 240)

But what does 'grasp' mean here, more exactly? Inferentialists may be tempted to say that to grasp $\$$'s rules just is a matter of being willing to infer according to them. Yet, there are reasons for thinking that they should resist this temptation.[13]

Consider Michael Dummett's example of the word 'Boche', used by French soldiers during the First World War as a derogatory way of referring to Germans (Dummett, 1973a, p. 454). Dummett's proposed introduction and elimination rules for "Boche" are, respectively, as follows:

[13] I develop a version of the argument to be given below in my Murzi (2010b).

(B-I) From 'x is German', one may infer 'x is Boche'

(B-E) From 'x is Boche', one may infer 'x is cruel' (or, as Dummett has it, 'x is more prone to cruelty than any other European').

Suppose one is willing to offer an inferentialist account of the meaning 'Boche'. Then, if these are the correct rules for 'Boche', it would follow that one understands 'Boche' only if one is willing to infer according to the above rules. But, Williamson (2003, pp. 257-9) argues, most speakers are *not* willing to infer according to these rules, even though they perfectly understand what "Boche" means. Worse still, they are not willing to infer according the the foregoing rules *precisely* because they know what 'Boche' means. This is, roughly, the shape of Williamson's arguments against inferentialist accounts of understanding: we are presented with cases of competent speakers who understand some expression *E*, but are nonetheless unwilling to infer according to (what are taken to be) its meaning-constitutive rules.

I take the 'Boche' objection to be the least controversial of Williamson's cases. It is clear that 'Boche' is an expression we do understand, and it is equally clear that we are not willing to infer according to what Dummett takes to be its introduction and elimination rules. Hence, our understanding of 'Boche' *cannot* require, let alone consist in, our willingness to infer according to B-I and B-E. On the other hand, the 'Boche' example is not directly an objection against *logical* inferentialism, given that 'Boche' is arguably not a logical expression. All the same, logical inferentialists had better be able to give a precise diagnosis of what, if anything, has gone wrong in the case of 'Boche'. For one thing, the objection threatens to undermine any account of understanding according to which our willingness to use an expression in a certain way is a *necessary* condition for understanding that expression—a relatively minimal assumption, to which most inferentialists are most likely to be committed. For another, although the objection does not directly threaten logical inferentialism, it nevertheless shows, if successful, that there are areas of discourse which *cannot* be accounted for in inferentialist terms. And why, one might ask, should we give an inferentialist account of logic, if we already know that there are areas of discourse in which the account fails? We shall return to the 'Boche' objection in § 4.3.3 below. For the time being, let us turn to Williamson's more direct objections against logical inferentialism.

4.3.1 McGee, Peter, and Stephen

Williamson has recently urged that even understanding *logical* expressions cannot be a matter of grasping inference rules.[14] Consider the inferentialist's claim that our understanding of 'if' consists in our willingness to infer according to its introduction and elimination rules. The problem is that there seem to be very competent speakers of English who appear to perfectly understand 'if', and yet are prepared to reject arrow introduction (conditional proof) or arrow elimination (*modus ponens*). It might be objected that these rules cannot be plausibly rejected: they are as basic as any logical rule can be. However, there are *prima facie* compelling grounds for rejecting them both.

Here is a very famous example by Vann McGee, aiming at showing that there are counterexamples to *modus ponens*:

> Opinion polls taken just before the 1980 election showed the Republican Ronald Reagan decisively ahead of the Democrat Jimmy Carter, with the other Republican in the race, John Anderson, a distant third. Those apprised of the poll results believed, with good reason:
>
> > If a Republican wins the election, then if it's not Reagan who wins it will be Anderson.
> > A Republican will win the election.
>
> Yet they did not have reason to believe
>
> > If it's not Reagan who wins, it will be Anderson.
>
> (McGee, 1985, p. 462)

As is well known, McGee himself takes this to be a counterexample to *modus ponens*. But then, one might ask, how could he understand 'if' *and* reject some instances of *modus ponens*, if, as inferentialists claim, his understanding of 'if' is at least partly constituted by his willingness to infer according to *modus ponens*? Williamson voices the concern:

> Vann McGee, a distinguished logician, has published purported counterexamples to *modus ponens*. Presumably, he refuses to make some inferences by *modus ponens*. Does McGee lack the concept *if*? [...] In conversation with McGee, he appears to understand the word 'if'

[14] See Williamson (2003), Williamson (2006), and Williamson (2008, Chapter 4).

> quite well by ordinary standards. He certainly seems to know what we other English speakers mean when we use the word 'if'. Before he had theoretical doubts about modus ponens, he understood the word 'if' if anyone has ever understood it; surely his theoretical doubts did not make him cease to remember what it means. We may therefore assume that McGee has the concept *if*, just like everyone else. (Williamson, 2003, pp. 251-2)

Notice that it is not open to argue that McGee may not be competent enough in the use of 'if'. As Williamson writes, "McGee is an expert on conditionals. He publishes on them in the best journals" (Williamson, 2003, p. 253). What is more, McGee is not the only philosopher of logic who has questioned the standard rules for 'if'. For instance, a long standing tradition, originated with Kripke (1975) and recently revamped by Hartry Field (2008), locates the source of the semantic paradoxes in the invalidity of classical rules such as conditional proof and negation introduction. Field does not believe that conditional proof is unrestrictedly valid: he rejects the standard introduction rule for 'if'.

The argument equally applies to logical truths. Consider the following sentence:

(6) Every vixen is a vixen.

This is an elementary logical truth. Hence, if our understanding of the logical constants is constituted, at least in part, by our willingness to infer according to their meaning constitutive rules, then any speaker who understands 'every' should be willing to assent, at least on reflection, to (6). Williamson introduces two characters, Peter and Stephen, who, he claims, perfectly understand 'every', and yet are not willing to assent to (6). In Peter's view, (6) presupposes

(7) There is at least one vixen.

Oddly enough, however, Peter thinks that presupposition is a logical entailment: (6) presupposes (7) if and only if (6) entails (7). Furthermore, Peter thinks that (7) is false: he is convinced that there are no foxes. Stephen, on the other hand, subscribes to a supervaluationist account of vagueness, according to which predications of borderline cases are gappy, or neither true nor false. He also believes that some fox ancestors were borderline cases for 'fox', and therefore 'vixen'. As a result, (6) comes out gappy on Stephen's semantics; but, since Stephen believes that one should only assent to true sentences, this means that Stephen, just like Peter, does not assent to (6).

Williamson submits that Peter and Stephen both perfectly understand the word 'vixen', just as McGee understands 'if' (Williamson, 2008, p. 88). Furthermore, he argues, Peter and Stephen's logical deviance is not really manifested in their use of 'every'. Hence, we really have no grounds for attributing them a deviant understanding of this word. Peter and Stephen "seem like most philosophers, thoroughly competent in their native language, a bit odd in some of their views" (Williamson, 2008, p. 88). Experts, Williamson writes "can make deviant applications of words as a result of theoretical errors and still count as fully understanding their words" (Williamson, 2008, p. 91).

Here the central assumption in play is the doctrine, sometimes referred to as *semantic externalism*, that to understand an expression just is, for the semantic externalist, to participate to a social practice: understanding does not require acceptance of any fixed set of linguistic uses. As Williamson puts it, following Quine, "*[n]o* given argument or statement is immune from rejection by a linguistically competent speaker" (Williamson, 2008, p. 97) and "[w]hat strikes us today as the best candidate for analytic or conceptual truth some innovative thinker may call unto question tomorrow for intelligible reasons" (Williamson, 2008, p. 126). But, Williamson thinks, this need not disrupt our linguistic understanding, since "[s]ufficiently fluent engagement in the practice can take many forms, which have no single core of agreement" (Williamson, 2008, p. 126). Williamson writes:

> Each individual uses words as words of a public language; their meanings are constitutively determined not individually but socially, through the spectrum of linguistic activity across the community as a whole. The social determination of meaning requires nothing like an exact match in use between different individuals; it requires only enough connection in use between them to form a social practice. Full participation in that practice constitutes full understanding. (Williamson, 2008, p. 91)

I will return to Williamson's semantic assumptions in § 4.3.5 below. For the time being, we shall look at a first possible inferentialist response to Williamson's argument—one that has been recently been advanced by Cesare Cozzo, and that Williamson himself considers, and dismisses. The reply concedes Williamson's point, but objects that it only undermines too crude an inferentialist account of understanding. We shall examine a less concessive response in §§ 4.3.5-6.

4.3.2 Knowledge of rules

Inferentialists like Boghossian are prepared to equate our understanding of a logical expression \$ with our willingness to infer according to \$'s meaning-constitutive rules. But, it might be suggested, this is a mistake: understanding really is a matter of *knowing* \$'s rules, and one might know that without thereby being willing to use \$, just as one may know the rules of a game without being willing to play it. Thus Cozzo:

> Does Williamson refute [logical inferentialism] in general? He does not, but he shows that the [inferentialist] should: i) emphasize the distinction between knowing a rule and accepting it; ii) explain understanding in terms of knowledge of rules and not in terms of acceptance. If *W* is a meaningful word, [...] a speaker *S* understands *W* if, and only if, *S* *knows* the constitutive rules, i.e. knows that *W* should be used in a certain way, e.g. in accordance with a pattern of inference *P*. Suppose that *S* understands *W* in this sense. It does not follow that *S* will use *W* or will accept uses of *W*. It follows only that *S* has the ability to use *W* according to *P*. (Cozzo, 2008, p. 315)

Understanding an expression *E*, Cozzo maintains, is not a matter of being willing to infer according to its meaning-constitutive rules. Rather, it is an *ability* to infer according to these rules—an ability grounded in the speakers' *knowledge* of the rules.

Williamson considers this possible move, and offers two counter-objections. First, he claims that the move would backfire, on the grounds that it is difficult to see how the kind of knowledge the inferentialist now appeals to can be "more practical than the semantic knowledge that the referentialist invokes" (Williamson, 2009a, p. 143). Second, he argues that not even knowledge how to infer according to B-I can be a precondition for understanding 'Boche'. He writes:

> Someone might grow up in a narrow-minded community with only pejorative words for some things, in particular with only the pejorative 'Boche' for Germans. He might understand 'Boche' as other xenophobes do without understanding 'German' or any equivalent non-pejorative term. He would be unacquainted with Boche-Introduction and any similar rule. Thus not even knowing how to infer according to Boche-Introduction is necessary for understanding 'Boche', or for having the concept that it expresses. (Williamson, 2009a, p. 143)

Let us consider these two arguments in turn.

Williamson's first argument relies on the assumption that knowledge of rules is not practical knowledge. But this is problematic. If to know E's rules is to *know how* to infer according to E's rules, then Williamson's objection assumes that a subject knows how to infer according to E's rules only if she *knows that* E's rules are valid. That is, the objection relies on the controversial assumption that knowledge-how is a special case of knowledge-that—an assumption that Williamson has influentially defended (see Stanley and Williamson, 2001), but which has also been been forcefully criticized.[15]

As for Williamson's second argument, it is difficult to see how it can generalize to the case of *logical* concepts. Moreover, I do not think that the objection works, even in the case of 'Boche'. On the global inferentialist semantics Williamson is attacking, 'German' will *itself* have a set of I- and E-rules, of the form

(G-I) From 'x is F', infer 'x is German';

(G-E) From 'x is German', infer 'x is F'.

But then, inferentialists may object that, even if 'German' is not present in the language, one *can* give an inferentialist account of 'Boche': one only needs to substitute 'Boche' for 'German' in G-I, and keep the original B-E, which does not involve 'German'.

Williamson might insist that one can understand what 'Boche' mean in languages where neither 'German' nor 'F' are present. But this will not solve the problem. The expressions in 'F' will themselves have I- and E-rules, call them the F-rules, and, in the envisaged scenario, it seems possible to formulate a new introduction rule for 'Boche' by means of the F-rules. Presumably, Williamson will object that this process can be iterated: one can further and further impoverish the language, so that no plausible rules for 'Boche' could be given. Yet, it is hard to see how Williamson can be correct in claiming that one could be competent in the use of 'Boche' without knowing anything like a suitable introduction rule. If 'Boche' were a term for an observable feature of things, one might claim that one's competence in the use of the term just consisted in one's ability to recognise cases in which it applied and tell them apart from cases in which it does not apply. But the plain fact is that 'Boche' is not an observation term: one would need to know articulable conditions for its application, and if there are such, there is scope for an introduction rule.

[15]See e.g. Rumfitt (2003) and Sgaravatti and Zardini (2008).

All the same, the foregoing response to Williamson requires that knowledge-how does not collapse on knowledge-that—an assumption that is not common ground between Williamson and his inferentialist opponent. Can inferentialists do better?

4.3.3 Inferentialism and 'Boche'

Let us consider the case of 'Boche' first. In response to Williamson, inferentialists need to argue that one can understand 'Boche' without thereby being willing to infer according to its meaning-constitutive rules. But, if one's understanding of 'Boche' is not constituted by one's willingness to infer according to its introduction and elimination rules, how can an inferentialist account of the meaning of 'Boche' be correct? It might be thought that inferentialists may take the meaning of 'Boche' to be determined by the following set of *indefeasible* rules (see Williamson, 2009a, p. 147):

(B-I*) From 'x is German', infer 'x is Boche';

(B-E*) From 'x is Boche', infer 'x is German'.

They may take these rules to fix the reference, and the meaning, of 'Boche', and maintain that Dummett's defeasible rules have a merely *pragmatic* significance. They do not affect the meaning of 'Boche', but they explain why 'Boche' is offensive. This account of the meaning of 'Boche', I take it, is vastly more plausible than the one suggested by Dummett. However, it should be noted that it does not yet solve the problem. Even granting B-I* and B-E*, 'x is Boche' still pragmatically implicates 'x is cruel'; an implicature that most speakers will be unwilling to convey.

Cozzo's suggestion was that one may know the meaning of a word, without thereby being willing to use it:

> We can have an ability without being willing to exercise it. Rejecting a recognized instance of a constitutive rule, therefore, does not necessarily show that one does not understand the relevant word. (Cozzo, 2008, p. 315)

This is a helpful observation, although, on its own, it does not yet address the question. To be sure, in view of examples such as "Boche", inferentialists need to concede that understanding is not a matter of being willing to infer according to rules. But *how* can this concession be reconciled with the inferentialist thought that understanding an expression requires that one grasp the rules for its correct use, without assuming that knowledge-how does not ultimately reduce to knowledge-that?

My suggestion is that Boghossian's account is indeed too crude. Subjects may not be willing to use words they do understand in a certain way for a variety of reasons. Certain words may be inappropriate because they are derogatory, as in the case of 'Boche', because they are gross, vulgar, etc. A more natural thought would be to say that a speaker understands an expression E only if, *were she under the obligation to use E, she would be use it according to the rules for its correct use*. More precisely:

> (INF) A speaker understands what E means in a language L on a given semantics S only if, were she under the obligation to use E, she would use it according to the S-rules for its correct use.

In order to undermine *this* claim, the inferentialist's opponent would have to find a situation where, although speakers are obliged to use some expression E, they do not use it according to the rules for their correct use. Such are indeed the cases of McGee and Williamson's characters, Peter and Stephen. But are these 'deviant logicians' counterexamples to *logical inferentialism*?

4.3.4 Theoretical and radical revisionism

Before we answer this question, let us first briefly pause on the very notion of *logical revision*. I take it that there are at least two ways one can be a logical revisionist, only one of which involves are revision of logic itself. On the one hand, two logicians may disagree as what are the logical rules we are actually following. I call this *theoretical revisionism*. On the other, the may disagree as to which logical rules we *should* be following, even if they agree on which rules we are actually following. I call this *radical revisionism*. For instance, relevant logicians typically do not advocate a *revision* of our use of 'if': quite the contrary, they claim that paraconsistent relevant logics better account for the way 'if' is actually used in English. They are, in our terminology, theoretical revisionists. By contrast, intuitionists like Michael Dummett and Dag Prawitz are willing to concede that the logical rules we actually follow are those classical logic, but ask for our revision of our logical practice. For instance, they claim that one may not unrestrictedly infer 'A' from 'It is not the case that not A', and they demur from asserting certain instances of the Law of Excluded Middle. They are, in our terminology, radical revisionists.[16] These two forms of revisionism are often conflated under the common slogan that some non-classical logic is the *correct logic*. There are important differences, however. Theoretical revisionists promote a revision of our belief that some logic, say classical logic, is the correct logic, but do not advocate a change of our actual logical practice. Radical revisionists concede that some logic, say classical logic, is the logic we are actually using, but call for a revision of logical practice itself.

Now back to Williamson's deviant cases. These cases all rely on the plausible assumption that it is possible to rationally disagree as to what the correct rules for using an expression E should be. But this assumption is shared by the inferentialist. Even conceding that we *know* what, say, 'if' means in English, inferentialists to allow for the possibility that there be speakers who are willing to give a different account of the meaning of 'if'. Likewise, inferentialists are happy to countenance the possibility that there be speakers who rationally suggest that we *change* the meaning of 'if'. To be sure, the point of Williamson's arguements is to show that inferentialists are not in a position to account for logical disagreement. However, I will suggest, this is mistaken. In what follows, I will argue that, irrespective of how one classifies McGee's, Peter's, and Stephen's revisionary inclinations, they do not constitute a counterexample to logical inferentialism.

4.3.5 The Quine-Williamson challenge

According to logical inferentialists, our understanding of 'if' is grounded in our willingness to infer according to its meaning-constitutive rules, whatever these may be. If *this* is Williamson's target, though, his cases at best show that some speakers have an idiosyncratic understanding of some logical expressions—their understanding is not grounded in a willingness to infer according to what are taken to be their standard introduction and elimination rules. In order for Williamson's argument to go through, at least two more assumptions are needed.

To begin with, the argument requires that McGee not only *thinks* that 'if' does not validate *modus ponens*, but that he actually *rejects* some instances of the rule. For suppose that McGee is a *theoretical* revisionist.[17] That is, suppose McGee just

[16]Graham Priest (2006a, p. 155) correctly stresses that what I call 'theoretical revisionism' just is a special case of belief-revision, and argues that it is "very misleading" to call this a revision of *logic*. Indeed it is: it is crucial to distinguish the revision of a logical *belief* from the revision of logic itself.

[17]This is likely to be Williamson's own interpretation of the case. Williamson writes: "[b]efore [McGee] had *theoretical doubts* about modus ponens, he understood the word 'if' if anyone has ever understood it; surely his *theoretical doubts* did not make him cease to remember what it means" (Williamson, 2003, p. 252; italics added).

thinks that our inferential uses of 'if' are best described as being applications of a restricted rule of *modus ponens*. Then, inferentialists may insist that McGee understands 'if' just like the rest of us, but disagrees with some of us about how that understanding is to be characterized. He may still hold, consistently with inferentialism, that it consists in acceptance of certain inference rules, but he thinks some of us have gone wrong about what those rules are. If McGee is right, we are, as a matter of fact, not willing to unrestrictedly infer according to *modus ponens*. Competent speakers, such as McGee and Williamson, are following some *restricted* rule, and share the same understanding of 'if'. However, this is consistent with the inferentialist account of understanding: inferentialists by no means require that the rules we are actually following be transparent to us. If McGee is wrong, on the other hand, we *are*, as a matter of fact, and *pace* McGee, willing to infer according to *modus ponens*. Competent speakers like McGee and Williamson are following the same unrestricted rule, but McGee is misdescribing the rule he is actually following. Yet, again, there is no objection so far to the inferentialist account understanding: inferentialists are not committed to the infallibility of their semantic beliefs.

Second, Williamson needs to assume that McGee understands 'if' precisely as the majority of the competent speakers of English do. Williamson explicitly makes the assumption:

> In conversation with McGee, he appears to understand the word 'if' quite well by ordinary standards. He certainly seems to know what we other English speakers mean when we use the word 'if'. [...] We may therefore assume that McGee has the concept *if*, just like everyone else. (Williamson, 2003, p. 252)

It is now clear what the problem is supposed to be: if McGee and Williamson share the *same* understanding of 'if', and if McGee is a *radical* revisionist, i.e. if he and Williamson use 'if' *different* ways as a result of following different rules for its correct use, it is hard to see how their understanding could be grounded in their use of 'if'.

One more observation before we proceed. The problem, if it is one, is more general than Williamson would make it seem. Suppose one thought, with Frege, that to understand a subsentential expression is to know its contribution to the truth-conditions of the complex sentences in which it may occur. Then, if 'if' means what classical logicians take it to mean, to understand 'if' is to know that "If A, then B" is true if and only if either A is false or B is true. Now suppose we wish to reject *modus ponens*. In order to do so, we must be able to exhibit some true conditional with a true antecedent and a false consequent. (McGee's examples precisely attempt to do as much.) However, if there are such examples, it is easy to check that the meaning of 'if' cannot be given by its classical truth-table. If 'or' and 'not' are to mean what they mean, a counterexample to *modus ponens* would require the truth of a disjunction, "Either not-A or B", both of whose disjuncts are *false*. If there are true conditionals with true antecedents and false consequents, our understanding of 'if' may not be constituted by a knowledge of its classical truth-condition. It follows that, if sound, Williamson's argument against the inferentialist account of understanding validates a parallel argument against the Fregean account of understanding as knowledge of truth-conditions.

The way out of the paradox, I suggest, is to reject Williamson's first assumption, viz. that he and McGee share the same understanding of 'if'. There are two cases to consider. First, suppose McGee is wrong: 'if' really satisfies the unrestricted rule of *modus ponens*. Then, if McGee insists in following a restricted rule of *modus ponens*, he would adopt, perhaps for the wrong reasons, a new understanding of 'if'. Now suppose McGee is right. Then, if Williamson insists that we should follow the unrestricted rule, *he* would adopt, perhaps for the wrong reasons, a new understanding of 'if'. But, the inferentialist will insist, neither scenario constitutes a counterexample to logical inferentialism, since, in either case, McGee's and Williamson's understanding of 'if' is still grounded in their willingness to infer according to *some* rule—respectively, the restricted and the unrestricted rule.

Williamson will presumably concede that this is what inferentialists *should* say, and object that this is just a *reductio* of the view. His first assumption, that his deviant logicians do not exhibit a deviant understanding, cannot really be disputed—the fact that it is inconsistent with the inferentialist account of understanding simply shows that the account is mistaken. McGee's rejection of *some* instances of *modus ponens*, at least from the perspective of a *semantic externalist*, does not count as evidence that McGee has a different understanding of 'if'. It is just a plain fact that McGee understands 'if' the way we do. McGee *is* a competent user of 'if',. To suppose otherwise just is to deny the data. Indeed, Williamson might add, for there to *be* a disagreement between him and McGee, him and McGee must talk *about the same thing*, viz *if*. If McGee means if_{VMG} and Williamson means if_{TW} by 'if', surely there cannot be a disagreement between them as to how these concepts are to be applied. Quine famously made the point:

> To turn to a popular extravaganza, what if someone were to reject the law of non-contradiction and so accept an occasional sentence and its

> negation as both true? [...] My view [...] is that neither party knows what is talking about. They think they are talking about negation, '$\neg$', 'not'; but surely the notion ceased to be recognisable as negation when they took to regarding some conjunction of the form '$p \wedge \neg p$' as true [...]. (Quine, 1970, p. 81)

Quine's conclusion, as is well-known, was to stick to a broadly dispositionalist account of understanding, on the one hand, and deem logical disagreement, and indeed logical revision, to be impossible, on the other:

> Here, evidently, is the deviant logican's predicament: when he tries to deny the doctrine he only changes the subject. (Quine, 1970, p. 81)

This, though, not only conflicts with Quine's statement that everything, including logical laws, can be revised in the face of recalcitrant experience:[18] it just seems wrong to say that different logicians cannot really disagree. Williamson's reaction to Quine's puzzle is to tolerate logical disagreement, and give up the inferentialist premises on which Quine's argument depends. Inferentialists, by contrast, seem faced with a harder task. They must insist that their account of understanding best accounts for the data, and show, at the same time, how logical disagreement can be, *pace* Quine, possible. I will consider these two issues in turn.

4.3.6 Inferentialism and understanding

Let me begin with an example. Suppose I were to systematically apply the word 'blue' to some (not many, perhaps) red things, on the grounds that this is how 'blue' ought to be used. Suppose, too, that my senses are perfectly working: my linguistic deviance is not due to the fact that, say, at certain times of the day I am subject to some temporary red-blue colour-blindeness. Then, my fellow speakers would presumably rightly surmise that I do not quite mean by 'blue' what *they* mean by that word. Similarly, I submit, if in some cases McGee does not think that Q follows from P and $P \rightarrow Q$, the natural assumption to make is that McGee has a deviant understanding of 'if': to claim that McGee understands 'if' just like the rest of us seems like insisting that, in the above example, I understand 'blue' just as my fellow speakers do. Semantic externalists are forced to either treat the 'blue' case as they would treat the McGee case, which seems implausible, or to give a different treatment of the two cases, which, again, seems hard to justify.

[18]The *locus classicus* here is, of course, Quine (1951).

Williamson will presumably insist, as he does, that McGee displays a perfectly good understanding of what 'if' means, and that if *he* does not understand 'if', this would also have to be true of the majority of English speakers (see e.g. Williamson, 2003, p. 253). The majority of English speakers make all kinds of mistakes and logical fallacies. Hence, if we take McGee's slight deviance in the use of 'if' as grounds for thinking that his understanding differs from ours (after all, the disputed uses all involve nested and relatively uncommon uses of 'if'), we are inevitably forced to conclude that virtually *every* speaker of English has an idiosyncratic understanding of 'if'.

This reply is unconvincing, though. To begin with, tutored speakers are typically unaware of the logical rules they follow. By contrast, the tutored McGee is very well aware of his choice of the rules for 'if', as testified by his publications on conditionals: one cannot equate the occasional deviant *performance* of a speaker with McGee's self-avowed idiosyncratic *competence*. Williamson might stress that McGee is both a competent speaker *and* a deviant user of 'if', and that this suffices for his case. But it does not. The inferentialist will respond that McGee looks competent to Williamson because he is competently inferring according to some revised rules for 'if', which perhaps include a restricted version of *modules ponens*; not because he and Williamson share the same understanding of 'if. To infer sameness of understanding from the fact that McGee appears to be using 'if' more or less as we do just seems to be a bad piece of reasoning.

Williamson (2008, p. 89) objects, in keeping with semantic externalism, that small differences do not make a difference:

> Peter's and Stephen's eccentricities [are not] sufficiently gross and extensive to constitute defeating circumstances [...] although their rejection of (6) might on first acquaintance give an observer a defeasible reason to deny that they understood ['every'], any such reason is defeated by closer observation of them. (Williamson, 2008, pp. 90-1)

However, it is difficult to see why a closer observation of a deviant speaker's non-deviant uses can help us alleviating the feeling that there is something wrong with her deviant uses. Consider again our blue-example: in the overwhelming majority of cases, I apply 'blue' to blue things, but sometimes I systematically apply it to red things—say only between 5 and 5:05 pm, and only if my interocutor's name begins with a 'S'. As a matter of fact, my idiosyncratic understanding of 'blue' will be very rarely manifested in my linguistic practice. Yet, this does not mean that we share the same understanding of 'blue'. If I were to teach my students how to use 'blue', I would make sure they understand what 'blue' *means*: specifically, I would make sure they very well understand that 'blue' applies to red things between 5 and 5:05 pm, if the name of our interlocutor begins with an 'S'. But notice that, if we follow Williamson in thinking that small differences do not make a difference, we would have to conclude that I have imparted to my students a perfectly ordinary understanding of 'blue'!

Williamson stresses that his deviant characters are all very competent English speakers, and that it would be very odd to correct their deviant uses, as we would do with "young children or native speakers of other languages who are in the process of learning English": "to stop our logical debate with Peter and Stephen in order to explain to themwhat the word 'every' means in English would be irrelevant and gratuitously patronizing" (Williamson, 2008, pp. 91). This much is certainly correct: we do not interrupt our conversations with speakers like McGee, Dummett, and Williamson's Peter and Stephen in order to explain to them what 'if', 'not', and 'every' really mean. That would indeed be irrelevant and patronizing. But it would also be inappropriate. There is no point in correcting these speakers, since we *know* that they have reasons for using these words in a deviant way: they have all published in refereed philosophy journals their views about 'if', 'not', and 'every', and we have all read their articles and books. Someone who had *not* read their work might sensibly stop them, and correct them. However, she would soon learn that McGee, Dummett, Peter, and Stephen's deviance is not due to a lack of linguistic competence: quite the contrary, it is motivated by theoretical considerations, possibly together with other beliefs. The fact that we respect—or at least tolerate—logical deviance is not evidence that we have the same understanding of 'if', 'not', and 'every': it is only evidence that we do not regard as irrational the thought that our beliefs about logic, or logic itself, can be revised.

Williamson further insists that his deviant characters have "acquired their non-standard views as adults". On the assumption that "before that, nothing in their use of English suggested semantic deviation", and that "the process by which they acquired their eccentricities did not involve forgetting their previous semantic understanding" (Williamson, 2008, p. 90), he concludes that Peter and Stephen's understanding cannot have changed. This indeed follows from Williamson's assumptions. But what are the grounds for assuming that Williamson's deviant characters did not change their understanding of 'if' and 'every' as a result of the adoption of their non-standard views? Williamson does not say. He claims that "the understanding which they lack is logical, not semantic" and suggests that "their attitudes [...] manifest only some deviant patterns of belief" (Williamson, 2008, p. 91). Yet, this is just to *state* that logical deviance has no semantic consequences, and that the deviant logical uses under considerations just are are the result of deviant beliefs: quite a question-begging assumption, in a context in which Williamson's opponent precisely takes one's understanding of a logical expression $ to be constituted by one's willingness to use $ in a certain way. I conclude that, *pace* Williamson, and semantic externalists with him, Williamson and McGee have a different understanding of 'if'.

Inferentialists, however, must still confront the task of explaining how, if this is true, logical disagreement is possible. There are two different aspects of the problem (see e.g. Dummett, 1978b, p. 119). First, inferentialists need to explain how rival logicians can *communicate* with each other, if they attach different meanings to some of our logical expressions. This is known as the problem of *shared content*. Second, they need to make sense of their *disagreement*: what, if anything, are different logicians disagreeing about?

In *The Philosophical Basis of Intuitionistic Logic*, Dummett writes:

> 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. (Dummett, 1973b, p. 119)

In the same article, Dummett goes on to put forward a solution to such problem, based on the fact that the intuitionist holds that there is a class of statements, both mathematical and non-mathematical, that obeys classical logic, namely the class of *decidable statements*. These statements, Dummett argues, can be used by the intuitionist to convey to the realist her conception of the meaning of non-decidable statements, whose semantics she takes to be intuitionist (see Dummett, 1978b, pp. 119-20).

Dummett's approach to the problem of shared content strikes me as being along the right lines. It can be straightforwardly applied to Williamson's cases. For instance, inferentialists can say that all McGee and Williamson can successfully communicate provided they confine themselves to the uses of 'if' on which there is agreement between them, e.g. the non-nested uses of 'if. Similarly for Peter, Stephen, and other deviant logicians. As for the problem of logical disagreement, if rival logicians can communicate, they can also communicate thoughts about how English words are to be used. Thus, they may say things to the effect that 'if' is to validate all instances of *modus ponens*, that 'not' is to satisfy Double Negation Elimination etc. If rival logicians can communicate, as I have suggested, logical disagreement can be accounted for as disagreement about which rules we take, or should take, logical expressions to be subject to.

Summing up, Williamson's arguments against logical inferentialism all involve subjects who are unwilling to infer according to (what are standardly taken to be) the basic rules for the use of certain logical expressions. There are at least two ways of being a deviant logician, however: two subjects may disagree as to how our actual logical practice is to be interpreted, but they may also disagree as to which logical rules we should, and could, rationally follow. Either way, I have argued, deviant logicians are no counterexample to logical inferentialism. On the one hand, theoretical disagreement is consistent with inferentialism: inferentialists are not committed to the infallibility of their semantic views. On the other, disagreement about logic itself can only be cause of concern on the assumption that deviant logicans understand logical expressions the way *we* do. This assumption ultimately rests on Williamson's intuition that the meaning of a word is not tied to its correct use, contrary to what inferentialists claim. But this is not to offer an *argument* against the inferentialist account of understanding: it is to presuppose the negation of the view Williamson is seeking to undermine. *Pace* Williamson, subjects who follow different logical rules have a different understanding of at least part of the logical vocabulary—irrespective of whether this understanding is to be accounted in inferentialist or broadly truth-condtional terms. *Pace* Quine, even on an inferentialist account of understanding, subject can disagree about the interpretation of the logical vocabulary *and*, at the same time, successfully communicate.

4.4 Prior's `tonk`

Let us now turn to one final objection. Like Williamson's alleged counterexamples, the objection only undermines rather naïve brands of inferentialism: it leaves non-naïve forms unscathed. Nevertheless, if correct, the objection cuts deeper than Williamson's arguments. It is very hard to spell out exactly what non-naïveté amounts to, as we shall see in Chapter 5 and Chapter 6.

4.4.1 Liberal inferentialism and tonk

Logical inferentialism became increasingly popular between the 30's and the 50's. Here are four representative quotes from, respectively, Ludwig Wittgenstein, Rudolf Carnap, Karl Popper and William Kneale:

> we can conceive the rules of inference [...] as giving the signs their meaning, because they are rules for the use of these signs. (Wittgenstein RFM, VII, § 30)
>
> Let any postulates and any rules of inference by chosen arbitrarily; then this choice, whatever it may be, will determine what meaning is to be assigned to the fundamental logical symbols. (Carnap, 1934, p. XV)
>
> The meaning of [the logical constants] can be exhaustively determined by the rules of inference in which these signs occur; this fact is established by defining our definitions of these formative signs explicitly in terms of rules of inference. (Popper, 1947, p. 220)
>
> Formal (or logical) signs are those whose full sense can be given by laying down rules of development for the propositions expressed by their help. (Kneale, 1956, pp. 254-5)

These first logical inferentialists subscribed to a very crude form of inferentialism, according to which *any* set of rules can be meaning-constitutive. This crude brand of inferentialism endorsed by Carnap, Popper, and Kneale received a jolt with the publication of Arthur Prior's *The runabout inference ticket* in 1960.

Prior (1960) famously showed that there is something deeply wrong with the early inferentialist's liberality. Consider a connective, tonk, with the following introduction and elimination rules:

$$\text{tonk-I}\,\frac{A}{A \texttt{ tonk } B} \qquad \text{tonk-E}\,\frac{A \texttt{ tonk } B}{B}\,.$$

If transitivity holds, and if we can prove at least one formula, it is easy to see that these rules allow us to prove *any* formula in the language, thereby yielding triviality and, provided the language includes negation, inconsistency.[19] Prior himself took his example to refute inferentialist accounts of the meanings of the logical constants in general.[20] In his 1960 paper, he introduces logical inferentialism as

[19] Of course, if we had *reductio* and double negation elimination in our proof-system, we would not even need to assume that at least one formula is provably in the system: given transitivity, tonk would then allow us to prove *any* formula.

[20] See e.g. Prior (1960, pp. 38-9) and Prior (1964, p. 194).

the thesis that some inferences are *analytically valid*, in the sense that they are valid in virtue of the meanings of the logical vocabulary occurring in them. He then writes:

> I want now to draw attention to a point not generally noticed, namely that in this sense of 'analytically valid' any statement whatever may be inferred, in an analytically valid way, from any other. Prior (1960, pp. 38-9)

Indeed. Prior's example shows that some choices of meaning-constitutive rules would be quite infelicitous. It is less clear, however, whether it follows from this that rules in general cannot determine meanings.

It it is widely thought that `tonk` is clearly a problem for logical inferentialism. Here is a recent quote by Graham Priest:

> One might say that the introduction and elimination rules for a connective in a system of natural deduction specify its meaning. The problem with this was pointed out by Prior (1960). (Priest, 2006a, p. 178)

But where does the problem exactly lie? It would seem that `tonk` undermines logical inferentialism only if *either* (i) it is assumed, as Carnap, Popper, and Kneale did, that any set of rules can be meaning-constitutive, and (ii) `tonk` lacks a meaning, *or* (iii) it is assumed that any set of rules can determine the meaning of a *logical* expression, but (iv) `tonk` is not logical.

In the first case, inferentialists may reject (ii) and insist that there are possible contexts in which `tonk` discriminates between correct and incorrect uses—contexts, for instance, in which logical consequence is not unrestrictedly transitive.[21] Alternatively, they might reject (i), on the grounds that only *logical* meanings are fully determined by the rules for their correct use. This brings us to the second case. Here the contentious assumption is clearly (iii). For why should inferentialists think that *any* set of rules define the meaning of a *logical* connective? As we have seen, Gentzen had already dismissed the view, on the grounds that admissible E-rules must be 'consequences' of the corresponding I-rules—they cannot be chosen randomly. To be sure, it must be conceded that Gentzen's remarks hardly solve the problem. For what does it mean to say that E-rules must be consequences of the corresponding I-rules? And, even if this can be clarified, *why* should this be the case? Should the converse direction also hold? Prior's `tonk` need not undermine the inferentialist approach to logic, but it nevertheless raises a crucial challenge:

[21] See Cook (2006).

that of justifying the choice of the meaning-constitutive rules, and hence the choice of logic.

It may be objected that there is no *need* to justify our logical choices in the first place. On a broadly Quinean approach to logic, `tonk` may be simply discarded because it is not *useful*. Thus, the Quinean may insist, we do not *need* a principled reason to rule out `tonk`, because we already have one.[22] My answer to this quick Quinean argument will also be quick. While I agree that one's choice of logic may be ultimately informed by pragmatic considerations, I do not believe that pragmatic considerations *alone* can provide a fully satisfactory justification of the choice of logic. On the Quinean view, there is no intrinsic difference between Prior's `tonk` and a well-behaved connective such as conjunction. The only difference between these two connectives is that, unlike conjunction, if we were to accept `tonk`, our inferential practices would be seriously compromised. It seems to me, though, that there *are* differences between `tonk` and conjunction—differences that are worth studying, and that, I will argue, shed light on the nature of logical concepts.

4.4.2 Towards a less liberal form of logical inferentialism

It is in this context that inferentialists like Dummett, Prawitz, and Tennant mount their challenge to classical logic. Their contention is that, if self-justifying meaning-constitutive rules must satisfy proof-theoretic requirements such as harmony, `tonk` *and classical negation* are, so to speak, in the same ballpark. Were we to find out that our current logical practice is not governed by proof-theoretically acceptable rules, we would have to conclude that some logical laws are not justified. Thus, Dummett writes that

> we are [not] obliged uncritically to accept the canons of [inference] as conventionally acknowledged. On the contrary, as soon as we re-construe the logical laws in terms of [an inferentialist] conception of meaning, we become aware that certain forms of reasoning which are conventionally accepted are devoid of justification. (Dummett, 1973b, p. 226)

But how can one proof-theoretically justify logical laws?

Proof-theoretic requirements fall into two main groups: *local* ones, concerning the form of acceptable rules, or pairs of rules, independently of the deductive

[22]Many thanks to Stewart Shapiro for pressing me on these Quinean points during the Academic Year 2009/10.

systems to which they may belong, and *global* ones, concerning the relations between rules and deductive systems.

The requirement of *harmony* belongs to the first group: it is a constraint on admissible pairs of rules, to the effect that there should be a kind of balance between admissible I- and E-rules. Clearly, the `tonk` rules are out of balance: its E-rule appears to be disproportionally strong—it tells us that *anything* can be inferred from `tonk`-statements. Thus, Dummett writes that Prior's error lies "in the failure to appreciate the interplay between the different aspects of 'use', and the requirement of *harmony* between them" (Dummett, 1973a, p. 397; emphasis added). It is worth stressing, though, that Dummett is giving a new name, harmony, to an old thought—a thought, that E-rules must be 'consequences' of the corresponding I-rules (and perhaps *vice versa*), that was nearly 40 years old at the time he was writing.

The requirements of *separability* and *conservativeness*, on the other hand, belong to the second group: roughly, they amount to requiring that admissible rules defining new vocabulary do not license new inferential relations among the expressions of the old vocabulary. Clearly, Prior's `tonk` does not respect this requirement either. It allows us to derive, on very minimal assumptions, $A \vdash B$.

Our task in the next two chapters will be to examine the inferentialist's justification for the requirements of harmony, separability, and conservativeness, to explore the relation between these, and other, proof-theoretic requirements, and to investigate their revisionary implications.

4.5 Conclusions

Logical inferentialism is, at least *prima facie*, an attractive approach to the meanings of the logical constants, and to logic more generally. It allows for a broadly internalist account of deductive knowledge—one that does not fall prey of Carroll's regress. And it promises us to provide means for selecting admissible rules, admissible logical meanings, and admissible models. What is more, if our arguments are successful, we have shown that some influential arguments against logical inferentialism, such as Williamson's McGee-like examples and Prior's `tonk`, falter on closer inspection. In responding to Prior's argument, however, we observed that inferentialists need to be able to discriminate between admissible and inadmissible meaning-constitutive rules. We have seen that Gentzen had already pointed out, back in 1934, that E-rules must respect the meanings defined by the I-rules—as we would say in a more contemporary jargon, I- and E-rules should be in *harmony* between each other. But what *is* harmony? This is where the problems begin. All the known accounts of harmony face difficulties, and it is not clear in the first place that they all aim at characterizing the same informal notion. Chapter 5 will be entirely devoted to the notion of harmony, to its justification, and to its alleged revisionary consequence. Chapter 6 will introduce and discuss more proof-theoretic requirements, and, with them, more arguments for logical revision.

Chapter 5

Proof-theoretic harmony

If our considerations in the previous section are correct, there are some *prima facie* compelling reasons for adopting a broadly inferentialist account of the meaning of the logical constants. Furthermore, I have argued, at least some would-be knock-down objections to logical inferentialism falter on closer inspection. It remains to be seen, however, whether inferentialists can satisfactorily respond to Prior's challenge. In this section, we shall consider three main accounts of harmony: harmony as reducibility, general elimination harmony, and what I shall call harmony as full invertibility. The connections between harmony and the global requirements of separability and conservativeness will be explored in Chapter 6. My main claim will be that the three accounts of harmony to be presented below are all equally viable, although the third one, I will suggest, is not particularly hospitable to intuitionistic logic. On the other hand, as we shall see, all three accounts sanction as non-harmonious the classical rules for negation.

The structure of the chapter is as follows. Section 5.1 introduces two arguments for harmony. Section 5.2 discusses Prawitz's account of harmony as reducibility, as well as its most natural strengthening, *strong* intrinsic harmony. Section 5.3 introduces the so-called *Generalized Inversion Principle*, and the elimination rules, General Elimination rules, it is usually taken to justify. Section 5.4 develops an account of harmony, harmony as *full invertibility*, as I shall call it, according to which harmonious E-rules can be generated by arbitrary I-rules, *and vice versa*. Section 5.5 briefly rehearses Dummett's and Prawitz's proof-theoretic reservations about classical negation. Section 5.6 offers some concluding remarks.

5.1 Two arguments for harmony

As far as I can see, there are two main arguments for harmony, one *epistemic*, the other *semantic*. Let us consider them in turn.

The epistemic argument

The epistemic argument proceeds from two main assumptions: (i) that logic alone should not create knowledge—one may not come to know, by logic alone, atomic statements that one could not otherwise have known, and (ii) that logic alone may not *destroy* knowledge either—introducing and immediately eliminating a logical operator should never yield a loss of knowledge. In a slogan: logic should be *epistemically conservative*.[1] These two thoughts lead to the claim that, on pain of compromising the epistemic neutrality of logic, E-rules should be neither stronger nor weaker than the corresponding I-rules.[2]

Admittedly, the requirement that basic logical rules be not too strong seems in better standing than the demand that logical rules be not too weak. After all, unlike E-rules that are too strong, weak E-rules will only produce a limited damage: they will never allow us to deduce falsities from truths. This objection, however, presupposes that only I-rules can fix meanings. But, one would want to ask, why should it be so? In principle, it would be preferable to have a more liberal approach to meaning-constitution—one on which I- and E-rules "are alternative in that either is sufficient to determine the meaning of a sentence uniquely" (Dummett, 1993a, p. 142). As Tennant puts it:

> any introduction rule, taken on its own, succeeds in conferring on its featured connective a precise logical sense. That sense in turn dictates what the corresponding elimination rule must be. *Mutatis mutandis,*

[1] See e.g. Wright (2003b), Wright (2007b), and Wright (2009) for an argument along similar lines.

[2] It may be objected that the classical rule of *classical reductio*

$$\dfrac{\begin{matrix}\Gamma, [\neg A]^n \\ \vdots \\ \bot\end{matrix}}{A}\ \text{CR}, n$$

is a clear counterexample to this requirement. After all, one might argue, does not this rule allow us to assert atomic statements which were previously not assertible, contrary to the requirement that logic should not create knowledge? This objection, however, does not work. For notice that, in order to derive a contradiction from $\neg A$, one already needs to be in a position to assert A. Hence, the rule does not seem able to allow us to assert *new* atomic statements which we were not previously in a position to assert.

> any elimination rule, taken on its own, succeeds in conferring on its featured connective a precise logical sense. That sense in turn dictates what the corresponding introduction rule must be. (Tennant, 2005a, p. 628)

But then, it would seem, if harmony requires that E-rules be not too strong, it should also require that, for any E-rule, its corresponding set of I-rules be not too strong either.

The semantic argument

Now to the semantic argument. Recall Gentzen's argument that E-rules cannot be chosen arbitrarily. His argument was in two steps. First, he assumed that I-rules 'define' the meanings of the expressions they introduce—at least in the case of *logical* expressions. He then claimed that E-rules must be faithful to these 'definitions'. As we have seen in § 4.1.2, inferentialists interpret Gentzen's first assumption as the claim that I-rules specify not only sufficient conditions for asserting complex statements, but also—at least in an idealized sense—*necessary* ones. This was essentially the content of the Fundamental Assumption. Thus, the rule of, say, conjunction introduction is meaning-constitutive because it fully determines the correct use of $\wedge$: *whenever we can introduce* $\wedge$, *we could have introduced it, in some sense of 'could', by means of a (closed) argument ending with an application of* $\wedge$*-I*. For, one might want to ask, how could we be in a position to assert $A \wedge B$ without being, in principle, in a position to assert both A and B?

Now suppose we are asking ourselves what the rule of $\wedge$-E should look like, on the assumption that $\wedge$ has been 'defined' by $\wedge$-I. What we know is that $A \wedge B$ can in principle *only* be derived from A and B. This means, however, that the inference *from* $A \wedge B$ *to* A and B will always be justified, i.e. $\wedge$-E can be justified with respect to $\wedge$-I. More generally, if I-rules exhaust in principle the possible grounds for asserting the complex statements they allow us to introduce, E-rules must give us back, so to speak, no more, and no less, than was required to introduce the complex statements they allow us to eliminate. This can be easily *proved* as follows. Let $\mathcal{CG}[A]$ be the canonical grounds for a complex statement A. Then, by the Fundamental Assumption, B follows from $\mathcal{CG}[A]$ if and only if B follows from A itself.

> *Proof*: Suppose B follows from A. Since A also follows from $\mathcal{CG}[A]$, B itself follows from $\mathcal{CG}[A]$. Now suppose B follows from $\mathcal{CG}[A]$.

Assume A. By FA, $\mathcal{CG}[A]$ itself follows. Hence, on our assumption that B follows from $\mathcal{CG}[A]$, we may conclude B, as required. ■

In short: it is a consequence of the Fundamental Assumption that complex statements and their grounds, as specified by their I-rules, must have the same set of consequences. I- and E-rules must be, in Dummett's phrase, in *harmony* between each other: one may infer from a complex statement nothing more, and nothing less, than that which follows from its I-rules.

The Fundamental Assumption conclusively refuted?

It may be objected that the semantic argument is at best unsound, because its main premise, the Fundamental Assumption, is either undermined by counterexamples, or it is, at best, question-begging.

Thus, Read (2000, p. 129) claims that the Fundamental Assumption is "conclusively refuted" by the I-rule for the possibility operator. In a nutshell, his argument is that the assumption collapses the distinction between A and $\Diamond A$.[3] If we apply the assumption to $\Diamond A$, we seem to be committed to saying that whenever $\Diamond A$ can be introduced, it could have been asserted *canonically*, i.e. by means of an argument ending with one step of $\Diamond$-I:

$$\Diamond\text{-I}\ \frac{A}{\Diamond A}$$

If correct, this reasoning would imply that whenever we can assert $\Diamond A$, we can also assert A, which is surely unacceptable.

But this conclusion is far too hasty. Read (2008) himself has recently rejected the main premises of his argument, viz. that $\Diamond$-I is the correct I-rule for $\Diamond A$. The idea is to index formulae to worlds, and to supplement the rules with constraints on the accessibility relations. For instance, the introduction rule for $\Diamond A$ is as follows:

$$\Diamond\text{-I}^*\ \frac{A_j \qquad i < j}{\Diamond A_i}$$

where 'i' and 'j' are labels indicating the worlds at which formulae are true, and '$i < j$' says that j is accessible from i. The rule intuitively says that, if A is true at j and j is accessible from i, then $\Diamond A$ is true at i. No modal collapse ensues from this rule. When we apply the Fundamental Assumption to it, we get the perfectly

[3]Dummett himself acknowledges the problem. See Dummett (1991b, p. 265). His own solution seems to be that either we cannot expect the Fundamental Assumption to apply to the whole of logic, or modal logic is not really logic. I find both horns of this dilemma rather problematic, however.

acceptable result that, if $\Diamond A$ is true at i, A must be true at some world accessible from i.

One might perhaps insist that the Fundamental Assumption is clearly at odds with a number of everyday uses of 'or'. Even intuitionist logicians must concede that we are often entitled to assert disjunctions even though we do not know which of the disjuncts is true. Call these *non-constructive* uses of 'or'. Dorothy Edgington offers the following example:

> A house has completely burnt down. The wiring was checked the day before, and two, independent grave electrical faults [call them X and W] were noted. Other possible explanations having been ruled out, we can (it seems) assert confidently "Either fault X caused the fire, or fault Y did". (Edgington, 1981, p. 181)

Moreover, the assumption may be accused of begging the question against the classical logician, since the standard proofs of *classical* Law of Excluded Middle, such as the following

$$\dfrac{\dfrac{\dfrac{{\scriptstyle(1)}\,\dfrac{\dfrac{\dfrac{[A]^1}{A \vee \neg A} \quad [\neg(A \vee \neg A)]^2}{\bot}}{\neg A} \qquad {\scriptstyle(1)}\,\dfrac{\dfrac{\dfrac{[\neg A]^1}{A \vee \neg A} \quad [\neg(A \vee \neg A)]^2}{\bot}}{\neg\neg A}}{\bot}}{{\scriptstyle(2)}\;\neg\neg(A \vee \neg A)}}{A \vee \neg A}$$

are bound to be counterexamples to it.

It seems to me that these objections are also too quick, however. They only show that *either* the Fundamental Assumption is incompatible with classical logic, and with some ordinary uses of 'or', *or* classical logicians should adopt different rules for disjunction. I shall argue in Chapter 7.4.1 that classical inferentialists have independent reasons for adopting *classical* rules for disjunction—rules that, as we shall see, satisfy the Fundamental Assumption, and validate the non-constructive uses of 'or'.

To be sure, these observations fall short of providing a full defense of the Fundamental Assumption. One major problem is that the Assumption sits very poorly with our ordinary use of universally quantified statements—a difficulty that Dummett himself recognizes in *The Logical Basis of Metaphysics*. Dummett writes:

> the universal quantifier, as ordinarily understood, appears not to fit [the Fundamental A]ssumption at all, which amounts to saying that

> we are entitled to say that something holds of *everything* only when we can show that it must hold of *anything*. It seems highly doubtful that we can hit on a genuine sense in which anyone entitled to assert a universally quantified statement could have arrived at it from the corresponding free-variable statement. (Dummett, 1991b, p. 274)

The difficulty, Dummett argues, is that, while $\forall$-I provides one type of ground for introducing universally quantified statements, "inductive procedures form the most obvious alternative type" (Dummett, 1991b, p. 275). And, Dummett suggests, $\forall$-I makes no provision for inductive reasoning.

But again, I do not find this objection irresistible. It is not "obvious" that inductive generalizations and proper applications of $\forall$-I are radically different. After all, in the inductive case, we can legitimately introduce $\forall x F(x) \to G(x)$ if we have verified that, for a finite sample of objects C, $F(a) \to G(a)$ holds, for any a in C. Similarly, we can introduce $\forall x F(x) \to G(x)$ by an application of $\forall$-I if we are able to prove, for an arbitrary a, $F(a) \to G(a)$.[4]

Be that as it may, I will assume, with the inferentialist, that harmony can be adequately motivated, in keeping with the general argumentative line of this thesis, which is to grant the logical reformist her metaphysical and semantic assumptions. Standard formalizations of classical logic, as we shall see, are typically not harmonious: their rules for eliminating negations are not justified by the corresponding I-rule, and *vice versa*. But, before we turn to this issue, we need to make the intuitive requirement of harmony more precise. This task will occupy us for the next six sections.

5.2 Intrinsic harmony

It will prove useful to start where everything started 75 years ago: from Gentzen's 1934 paper *Untersuchungen über das logischen schliessen*. Gentzen's remarks inspired at least two of the main contemporary accounts of harmony: *intrinsic harmony* and *general elimination harmony*. In this section, we shall focus on the first. We have seen in § 4.1 that, in Gentzen's view, I-rules define the expressions they introduce, and E-rules are just 'consequences' of these definitions. Gentzen adds that

> this fact may be expressed as follows: in eliminating a symbol we are dealing 'only in the sense afforded it by the introduction of that symbol'. [...] By making these ideas more precise it should be possible

[4] Dummett (1991b, pp. 274-7) himself develops an argument along these lines.

> to display the E-inferences as unique functions of their corresponding I-inferences, on the basis of certain requirements. (Gentzen, 1934, p. 80)

Gentzen points out that E-rules must be faithful to the meanings of the expressions they eliminate, as defined by their corresponding I-rules. Moreover, he argues, E-rules must be functions of the corresponding I-rules. But *how* to make these ideas more precise? In this section, we shall consider Dummett's and Prawitz's proposed elucidation of Gentzen's inspiring, if cryptic, remarks.

5.2.1 Making Gentzen's ideas more precise

We said that intuitively harmonious E-rules should be neither too strong, nor too weak: they should allow us to infer from a complex statement A nothing more, and nothing less, than what is required to introduce A in the first place. One half of the requirement of harmony, therefore, amounts to the following: if B follows from A, then it should already follow from $\mathcal{CG}[A]$, the canonical grounds for A. But this means that, if the rules for a logical operator \$ are harmonious, derivations containing sentences that are at the same time the conclusion of a rule of \$-introduction and the major premise of a rule of \$-elimination should always be transformable into derivation s from the same or fewer assumptions that do not contain any such detour.[5] Dummett calls a sentence which is at the same time the conclusion of an I-rule and the major premise of one of the corresponding E-rules a *local peak*. A necessary condition for harmony, then, is that 'local peaks' can always be removed, or, in Dummett's terminology, 'levelled':

> [F]or an arbitrary logical constant **c**, [...] it should not be possible, by first applying one of the introduction rules for **c** and then immediately drawing a consequence from the conclusion of that introduction rule by means of an elimination rule of which it is the major premiss, to derive from the premisses of the introduction rule a consequence that we could not otherwise have drawn. Let us call any part of a deductive inference where, for some logical constant **c**, a **c**-introduction rule is followed immediately by a **c**-elimination rule a 'local peak for **c**'. Then it is a requirement, for harmony to obtain between the introduction rules and elimination rules for **c**, that any local peak for **c** be capable of being levelled, that is, that there be a deductive path from the premisses of

[5] We shall look at some examples in § 5.2.3.

> the introduction rule to the conclusion of the elimination rule without invoking the rules governing the constant **c**. (Dummett, 1991b, pp. 247-9)

Following Dummett, let us call the foregoing requirement *intrinsic harmony*. In the *Logical Basis of Metaphysics*, Dummett "provisionally identif[ies] harmony between the introduction and the elimination rules for a given logical constant with the possibility of carrying out [...] the levelling of local peaks" (Dummett, 1991b, p. 250). It should be clear at the outset, though, that intrinsic harmony can only be *one half* of a viable definition of harmony—at least insofar as harmony must not only ensure that E-rules be not too strong, but also that they be not too weak. But, before we turn to the missing half of Dummett's notion of intrinsic harmony, let us have a closer look at intrinsic harmony, and its source: Dag Prawitz's 1965 doctoral dissertation.

5.2.2 Prawitz's Inversion Principle

Intrinsic harmony is based on Prawitz's *Inversion Principle*.[6] Prawitz informally states the principle as follows:

> an elimination rule is, in a sense, the inverse of the corresponding introduction rule: by an elimination rule one essentially only restores what had already been established by the major premiss of the application of an introduction rule. (Prawitz, 1965, p. 33)

Prawitz's wording suggests that E-rules must restore the conditions for introducing their major premises, as expressed by the corresponding I-rules. But this is quite misleading. Prawitz's principle only requires that the consequences of a complex statement A may not exceed the consequences of its canonical grounds: it does not require that A's canonical grounds themselves follow from A. Here is Prawitz's official statement of the principle:

> *let α be an application of an elimination rule that has B as consequence. Then, deductions that satisfy the sufficient condition [...] for deriving the major premiss of α, when combined with deductions of the minor premisses of α (if any), already "contain" a deduction of B; the deduction of B is thus obtainable*

[6]The idea of an inversion principle is borrowed from Paul Lorenzen (1955). There are, however, important differences between Prawitz's Inversion Principle and Lorenzen's *Inversionprinzip*. See Moriconi and Tesconi (2008) for an excellent discussion of Lorenzen's and Prawitz's inversion principles.

directly form the given deductions without the addition of α. (Prawitz, 1965, p. 33)

The idea is simple enough: E-rules satisfying the Inversion Principle do not allow us to infer anything that was not already inferable from the grounds for introducing A specified by its I-rules, in keeping with the intuitive idea that logic alone should not be creative.

5.2.3 Reduction steps

Some examples may prove useful. Consider the standard introduction and elimination rules for $\to$:

$$
{\scriptstyle \to\text{-I},\, i}\ \frac{\begin{array}{c}\Gamma, [A]^i \\ \vdots \\ B\end{array}}{A \to B}
\qquad\qquad
{\scriptstyle \to\text{-E}}\ \frac{\begin{array}{c}\Gamma_0 \\ \vdots \\ A \to B\end{array} \quad \begin{array}{c}\Gamma_1 \\ \vdots \\ A\end{array}}{B}
$$

A local peak created by successive applications of, respectively, $\to$-I and $\to$-E can be removed—or, in Dummett's helpful terminology, levelled:

Example 1. *$\to$-reduction:*

$$
{\scriptstyle \to\text{-E}}\ \frac{{\scriptstyle \to\text{-I},\, i}\ \dfrac{\begin{array}{c}\Gamma_0, [A]^i \\ \Pi_0 \\ B\end{array}}{A \to B} \quad \begin{array}{c}\Gamma_1 \\ \Pi_1 \\ A\end{array}}{B}
\quad \rightsquigarrow_r \quad
\begin{array}{c}\Gamma_1 \\ \Pi_1 \\ \underbrace{\Gamma_0, A} \\ \Pi_0 \\ B\end{array}
$$

where $\rightsquigarrow_r$ reads 'reduces to'. Our proof of B via $\to$-I and $\to$-E can be converted into a proof from the same of fewer assumptions that avoids the unnecessary detour through the introduction and elimination of $A \to B$. Any formula, such as $A \to B$ in our example, that is at the same time the conclusion of an introduction rule and the major premise of an elimination rule for the same constant, is called a *maximum formula* (or, as we have seen, a local peak).

Definition 1. *(Maximum formula) A formula occurrence occurring in a derivation* Π *that is both the consequence of an application of a $\$$ I-rule and the major premise of an application of a $\$$ E-rule is a* maximum formula *in* Π.

Here is another example. Consider the standard I- and E-rules for disjunction:

$$\vee\text{-I}\,\dfrac{A_j}{A_0 \vee A_1} \qquad\qquad \vee\text{-E}, i\,\dfrac{A \vee B \qquad \begin{array}{c}\Gamma_1,[A]^i\\ \vdots \\ C\end{array} \qquad \begin{array}{c}\Gamma_2,[B]^i\\ \vdots\\ C\end{array}}{C}\,.$$

where $j = 0$ or 1. Similarly, we can reduce a local peak created by successive applications of $\vee$-I and $\vee$-E to a simpler derivation that avoids the unnecessary detour through the disjunction rules:

Example 2. $\vee$-*reduction:*

$$\vee\text{-}E, i\,\dfrac{\vee\text{-}I\,\dfrac{\begin{array}{c}\Gamma_0\\ \Pi_0\\ A\end{array}}{A\vee B} \qquad \begin{array}{c}\Gamma_1,[A]^i\\ \Pi_1\\ C\end{array}\qquad \begin{array}{c}\Gamma_2,[B]^i\\ \Pi_2\\ C\end{array}}{C} \qquad \rightsquigarrow_r \qquad \begin{array}{c}\Gamma_0\\ \Pi_0\\ \underbrace{\Gamma_1, A}\\ \Pi_1\\ C\end{array}$$

The foregoing reductions are standardly called *reduction steps*, or *detour conversions*. They are available for all the operational rules of minimal logic, and they collectively constitute the induction step of Prawitz's *normalisation theorem* for classical and intuitionistic logic.[7]

At first approximation, a normalization theorem for a deductive system S tells us that every proof in S can be transformed into a direct, or non-roundabout proof, of the same conclusion from the same or fewer assumptions.[8] Normalization

[7]For intuitionistic logic, the theorem had already been proved by Gentzen; see von Plato (2008).

[8]See Prawitz (1965). One should be careful not to conflate the notion of a normal proof with the notion of a proof that does not contain local peaks, or maximum formulae. As Florian Steinberger (2009a, pp. 79-81) points out, there is more to normalization than the process of removing local peaks, and there is more to the notion of a normal proof than that of a proof with no maximum formulae. For consider the following derivation of C from $(A \wedge B) \vee (B \wedge E)$ and $C \wedge D$:

$$\wedge\text{-E}\,\dfrac{\vee\text{-E}, 1\,\dfrac{\wedge\text{-I}\,\dfrac{\wedge\text{-E}\,\dfrac{[A\wedge B]^1}{B} \quad \wedge\text{-E}\,\dfrac{[C\wedge D]}{C}}{B\wedge C} \qquad \wedge\text{-I}\,\dfrac{\wedge\text{-E}\,\dfrac{[B\wedge E]^1}{B} \quad \wedge\text{-E}\,\dfrac{[C\wedge D]}{C}}{B\wedge C} \qquad (A\wedge B)\vee(B\wedge E)}{B\wedge C}}{C}$$

This derivation contains no maximum formulae, or local peaks. Yet, it is not normal: $B \wedge C$ is needlessly introduced and successively eliminated. Our derivation can nevertheless be turned into a more tractable non-normal derivation by applying what Dummett (1977, p. 112) calls a *permutative reduction procedure*, i.e. a reshuffling of the order of the rules for conjunction and disjunction:

$$\vee\text{-E}, 1\,\dfrac{\wedge\text{-E}\,\dfrac{\wedge\text{-I}\,\dfrac{\wedge\text{-E}\,\dfrac{[A\wedge B]^1}{B} \quad \wedge\text{-E}\,\dfrac{[C\wedge D]}{C}}{B\wedge C}}{C} \qquad \wedge\text{-E}\,\dfrac{\wedge\text{-I}\,\dfrac{\wedge\text{-E}\,\dfrac{[B\wedge E]^1}{B} \quad \wedge\text{-E}\,\dfrac{[C\wedge D]}{C}}{B\wedge C}}{C} \qquad (A\wedge B)\vee(B\wedge E)}{C}$$

theorems typically (though not always) entail that the logic satisfies a number of pleasing proof-theoretic properties, such as separability and the subformula requirement. We will discuss these properties, and their philosophical motivations, in Chapter 6, and we will prove a normalization theorem for a formalization of classical logic in Chapter 7.

5.2.4 A necessary but not sufficient condition for harmony

Intrinsic harmony requires that the consequences of a complex statement may not exceed the grounds for that statement. In Prawitz's words: "nothing is gained by inferring a formula through introduction for use as a major premiss in an elimination" (Prawitz, 1965, pp. 33-4). This is an eminently plausible requirement, if logic alone is not to yield new knowledge of atomic statements that we could not have otherwise acquired. Thus, unsurprisingly, the `tonk` rules are not intrinsically harmonious. There is no way one can, in general, transform the following derivation

$$\dfrac{\dfrac{\Gamma, A}{A \text{ tonk } B}\,\text{tonk-I}}{B}\,\text{tonk-E}$$

in a derivation of B from the same or fewer assumptions that does not resort to the `tonk` rules. But can harmony be *identified* with intrinsic harmony?

Not quite. Intrinsic harmony does not prevent E-rules to be weaker than the corresponding I-rules. For instance, consider the following connective, obtaining by conjoining $\wedge$-I with *one half* of $\wedge$-E:

Example 3. *The* $\circ$*-rules:*

$$\circ\text{-I}\ \dfrac{A \quad B}{A \circ B} \qquad \circ\text{-E}\ \dfrac{A \circ B}{A}$$

The rule of $\circ$-E is intrinsically harmonious with the corresponding introductions. And yet, the rules for $\circ$ are *not* intuitively harmonious: the E-rule is too weak—it does not fully exploit the meaning conferred by the corresponding I-rule.

What is needed, then, is something stronger: E-rules should be neither stronger nor *weaker* than the corresponding introductions. Dummett calls this relation *stability*. He writes:

This non-normal derivation can now be reduced in normal form by applying the standard reduction procedures, or conversion steps. Normalization is sometimes identified with intrinsic harmony—the levellability of local peaks or maximum formulae. Thus Read: "Normalization is the requirement that maximum formulae be eliminable, where a maximum formula in a proof is any occurrence of a formula which is both the conclusion of an I-rule and major premise of an E-rule" (Read, 2008, p. 5). This terminology is misleading, however. As we have just seen, there is more to normalization than intrinsic harmony alone—at least if our system contains rules such as $\vee$-E and $\exists$-E.

> A little reflection shows that harmony is an excessively modest demand. [...] The fact that the consequences we conventionally draw from [a statement] are in harmony with these acknowledged grounds shows only that we draw no consequences its meaning does not entitle us to draw. It does not show that we fully exploit that meaning, that we are accustumed to draw all those consequences we should be entitled to draw. [...] Such a balance is surely desirable [...]: we may call it 'stability' (Dummett, 1991b, p. 287)

However, Dummett never really says what stability is: he dedicates one chapter of the *Logical Basis of Metaphysics* to the topic, but there is no proper account of stability there to be found.[9] This leaves inferentialists with three main known alternatives: Tennant's account of harmony as *reflective equilibrium*,[10]what I shall call *strong intrinsic harmony*, and the so-called *general elimination* account of harmony (GE harmony, for short).[11] A recent result by Florian Steinberger (2009b) suggests that Tennant's account may not be a viable one: it sanctions as harmonious obviously unsound rules for the quantifiers; rules which lack the usual restrictions on the parameters.[12] For this reason, we will set aside Tennant's account of harmony for present purposes. In what follows, we will focus on strong intrinsic harmony (§ 5.3) and GE harmony (§ 5.4).

5.3 Strong intrinsic harmony

In two recent papers, Rowan Davies and Frank Pfenning have suggested a natural strengthening of Dummett's and Prawitz's notion of intrinsic harmony. They define two key notions: *local soundness* and *local completeness*.[13] Local soundness just is intrinsic harmony. Local completeness, on the other hand, is the requirement that "we can apply the elimination rules to a judgment to recover enough knowledge to permit reconstruction of the original judgment" (Pfenning and Davies, 2001, pp. 3). At first approximation: E-rules must allow us to reintroduce the complex statements they eliminate (we shall give a more precise definition in a

[9]See Dummett (1991b, Chapter 13). See also Steinberger (2009b, p. 656).

[10]See Tennant (1997) and Tennant (forthcoming).

[11]See Read (2000) and Negri and von Plato (2001).

[12]See Steinberger (2009b, pp. 559-61) for details. Tennant (2010) has recently responded to Steinberger. It seems to me that Tennant's response misses Steinberger's point, but I do not have space to expand on this issue here.

[13]See Davies and Pfenning (2001) and Pfenning and Davies (2001).

moment). Intuitively, local soundness guarantees that E-rules be not too strong. Local completeness aims at guaranteeing that they be not too weak.

This is certainly an improvement on intrinsic harmony. The requirement is satisfied by the rules for $\wedge$, as the following *expansion* shows:

$$\begin{array}{c}\Pi\\ A\wedge B\end{array} \quad \rightsquigarrow_e \quad \dfrac{\dfrac{\begin{array}{c}\Pi\\ A\wedge B\end{array}}{A}\,\wedge\text{-E} \qquad \dfrac{\begin{array}{c}\Pi\\ A\wedge B\end{array}}{B}\,\wedge\text{-E}}{A\wedge B}\,\wedge\text{-I}$$

where $\rightsquigarrow_e$ reads 'can be expanded into'. However, it is not satisfied by our modified rules for $\circ$: its E-rules are too weak, since *both halves* of $\wedge$-E are needed in order to reintroduce $A \wedge B$. Consider now implication. We have already seen that it is locally sound:

$$\to\text{-E}\,\dfrac{\to\text{-I},i\,\dfrac{\begin{array}{c}\Gamma_0,[A]^i\\ \Pi_0\\ B\end{array}}{A\to B} \qquad \begin{array}{c}\Gamma_1\\ \Pi_1\\ A\end{array}}{B} \quad \rightsquigarrow_r \quad \begin{array}{c}\Gamma_1\\ \Pi_1\\ \underbrace{\Gamma_0, A}\\ \Pi_0\\ B\end{array}$$

where $\rightsquigarrow_r$ reads 'is reducible to'. The following expansion shows that it is also locally complete:

$$\begin{array}{c}\Pi\\ A\to B\end{array} \quad \rightsquigarrow_e \quad \to\text{-I},i\,\dfrac{\to\text{-E}\,\dfrac{\begin{array}{c}\Pi\\ A\to B\end{array} \qquad [A]^i}{B}}{A\to B}$$

What about disjunction? Again, we already know that its standard rules are locally sound, as the following reduction reminds us:

$$\vee\text{-E},i\,\dfrac{\vee\text{-I}\,\dfrac{\begin{array}{c}\Gamma_0\\ \Pi_0\\ A\end{array}}{A\vee B} \qquad \begin{array}{c}\Gamma_1,[A]^i\\ \Pi_1\\ C\end{array} \qquad \begin{array}{c}\Gamma_2,[B]^i\\ \Pi_2\\ C\end{array}}{C} \quad \rightsquigarrow_r \quad \begin{array}{c}\Gamma_0\\ \Pi_0\\ \underbrace{\Gamma_1, A}\\ \Pi_1\\ C\end{array}$$

Whether the standard disjunction rules satisfy local completeness, however, depends how local completeness is defined.

If it is defined as the requirement that an application of $\$$-E can always be followed by an application of $\$$-I, then $\vee$-I and $\vee$-E do *not* satisfy local completeness: an application of $\vee$-E immediately followed by an application of $\vee$-I does not in general allow us to reintroduce $A \vee B$. But if we do not ask that there be an *ordering* in the application of I- and E-rules, then local completeness can be defined in such a way that the disjunction rules *are* locally complete. Nissim Francez and Roy Dyckhoff offer the following definition:

Definition 2. *(Local completeness) The E-rules for \$ are locally complete if and only if "every derivation of a formula A with principal operator \$ can be expanded to one containing an application of an E-rule of \$, and applications of all I-rules of \$ each with conclusion A" (Francez and Dyckhoff, 2009, p. 9).*

The following expansion for $\vee$ shows that the rules for disjunction are also locally complete in Francez and Dyckhoff's sense:

$$\begin{array}{c}\Pi \\ A \vee B\end{array} \quad \rightsquigarrow_e \quad \dfrac{\begin{array}{c}\Pi \\ A \vee B\end{array} \quad \dfrac{[A]^i}{A \vee B}\,\vee\text{-I} \quad \dfrac{[B]^i}{A \vee B}\,\vee\text{-I}}{A \vee B}\,\vee\text{-E}, i$$

Let us call a notion of harmony requiring that I- and E-rules satisfy both local soundness and local completeness *strong intrinsic harmony*:

Definition 3. *(Strong intrinsic harmony) A pair of I- and E-rules for a logical operator \$ satisfies strong intrinsic harmony if and only if it satisfies both local soundness and local completeness.*

It may be objected that strong intrinsic harmony does not allow us to *derive* harmonious E-rules from the corresponding I-rules, nor does it allow us to derive harmonious I-rules from the corresponding E-rules. The reason is simple: reductions and expansions are procedures that may be applied to *existing* pairs of rules—they do not allow us to *produce* new rules. Hence, strong intrinsic harmony can at best *justify* rules with respect to basic rules: it is not a general procedure for generating harmonious rules.

This is not a decisive objection, however. Nothing in our intuitive notion of harmony suggests that E-rules must be *derivable* from the corresponding I-rules, and *vice versa*—although this would certainly be a very welcome feature for an account of harmony to have. We shall discuss a more serious objection to strong intrinsic harmony in § 5.5 below. For the time being, let us now turn to a second possible account of stability: GE harmony.

5.4 General elimination harmony

Unlike strong intrinsic harmony, GE harmony delivers a *procedure* for generating harmonious rules. The account can be traced back to the pioneering work by Paul Lorenzen, in the second half of the 50's. It was first introduced by Per Martin-Löf in the mid-seventies, and it has more recently been developed by a number of authors: Roy Dyckhoff, Nissim Francez, Jan von Plato, Stephen Read, Peter

Schroeder-Heister, and Neil Tennant, to mention but a few.[14] Following Lorenzen (1955), Sara Negri and Jan von Plato (2001) suggest the following recipe for deriving harmonious—and indeed stable—E-rules from arbitrary I-rules: arbitrary consequences of A's canonical grounds should follow from A itself. This yields what Negri and von Plato call the *Generalized Inversion Principle*:

Generalized Inversion Principle

Whatever follows from the canonical grounds for asserting A must also follow from A.

The principle is standardly taken to be formally represented by what we may call the *GE schema*. Let \$ be A's main logical operator, and let $\pi_1, \ldots \pi_m$ be the severally sufficient and jointly necessary grounds for asserting A, where each π_i schematically represents either a sentence or a derivation. Then, \$'s I-rules are as follows:

$$\frac{\pi_1 \quad \ldots \quad \pi_j}{A}\,\text{\$-I}_1 \ldots \frac{\pi_k \quad \ldots \quad \pi_m}{A}\,\text{\$-I}_o$$

In order to find the corresponding GE harmonious rule, we now need to require that everything that follows from each of the π_i's also follows from A itself. If π is a derivation,

$$\pi := \begin{array}{c} \alpha \\ \vdots \\ \gamma \end{array}$$

let γ be a *ground* for A, and α be the *support* of this ground. Now let Σ be the collection of all the supports in Π_i. Then, a first approximation of the GE schema can be given as follows:

$$\frac{A \quad \Sigma \quad \begin{array}{c}[\pi_1]^k \\ \vdots \\ C\end{array} \quad \ldots \quad \begin{array}{c}[\pi_m]^k \\ \vdots \\ C\end{array}}{C}\,\text{\$-E}, k$$

The intuitive idea is that, given an assertion of A, and derivation(s) of C from each of the grounds for A, one may infer C and discharge those grounds. This yields the following definition of harmony:

Definition 4. *(GE harmony) A pair of I- and E-rules is GE harmonious if and only if the E-rule has been induced from the I-rule by means of (a suitable representation of) the GE schema.*

[14]See e.g. Francez and Dyckhoff (2009), von Plato (2001), Read (2000), Schroeder-Heister (1984), and Tennant (1992).

Some examples may prove useful.

Consider conjunction first. Its canonical grounds are A, B. Accordingly, the GE schema yields the following harmonious general elimination rule:

$$\dfrac{A \wedge B \qquad \begin{matrix}[A, B]^k \\ \vdots \\ C\end{matrix}}{C}\ \wedge\text{-E}_{GE}, k$$

The standard rule of $\wedge$-E can be derived as a special case, setting C equal to A, B.

Consider now implication. A conditional $A \rightarrow B$ may be canonically introduced if we have a derivation of B from A. Accordingly, the GE schema requires that whenever we have a proof of the support A, we may, given a proof of $A \rightarrow B$, infer whatever follows from our ground B. In symbols:

$$\dfrac{A \rightarrow B \qquad A \qquad \begin{matrix}[B]^k \\ \vdots \\ C\end{matrix}}{C}\ \rightarrow\text{-E}_{GE}, k$$

Again, we can easily derive *modus ponens* as a special case, by setting C equal to B.

It is easy to check that neither tonk nor $\circ$ are GE harmonious, as desired. Consider tonk first. The GE schema dictates the following harmonious E-rule:

$$\dfrac{A \texttt{ tonk } B \qquad \begin{matrix}[A]^k \\ \vdots \\ C\end{matrix}}{C}\ \texttt{tonk}\text{-E}_{GE}, k$$

By setting C equal to A, we see that the GE harmonious elimination for tonk allows us to infer from tonk-statements precisely what was required to introduce them in the first place. As for $\circ$, we have already seen that its I-rule, the introduction rule for conjunction, induces via the GE schema *conjunction* elimination—not $\circ$-E.

5.5 Quantum disjunction

It is now time to consider a common objection to both strong intrinsic harmony and GE harmony: that they do not respect Gentzen's requirement that eliminations be *functions* of the corresponding introductions. Section 5.5.1 introduces the problem. Section 5.5.2 shows where the objection goes wrong.

5.5.1 Harmony and the quantum rules

Let us consider strong intrinsic harmony first. Following Dummett (1991b, pp. 289-90), consider the rules for *quantum disjunction*. These rules are just like the standard ones, except from the fact that the E-rule disallows side assumptions in the assumptions for discharge, A and B:

Example 4. *Quantum disjunction:*

$$\frac{\Gamma, A}{A \sqcup B} \quad \frac{\Delta, B}{A \sqcup B}\,{\sqcup\text{-}I} \qquad \frac{A \sqcup B \quad \begin{matrix}[A]^i \\ \vdots \\ C\end{matrix} \quad \begin{matrix}[B]^i \\ \vdots \\ C\end{matrix}}{C}\,{\sqcup\text{-E}, i}$$

The *standard* E-rule for $\wedge$ allows to prove the *distributive law*

$$\frac{A \wedge (B \wedge C)}{(A \wedge B) \vee (A \wedge C)}$$

as follows:

$$\dfrac{\dfrac{A \wedge (B \vee C)}{B \vee C}\,{\wedge\text{-E}} \quad \dfrac{\dfrac{\dfrac{A \wedge (B \vee C)}{A}\,{\wedge\text{-E}} \quad [B]^1}{A \wedge B}\,{\wedge\text{-I}}}{(A \wedge B) \vee (A \wedge C)}\,{\vee\text{-I}} \quad \dfrac{\dfrac{\dfrac{A \wedge (B \vee C)}{A}\,{\wedge\text{-E}} \quad [C]^1}{A \wedge C}\,{\wedge\text{-I}}}{(A \wedge B) \vee (A \wedge C)}\,{\vee\text{-I}}}{(A \wedge B) \vee (A \wedge C)}\,{\vee\text{-E}, 1}$$

The proof crucially relies on the possibilities of wheeling in side assumptions—in our example, $A \wedge (B \vee C)$—in the application of $\vee$-E. The modified rule of $\sqcup$-E, however, disallows the possibility of adding side assumptions, and therefore invalidates the above proof, and, with it, the distributive law.

Perhaps surprisingly, the rules for $\sqcup$ appear to be strongly intrinsically harmonious if and only if the rules for $\vee$ are. For deductions where $A \sqcup B$ is introduced and then immediately eliminated are always transformable into simpler deductions from the same or fewer assumptions that do not pass through $A \sqcup B$. Moreover, the rule of $\sqcup$ does not prevent the possibility of *expanding* proofs of $A \sqcup B$, since the expansion procedure we have just given for $\vee$ also holds for $\sqcup$, since it does not rely on the possibility of wheeling in side assumptions in the assumptions for discharge:

$$\begin{matrix}\Pi \\ A \sqcup B\end{matrix} \quad \rightsquigarrow_e \quad {\sqcup\text{-E}, i}\,\dfrac{\begin{matrix}\Pi \\ A \sqcup B\end{matrix} \quad {\sqcup\text{-I}}\,\dfrac{[A]^i}{A \sqcup B} \quad {\sqcup\text{-I}}\,\dfrac{[B]^i}{A \sqcup B}}{A \sqcup B}$$

If harmony is strong intrinsic harmony, it would seem, both $\vee$-E and $\sqcup$-E are in harmony with *the same* I-rule. And yet, $\sqcup$-E is weaker than $\vee$-E. This may suggest that strong intrinsic harmony is stronger than intrinsic harmony, but not strong enough. Any viable account of harmony should validate one, and only one, set of harmonious E-rules for disjunction, given the standard rules of $\vee$-I.

The problem, it may be thought, equally afflicts GE harmony (see Steinberger, 2009a, p. 138). The point is disarmingly simple: the GE schema (let alone the Generalized Inversion Principle) appears to be simply silent on the issue whether one should allow side assumptions in the premises for discharge. Hence, both $\vee$-E and its quantum cousin $\sqcup$-E

$$\dfrac{A \vee B \qquad \begin{matrix}\Gamma, [A]^i \\ \vdots \\ C\end{matrix} \qquad \begin{matrix}\Delta, [B]^i \\ \vdots \\ C\end{matrix}}{C}\,{\scriptstyle \sqcup\text{-E}, i} \qquad\qquad \dfrac{A \sqcup B \qquad \begin{matrix}[A]^i \\ \vdots \\ C\end{matrix} \qquad \begin{matrix}[B]^i \\ \vdots \\ C\end{matrix}}{C}\,{\scriptstyle \sqcup\text{-E}, i}$$

satisfy the demand that anything that follows from either A lor B also follows from, respectively, $A \wedge B$ and $A \sqcup B$. Thus, it would seem, *both* rules are GE harmonious with respect to the *same* I-rule.

What are we to conclude? Following Dummett (1991b, p. 290), Steinberger (2009a) argues that one can *show* that the quantum rules for disjunction fail to confer to $\sqcup$ a stable meaning. He first observes that, if we start with a system S containing only $\wedge$ and $\sqcup$, and we successively add $\vee$ with its unrestricted $\vee$-E to S, the new system, call it S', yields a *non-conservative* extension of S: that is, S' licenses new inferential relations among the expressions of S's language. To see this, it is sufficient to observe that, in S', quantum disjunction collapses on the standard one, as the following derivation shows:

$${\scriptstyle \sqcup\text{-E}, i}\,\dfrac{A \sqcup B \qquad {\scriptstyle \vee\text{-I}}\,\dfrac{[A]^i}{A \vee B} \qquad {\scriptstyle \vee\text{-I}}\,\dfrac{[B]^i}{A \vee B}}{A \vee B}$$

As a result, the S-invalid distributivity law

$$\dfrac{A \wedge (B \sqcup C)}{(A \wedge B) \sqcup (A \wedge C)}$$

becomes now derivable in S'. Secondly, Steinberger (2009a, pp. 82-3) notices that the new system S' is not normalizable: although one can always level local peaks, the permutative reduction procedures we mentioned in § 5.2.1, fn. 8, do not always apply. Steinberger concludes that

> Dummett has produced a system composed exclusively of intrinsically harmonious pairs of sets of inference rules that is nonetheless not normalizable and does not display total harmony [i.e. conservativeness]. This shows that the [...] inference rules for $\sqcup$ failed to fix its meaning. Steinberger (2009a, pp. 82-3)

However, this conclusion appears to be mistaken, for at least two reasons. First, it is difficult to see why a failure of normalizability should have semantic implications, in absence of an argument linking meaning-constitution and the availability of a *normalization theorem*—an argument that, to my knowledge, has not yet been provided. Second, the fact that the $\vee$-rules are not conservative over the $\sqcup$-rules by no means imply that there is something amiss with the $\sqcup$-rules. Let $\sim$ and $\neg$ be, respectively, intuitionistic and classical negation. A well-known result by J. L. Harris (1982) shows that the addition of a classical negation $\neg$ to intuitionistic logic yields both $\sim A \vdash \neg A$ and $\neg A \vdash \sim A$. Moreover, as we shall see in Chapter 6, the rules for $\neg$ yield a non-conservative extension of intuitionistic logic—among other things, they allow us to prove the intuitionistically invalid Peirce's Law, $((A \rightarrow B) \rightarrow A) \rightarrow A$. Yet, it certainly does not follow from this that the intuitionistic rules for negation fail to confer to $\sim$ a stable meaning!

All the same, it would seem that there *is* something amiss with the quantum rules. On the one hand, the I-rule allows to introduce $A \vee B$ from either Γ, A or Δ, B. On the other, the E-rule allows to infer from $A \sqcup B$ whatever follows from both A and B *alone*. That is, our contexts Γ and Δ have now disappeared from the scene. As a result, $\sqcup$-E is weaker than $\vee$-E; as we have seen, only the latter allows us to prove $(A \wedge B) \vee (A \vee C)$. But was not harmony supposed to guarantee that E-rules be neither stronger *nor weaker* than the corresponding I-rules? How can the *same* I-rule justify E-rules of different strength?

5.5.2 What is wrong with quantum disjunction

I suggest that it is a mistake to think that quantum disjunction poses a problem for strong intrinsic harmony and GE harmony. If we take it that the grounds for asserting disjunctions are either Γ, A or Δ, B, then *these* grounds should figure in the elimination rule for disjunction. Analogously, if we take it that the grounds for asserting disjunctions are either A or B, then, again, *these* grounds should figure in the elimination rule for disjunction. Mixed cases, such as the quantum rules, are intuitively disharmonious. That is, intuitively, only the following two combinations should be admissible: either the standard rules for disjunction, or the following very weak rules:

$$\frac{A}{A \sqcup B} \quad \frac{B}{A \sqcup B}\ {\scriptstyle \sqcup\text{-I}^*} \qquad \frac{A \sqcup B \qquad \begin{matrix}[A]^i\\ \vdots \\ C\end{matrix} \qquad \begin{matrix}[B]^i\\ \vdots \\ C\end{matrix}}{C}\ {\scriptstyle \sqcup\text{-E}, i}$$

My contention is that, on closer inspection, strong intrinsic harmony and GE precisely tell us this much.

Let us start from GE harmony. The Generalized Inversion Principle tells us that whatever follows from the grounds from introducing $A \sqcup B$, as specified by the I-rules for $\sqcup$, should also follow from $A \sqcup B$. However, if the canonical grounds for $A \sqcup B$ are Γ, A and Γ, B, the Generalized Inversion Principle tells us that whatever follows from Γ, A and Γ, B should also follow from $A \sqcup B$. That is, the Generalized Inversion Principle, and the GE schema it validates, yield $\vee$-E, not $\sqcup$-E, as the harmonious E-rule for $\sqcup$. On the other hand, if the grounds for $A \sqcup B$ are A and B (without side assumptions), as I have suggested, then the Generalized Inversion Principle effectively yields $\sqcup$-E. In short, GE harmony validates the harmonious pairs $\{\vee\text{-I}, \vee\text{-E}\}$ and $\{\sqcup\text{-I}^*, \sqcup\text{-E}\}$, but not the hybrid $\{\sqcup\text{-I}, \sqcup\text{-E}\}$, as it should be.

Essentially the same reasoning applies in the case of strong intrinsic harmony. Recall, local completeness, i.e. the possibility of carrying out *expansions*, was supposed to guarantee that E-rules be not too weak. Our problem, then, was that the following expansion seemed to show that the disharmonious pair $\{\sqcup\text{-I}, \sqcup\text{-E}\}$ passes the test:

$$\begin{matrix}\Pi \\ A \sqcup B\end{matrix} \quad \rightsquigarrow_e \quad {\scriptstyle \sqcup\text{-E}, i}\ \frac{\begin{matrix}\Pi \\ A \sqcup B\end{matrix} \qquad {\scriptstyle \sqcup\text{-I}}\ \dfrac{[A]^i}{A \sqcup B} \qquad {\scriptstyle \sqcup\text{-I}}\ \dfrac{[B]^i}{A \sqcup B}}{A \sqcup B}$$

Or does it? Recall, $\sqcup$-I is *just like* $\vee$-I, i.e. it allows side assumptions. Hence, the expansion should rather read:

$$\begin{matrix}\Pi \\ A \sqcup B\end{matrix} \quad \rightsquigarrow_e \quad {\scriptstyle \sqcup\text{-E}, i}\ \frac{\begin{matrix}\Pi \\ A \sqcup B\end{matrix} \qquad {\scriptstyle \sqcup\text{-I}}\ \dfrac{\Gamma, [A]^i}{A \sqcup B} \qquad {\scriptstyle \sqcup\text{-I}}\ \dfrac{\Delta, [B]^i}{A \sqcup B}}{A \sqcup B}$$

But now, this expansion is still mistaken at it stands: $\sqcup$-E does *not* allow side assumptions. Hence, it should be rewritten as:

$$\begin{matrix}\Pi \\ A \sqcup B\end{matrix} \quad \rightsquigarrow_e \quad {\scriptstyle \underline{\vee}\text{-E}, i}\ \frac{\begin{matrix}\Pi \\ A \sqcup B\end{matrix} \qquad {\scriptstyle \sqcup\text{-I}}\ \dfrac{\Gamma, [A]^i}{A \underline{\vee} B} \qquad {\scriptstyle \sqcup\text{-I}}\ \dfrac{\Delta, [B]^i}{A \underline{\vee} B}}{A \underline{\vee} B}$$

This expansion, however, shows that the pair $\{\sqcup\text{-I}, \vee\text{-E}\}$, i.e. $\{\vee\text{-I}, \vee\text{-E}\}$ (recall, $\sqcup$-I and $\vee$-I are mere notational variants), is locally complete, as it should be. Similarly, what the first expansion shows is that the *harmonious* pair $\{\sqcup\text{-I}^*, \sqcup\text{-E}\}$ is also locally complete, as it should be. It should then be rewritten as follows:

$$\begin{array}{c}\Pi\\ A \sqcup B\end{array} \quad \rightsquigarrow_e \quad \sqcup\text{-E}, i\, \dfrac{\begin{array}{c}\Pi\\ A \sqcup B\end{array} \quad \sqcup\text{-I}^*\dfrac{[A]^i}{A \sqcup B} \quad \sqcup\text{-I}^*\dfrac{[B]^i}{A \sqcup B}}{A \sqcup B}$$

The rules for quantum disjunction, I conclude, do not pass the harmony test, irrespective of whether harmony is defined as GE harmony or strong intrinsic harmony.

5.6 Harmony as full invertibility

If the foregoing considerations are correct, both strong intrinsic harmony and GE harmony satisfy Gentzen's functionality requirement, that E-rules can be displayed as unique functions of the corresponding I-rules. Neither account of harmony, however, respects in general the requirement that logic alone may not license losses of knowledge. To see the problem, consider the standard rule of *modus ponens*. The rule tells us that, if we can assert $A \to B$, then we may infer B from A: precisely what was required for introducing $A \to B$ in the first place. This pleasing symmetry, though, is lost as soon as GE rules enter into the scene. Consider the standard elimination rules for $\vee$ and $\exists$:

$$\vee\text{-E}, i\, \dfrac{A \vee B \quad \begin{array}{c}\Gamma_0, [A]^i\\ \vdots\\ C\end{array} \quad \begin{array}{c}\Gamma_1, [B]^i\\ \vdots\\ C\end{array}}{C} \qquad \exists\text{-E}, i\, \dfrac{\begin{array}{c}\Gamma_0\\ \vdots\\ \exists x F(x)\end{array} \quad \begin{array}{c}\Gamma_1, [F[a/x]]^i\\ \vdots\\ C\end{array}}{C}.$$

Although both $A \vee B$ and $\exists x F x$ can occur in the conclusions of, respectively, $\vee$-E and $\exists$-E, they cannot in general be reintroduced by means of their I-rules. That is, the following derivations are both incorrect:

$$\vee\text{-I}\,\dfrac{\vee\text{-E}, i\, \dfrac{A \vee B \quad \begin{array}{c}\Gamma_0, [A]^i\\ \vdots\\ C\end{array} \quad \begin{array}{c}\Gamma_1, [B]^i\\ \vdots\\ C\end{array}}{C}}{A \vee B} \qquad \exists\text{-I}\,\dfrac{\exists\text{-E}, i\, \dfrac{\begin{array}{c}\Gamma_0\\ \vdots\\ \exists x F(x)\end{array} \quad \begin{array}{c}\Gamma_1, [F[a/x]]^i\\ \vdots\\ C\end{array}}{C}}{\exists x F(x)}.$$

Thus Alan Weir writes that "C need not be identical with A_i, the premise of the original application of $\vee$-I" (Weir, 1986, p. 464), and that "one cannot derive, by

$\exists$-I alone, $\exists x F(x)$ from an arbitrary conclusion, C, of $\exists E$" (Weir, 1986, p. 471). The E-rules for $\vee$ and $\exists$ do not extract all that is contained in their corresponding I-rules: in Weir's phrase, there is some "leakage" (Weir, 1986, p. 467). Successive application of the I- and E-rules for a constant \$ may yield a loss of information, in contrast with the epistemic argument for harmony we introduced in § 5.1. They do not allow us to recover for complex statements what was required to introduce them in the first place.

To be sure, inferentialists may learn to live with the idea that harmony does not always guarantee symmetry between I- and E-rules, in the sense that E-rules give us back, so to speak, the conditions expressed by the corresponding I-rules. But this does not mean that they have to. There *are* accounts of harmony on which I- and E-rules are mutually interderivable, as we shall see in this section. We shall consider two such accounts: one offered by Weir (§§ 5.6.1-3), and an improved version of it (§§ 5.6.4-5).

5.6.1 Weir's account of harmony

Weir's account of harmony is informed by one key idea: harmonious E-rules should give us back *precisely* the canonical grounds expressed by the corresponding I-rules. Accordingly, Weir requires that harmonious I- and E-rules obey a *bipartite inversion principle,* whose first half allows us to generate harmonious E-rules given arbitrary I-rules, and whose second half, conversely, allows us to generate harmonious I-rules given arbitrary E-rules. Let us have a closer look.

Weir's inversion principle: first half

Weir introduces the first half of his principle thus:

> When the sufficient condition for application of an I-rule obtains, application of that rule followed immediately by application of elimination rules for the relevant constant returns us to the sufficient condition for application of the I-rule. (Weir, 1986, p. 466)

He provides the following schematic representation. Let

$$\dfrac{\begin{matrix}[A^i_1] & & [A^i_{r(i)}] \\ \vdots & & \vdots \\ P^i_1 & & P^i_{r(i)}\end{matrix}}{\star(C_1,\ldots,C_n)}\,\star\text{-I}^i$$

be a schematic representation of a set of I-rules for a n-ary constant $\star$. The i's denotes the number of I-rules for $\star$-statements, and $r(i)$ denotes the number of premises for each I-rule. The P's denote the *grounds* for $\star$-statements, in the terminology we introduced in § 5.3.3, and the A's denote the possibly empty *supports* for those grounds. Now let

$$\begin{array}{c}\Pi \\ \star(C_1, \ldots, C_n)\end{array}$$

be

$$\dfrac{\begin{array}{c}[A^i_1] \\ \vdots \\ P^i_1\end{array} \quad \begin{array}{c}[A^i_1] \\ \vdots \\ P^i_{r(i)}\end{array}}{\star(C_1, \ldots, C_n)}\ \star\text{-I}^i.$$

Then, the first half of Weir's inversion principle requires that, if $\star(C_1, \ldots, C_n)$ is derivable from Π, the following holds:

$$\dfrac{\begin{array}{c}\Pi \\ \star(C_1, \ldots, C_n)\end{array} \quad [A^i_1] \ldots [A^i_{r(i)}]}{P^i_1}\ \star\text{-E}^i_1 \quad \vdots\vdots\vdots \quad \dfrac{\begin{array}{c}\Pi \\ \star(C_1, \ldots, C_n)\end{array} \quad [A^i_1] \ldots [A^i_{r(i)}]}{P^i_{r(i)}}\ \star\text{-E}^i_{r(i)}$$

That is, the principle requires that, for every rule of $\star$-I$_i$, there are $r(i)$ corresponding rules of $\star$-E^i, such that derivations of $\star(C_1, \ldots, C_n)$ according to $\star$-I^i, when combined with derivations of the minor premises of $\star$-E^i, return the grounds for asserting $\star(C_1, \ldots, C_n)$ according to $\star$-I^i.

Notice the multiplication of E-rules. For each logical operator $\star$, there can be several I-rules, each of which can have several premises. Accordingly, for each I-rule R, with n premises, there will be precisely n harmonious E-rules. Now, *this* is by no means a novelty: the standard rule of $\wedge$-I, for instance, has two premises (in our terminology—see *supra*, § 5.4—two grounds with no support) to which correspond two E-rules. What *is* striking, however, is that Weir requires that there be exactly as many E-rules for $\star$ as there are *I-rules*. We shall return to this point in §§ 5.6.2-4 below.

Weir's inversion principle: second half

Weir supplements the first half of his principle by a second half, to the effect that E-rules should be strong enough to allow us to immediately reintroduce the complex statement they allow us to eliminate. In Weir's words:

> When the sufficient condition for application of an E-rule obtains, application of that rule followed immediately by application of introduction rules for the relevant constant returns us to the sufficient condition for application of the E-rule. (Weir, 1986, p. 467)

Weir's schematic representation is forthcoming. Let

$$\begin{array}{c}\Pi \\ E_i\end{array}$$

be

$$\dfrac{\star(C_1,\ldots,C_n) \qquad \begin{array}{c}[A_1^i] \\ \vdots \\ M_1^i\end{array} \qquad \begin{array}{c}[A_{r(i)}^i] \\ \vdots \\ M_{r(i)}^i\end{array}}{E_i}\ \star\text{-E.}$$

Then, the second part of Weir's inversion principle requires that, if E_i is derivable from Π, the following holds:

$$\dfrac{\begin{array}{c}[M_1^1]^j\ [M_{r(i)}^i]^j \\ \dfrac{\Pi}{E_1}\ \star\text{-E}\end{array} \qquad \vdots\vdots\vdots\ \ \vdots\vdots\vdots \qquad \begin{array}{c}[M_1^1]^j\ [M_{r(i)}^i]^j \\ \dfrac{\Pi}{E_s}\ \star\text{-E}\end{array}}{\star(C_1,\ldots,C_n)}\ \star\text{-I}, j.$$

That is, for every t $(0 \leq t \leq v)$, the t^{th} E-rule for $\star$ must be strong enough to allow us to reintroduce $\star$-statements by means of their t^{th} I-rule. This is how Weir interprets the requirement that, on the one hand, the E-rules for $\star$ give us back *precisely* the conditions for introducing $\star$-statements, as given by their I-rules, while, on the other, the conditions for introducing complex statements expressed by the I-rules for $\star$ be *precisely* what one may infer from $\star$-statements according to the E-rules for $\star$. This allows us to effectively derive E-rules from arbitrary I-rules, and *vice versa*—the inferentialist's pipe dream.

Let us see, then, Weir's principle at work. By way of example, consider the standard rules for implication. The following derivation shows that they satisfy the first half of Weir's inversion principle:

$$\rightarrow\text{-E}, i\ \dfrac{\rightarrow\text{-I}\ \dfrac{\begin{array}{c}[A]^i \\ \vdots \\ B\end{array}}{A \rightarrow B} \qquad A}{B}\ .$$

An application of →-I immediately followed by an application of →-E gives us back the necessary and sufficient condition for applying →-I, viz. that B is derivable from A. The following derivation

$$\dfrac{\dfrac{A \to B \qquad [A]^i}{B}\ \to\text{-E}}{A \to B}\ \to\text{-I}, i \; .$$

shows that the rules for implication also satisfy the second part of Weir's principle: an application of →-E immediately followed by an application of →-I gives us back the necessary and sufficient condition for applying →-E, viz. that both $A \to B$ and A hold. It easy to check that the standard rules for conjunction, negation, and the universal quantifier are also harmonious in Weir's sense.

5.6.2 Weir's rules for disjunction

Given Weir's inversion principle, the rules for ∨ and ∃ do not return us the grounds for asserting, respectively, disjunctive and existential statements. The rule of ∨-E does not give us back either A or B, and the rule of ∃-E does not give us back our witness, so to speak. Hence, the standard rules for ∨ and ∃ are not harmonious in Weir's sense. How to solve the problem? Weir's idea is to take this to show that there is something wrong with the standard rules for ∨ and ∃: we must give *new* rules for these operators—rules that are harmonious in Weir's sense.

Weir's proposed rules for disjunction are as follows:

$$\vee\text{-I}^1_W, n\ \dfrac{\begin{matrix}[\neg B]^n \\ \vdots \\ A\end{matrix}}{A \vee B} \qquad \vee\text{-E}^1_W\ \dfrac{A \vee B \qquad \neg B}{A} \qquad \vee\text{-I}^2_W, n\ \dfrac{\begin{matrix}[\neg A]^n \\ \vdots \\ B\end{matrix}}{A \vee B} \qquad \vee\text{-E}^2_W\ \dfrac{A \vee B \qquad \neg A}{B} \; .$$

And here are his suggested revised rules for the existential quantifier:

$$\exists\text{-I}_W, i\ \dfrac{\begin{matrix}[\neg x = t \to \neg Fx]^i \\ \vdots \\ Ft\end{matrix}}{\exists x Fx} \qquad \exists\text{-E}_W\ \dfrac{\exists x Fx \qquad \neg x = t \to \neg Fx}{Ft} \; .$$

Unlike the standard ones, it is easily verified that these rules satisfy Weir's inversion principle: in both cases, the E-rule gives us back precisely what was required for introducing the complex statement it allows us to eliminate. These classical rules allow us to introduce $A \vee \neg A$ by means of an argument ending by one application of disjunction introduction (Weir, 1986, p. 469):

$$\cfrac{\begin{matrix}[\neg A]^1\\ \vdots\\ \neg A\end{matrix}}{A \vee \neg A}\ \vee\text{-I}^1_W, 1 \ .$$

The rule of double negation elimination is derived as follows:

$$\cfrac{\cfrac{\cfrac{}{\neg A}\,(1)}{A \vee \neg A}\,\vee\text{-I}^1_W\,(1) \qquad \neg\neg A}{A}\ \vee\text{-E}^1_W \ .$$

A first point to notice is that these rules are *only pairwise* intrinsically harmonious. Collectively, they give rise to local peaks that cannot be levelled, as the following derivation shows (see Weir, 1986, pp. 476-8):

$$\cfrac{\cfrac{\begin{matrix}\Delta, [\neg B]^i\\ \Pi_0\\ A\end{matrix}}{A \vee B}\,\vee\text{-I}^1_W, i \qquad \begin{matrix}\Gamma\\ \Pi_1\\ \neg A\end{matrix}}{B}\ \vee\text{-E}^2_W \ .$$

Here there is no way one can in general derive B from a derivation of A from $\neg B$, without appealing to Weir's rules for disjunction. Weir's notion of harmony does not guarantee intrinsic harmony—a serious problem, if it is thought that intrinsic harmony is a necessary condition for harmony. To be sure, Weir will reject this latter claim, on the grounds that he is providing an alternative conception of harmony. However, it is difficult to see why one should not interpret Weir's rules as defining *two* distinct connectives, call them $\vee$ and $\vee^*$, both of which are governed by genuinely harmonious rules. Presumably, Weir will insist that, on *his* account of harmony, only *one* connective is being defined; not two. Weir's insistence, though, quickly leads to disaster. His conception of harmony can be shown to collapse on the disastrous liberality recommended by the early, and naïve, inferentialists.

5.6.3 Tonk strikes back

If harmony is a cure for tonkitis, as it is sometimes said, then Weir's cure is terribly ineffective. A little reflection shows that Weir's bipartite inversion principle validates the following seemingly innocuous rules:

Example 5. *The $\oplus$-rules:*

$$\frac{A}{A \oplus B}\ \oplus\text{-}I_1 \quad \frac{B}{A \oplus B}\ \oplus\text{-}I_2 \qquad \frac{A \oplus B}{A}\ \oplus\text{-}E_1 \quad \frac{A \oplus B}{B}\ \oplus\text{-}E_2.$$

However, it is easy to see that these rules collectively yield disaster.[15] An application of $\oplus$-I$_1$ immediately followed by an application of $\oplus$-E$_2$ yields a derivation of B from A, for arbitrary A's and B's:

$$\dfrac{\dfrac{A}{A \oplus B}\,\oplus\text{-I}_1}{B}\,\oplus\text{-E}_2\,.$$

Prior's `tonk` strikes back! Weir's conception of harmony validates the rules for `tonk`. On the plausible assumption that the `tonk` rules cannot be harmonious, Weir's proposed account has been shown to be in adequate.

What has gone wrong? The obvious diagnosis, I take it, is that harmony indeed requires intrinsic harmony, and the $\oplus$-rules fail to define a *single* connective. They are really defining *two* connectives, whose harmonious I- and E-rules are, respectively, $\oplus$-I$_1$ and $\oplus$-E$_2$, and $\oplus$-I$_2$ and $\oplus$-E$_2$. Only on the further assumption that these two connectives mean the same do the foregoing rules enjoin disaster. Yet, it is difficult to see *why*, from an inferentialist perspective, these two connectives should mean the same, given that they are governed by different I- and E-rules. If, as I have suggested, Weir's conception of harmony is inadequate, we cannot but conclude that Weir's formalization of classical logic rests on implausible semantic assumptions, viz. that distinct pairs of (genuinely) harmonious I- and E-rules can define a *single* logical operator.

5.6.4 Harmony as full invertibility

Weir's account of harmony gives us a *procedure* for deriving 'harmonious' rules; one that can work two ways: from I-rules to E-rules, and *vice versa*. Unfortunately, as we have seen, the procedure fails to produce, as it stands, *harmonious* rules. Is it possible to solve this problem? Proof-theorists are generally sceptical. Thus, Wagner de Campos Sanz and Thomas Piecha have recently argued that symmetry

> is not present in natural deduction, which makes the formulation of an inversion principle based on elimination rules rather than introduction rules quite difficult. (De Campos Sanz, 2009, p. 551)

This scepticism is unjustified, however. We have seen that Weir's inversion principle is too liberal: it allows us to introduce logical expressions with multiple I-rules *and* multiple E-rules. One way to solve the problem is to require that, for

[15] Thanks to Dominic Gregory for helpful exchanges on this point. I am here applying to Weir an objection Dominic had raised against an account of harmony I have been exploring while writing my PhD.

each operator $\star$, there be at most *one* canonical way for introducing $\star$, i.e. one I-rule with r supports and r corresponding grounds ($0 \leq r \leq m$). The result is the following restricted version of Weir's original bipartite inversion principle.

Emended inversion principle: first half

The first part of the emended principle now tells us that, given a I-rule for $\star$-statements, with r grounds and at most r supports, the harmonious E-rules allow us to infer, given an assertion of a $\star$-statement, each ground from the corresponding support. More formally, we require that given an arbitrary I-rule for an arbitrary n-ary logical operator $\star$

$$\dfrac{\begin{matrix}[A_1]^j\\ \vdots\\ P_1\end{matrix} \qquad \begin{matrix}[A_r]^j\\ \vdots\\ P_r\end{matrix}}{\star(C_1,\ldots,C_n)}\,{\scriptstyle \star\text{-I}, j}$$

the harmonious E-rule for $\star$ allows us to derive $\star(C_1,\ldots,C_n)$'s grounds, given derivations of its supports:

$$\dfrac{\dfrac{\begin{matrix}[A_1]^j\\ \vdots\\ P_1\end{matrix} \qquad \begin{matrix}[A_1]^j\\ \vdots\\ P_r\end{matrix}}{\star(C_1,\ldots,C_n)}\,{\scriptstyle \star\text{-I}, j} \qquad A_1}{P_1}\,{\scriptstyle \star\text{-E}} \qquad \vdots\vdots\vdots \qquad \dfrac{\dfrac{\begin{matrix}[A_r]^j\\ \vdots\\ P_r\end{matrix} \qquad \begin{matrix}[A_r]^j\\ \vdots\\ P_r\end{matrix}}{\star(C_1,\ldots,C_n)}\,{\scriptstyle \star\text{-I}, j} \qquad A_r}{P_r}\,{\scriptstyle \star\text{-E.}}$$

In short: the E-rules for $\star$ must give us back precisely what was required to introduce $\star$-statements in the first place.

Emended inversion principle: second half

Similarly, the second half of the emended principle now tells us that, if we can assert a statement A with principal operator $\star$, then, given derivation of the grounds for asserting A, we may reintroduce A. More formally, we require that, given an arbitrary E-rule for an operator $\star$, a full application of $\star$-E allows us to reintroduce $\star(C_1,\ldots,C_n)$ by an application of $\star$-I:

$$\dfrac{\dfrac{\star(C_1,\ldots,C_n) \qquad [A_1]^j}{P_1}\,{\scriptstyle \star\text{-E}} \qquad \vdots\vdots\vdots \qquad \dfrac{\star(C_1,\ldots,C_n) \qquad [A_r]^j}{P_r}\,{\scriptstyle \star\text{-E}}}{\star(C_1,\ldots,C_n)}\,{\scriptstyle \star\text{-I}, j}$$

In short: the E-rules must be strong enough to give us back the necessary and sufficient condition to apply them again.

Our inversion principle is just like Weir's, except that it does not allow for multiple I-rules. Hence, it does not validate the tonk rules. On the foregoing sense of harmony, the harmonious rule justified by tonk-I tells us that we may infer A from A tonk B; not B. The principle thus provides an effective means of producing harmonious E-rules, given arbitrary I-rules, and *vice versa*. The requirement of symmetry, and Gentzen's requirement of functionality, are both met.

The first half of the modified principle guarantees intrinsic harmony (or local soundness), as the following reduction shows (the remaining $r - 1$ reductions are structurally identical):

$$\dfrac{\dfrac{\begin{matrix}[A_1] \\ \Pi_0 \\ P_1\end{matrix} \qquad \begin{matrix}[A_1]\\ \Pi_1 \\ P_r\end{matrix}}{\star(C_1,\ldots,C_n)}\,\star\text{-I} \qquad \begin{matrix}\Pi_2\\ [A_1]\end{matrix}}{P_1}\,\star\text{-E} \qquad \rightsquigarrow_r \qquad \begin{matrix}\Pi_2\\ A_1\\ \Pi_0\\ P_1\end{matrix}$$

There is no need to introduce and eliminate $\star$ in order to derive P_1, given that a derivation of P_1 is required in order to introduce $\star$-statements in the first place (the remaining cases are exactly analogous). But the principle also guarantees *strong intrinsic harmony*, as shown by the following expansion:

$$\star(C_1,\ldots,C_n) \rightsquigarrow_e \dfrac{\dfrac{\star(C_1,\ldots,C_n) \qquad [A_1]^j}{P_1}\,\star\text{-E} \qquad \vdots\vdots \qquad \dfrac{\star(C_1,\ldots,C_n) \qquad [A_r]^j}{P_r}\,\star\text{-E}}{\star(C_1,\ldots,C_n)}\,\star\text{-I}, j$$

It easy to check that the standard intuitionistic rules for $\wedge$, $\rightarrow$, $\neg$, and $\forall$ are all validated on the foregoing account of harmony.

I shall say that I- and E-rules satisfying the restricted version of Weir's original principle are *fully invertible*, and I shall call the foregoing account of harmony *harmony as full invertibility*.

Definition 5. *(Full invertibility) A pair of I- and E-rules is fully invertible if and only if it satisfies both halves of the emended version of Weir's inversion principle.*

It may be objected that the account of harmony as full unvertibility works in theory but not in practice, since it obviously does not validate the standard rules for $\vee$ and $\exists$. Indeed, on the foregoing account, $\vee$-I does not even count as a legitimate set of I-rules, given that it only applies to logical operators with at most *one* I-rule. However, I will argue in § 7.4.1 that this worry is ultimately ill-motivated: there are non-standard rules for disjunction (and the existential quantifier) that (i) satisfy the foregoing schemata, and (ii) are interderivable, given sufficiently strong structural assumptions, with the standard rules.

5.7 The Argument from Harmony

It is now time to put the the three accounts of harmony we have introduced to work, and start asking ourselves where the requirement of harmony leads us.

5.7.1 Minimal, intuitionistic, and classical logic

Let us begin with some bold, but true, claims. On any decent account of proof-theoretic harmony, the standard natural deduction rules of *minimal logic* are clearly harmonious, the standard natural deduction rules of *intuitionistic logic* are likely to be harmonious, and the standard natural deduction rules of *classical logic* are clearly *not* harmonious. Let us begin with the first two claims.

The rules for *minimal logic* consist of the standard I- and E-rules for conjunction, disjunction, implication, and negation:

$$
{\scriptstyle \neg\text{-I},\,i}\ \dfrac{\begin{array}{c}\Gamma,[A]^i\\ \vdots\\ \bot\end{array}}{\neg A}
\qquad\qquad
{\scriptstyle \neg\text{-E}}\ \dfrac{A \qquad \neg A}{\bot}\,.
$$

Intuitionistic logic may be obtained by adding either $\bot$-E (also called *ex falso quodlibet*) or disjunctive syllogism to minimal logic:

$$
{\scriptstyle \bot\text{-E}}\ \dfrac{\bot}{A}
\qquad\qquad
{\scriptstyle \text{DS}}\ \dfrac{\neg A \qquad A \land B}{B}\,.
$$

Classical logic may be obtained either by adding double negation elimination, classical *reductio*, or what we may call Peirce's Rule[16]

$$
{\scriptstyle \text{DN}}\ \dfrac{\neg\neg A}{\text{A}}
\qquad\qquad
{\scriptstyle \text{CR},\,i}\ \dfrac{\begin{array}{c}\Gamma,[\neg A]^i\\ \vdots\\ \bot\end{array}}{A}
\qquad\qquad
{\scriptstyle \text{Peirce's Rule},\,i}\ \dfrac{\begin{array}{c}\Gamma,[A \to B]^i\\ \vdots\\ A\end{array}}{A}
$$

to minimal logic, or by adding either DN, or CR, or one between classical dilemma and the Law od Excluded Middle

$$
{\scriptstyle \text{Dilemma},\,i}\ \dfrac{\begin{array}{c}\Gamma,[A]^i\\ \vdots\\ C\end{array} \qquad \begin{array}{c}\Delta,[\neg A]^i\\ \vdots\\ C\end{array}}{C}
\qquad\qquad
{\scriptstyle \text{LEM}}\ \dfrac{}{A \lor \neg A}
$$

[16]See Milne (2002, p. 511).

to intuitionistic logic. The rules for minimal logic all satisfy both strong intrinsic harmony and GE harmony. But so do the rules for intuitionistic logic, at least if we think that the rule of *ex falso quodlibet* satisfies both strong intrinsic harmony and GE harmony. As Francez and Dyckhoff write:

> Although there is no I-rule for $\bot$ (falsehood), the rules for Intuitionistic logic are locally intrinsically harmonious too, because the boundary case of no I-rules vacuously satisfies the requirement. The expansion obtained for $\bot$ is
>
> $$\begin{matrix}\mathcal{D}\\ \bot\end{matrix} \quad \rightsquigarrow_e \quad {\scriptstyle \bot\text{-E}}\,\dfrac{\begin{matrix}\mathcal{D}\\ \bot\end{matrix}}{\bot}$$
>
> containing, indeed, all the (non-existing) $\bot$ I-rules. (Francez and Dyckhoff, 2009, p. 10)

As for GE harmony, if there are *no* grounds for $\bot$, then, trivially, whatever proves $\bot$, and also proves C, which means that, by the Generalized Inversion Principle, C may be correctly inferred from $\bot$ (see Negri and von Plato, 2001, p. 6). On the other hand, it is worth pointing out that $\bot$-E does not satisfy full invertibility.[17]

What about the foregoing classical rules? One problem is that none of them, except from DN, clearly is an I- or an E-rule. Hence, it is hard to see how they could be justified on the assumption that only harmonious I- and E-rules can be meaning-constitutive, or, in Dummett's terminology, self-justifying. Let us focus, at least for the time being, on DN.

5.7.2 Double Negation Elimination

Consider the standard *intuitionistic* rule of $\neg$-E. The following reduction and the following expansion show that the pair $\{\neg\text{-I}, \neg\text{-E}\}$ satisfies strong intrinsic harmony:

$${\scriptstyle \neg\text{-E}}\,\dfrac{{\scriptstyle \neg\text{-I},\,i}\,\dfrac{\begin{matrix}\Gamma_0, [A]^i\\ \Pi_0\\ \bot\end{matrix}}{\neg A} \qquad \begin{matrix}\Gamma_1\\ \Pi_1\\ A\end{matrix}}{\bot} \quad \rightsquigarrow_r \quad \begin{matrix}\Gamma_1\\ \Pi_1\\ \underbrace{\Gamma_0, A}\\ \Pi_0\\ \bot\end{matrix} \qquad\qquad \begin{matrix}\Pi\\ \neg A\end{matrix} \quad \rightsquigarrow_e \quad {\scriptstyle \neg\text{-I},\,i}\,\dfrac{{\scriptstyle \neg\text{-E}}\,\dfrac{\begin{matrix}\Pi\\ \neg A\end{matrix} \qquad [A]^i}{\bot}}{\neg A}$$

[17] We will say more on the inferentialist's interpretation of $\bot$ in §§ 6.5.3-5 and § 7.4.2 below. For the time being, we can anticipate that there are several options available, depending on (i) whether one thinks that $\bot$ has content, and (ii) which $\bot$-rules one takes to be valid. On (i), Tennant (1999), Rumfitt (2000), and Steinberger (2009b) all suggest that $\bot$ is best treated as a logical *punctuation sign*. On (ii), Prawitz and Dummett offer two different I-rules for $\bot$ (see, *infra*, § 6.5.3).

But it also satisfies GE harmony. As applied to ¬-I, the GE schema yields the following elimination rule for negation:

$$\dfrac{\neg A \qquad A \qquad \begin{matrix}[\bot]^i \\ \vdots \\ C\end{matrix}}{C}\ \neg\text{-E}_{GE}, i$$

The standard rule is derivable by setting C equal to $\bot$. Likewise, it is easy to check that ¬-I and ¬-E are fully invertible, and hence harmonious in the sense defined in § 5.4.3.

Out there in the cold

Given ¬-I, the classical rule of Double Negation Elimination is left, so to speak, in the cold. Yet without it, or without some other classical rule, there is no way one could get classical logic. Consider the following derivation, where an application of negation introduction is immediately followed by one step of double negation elimination:

$$\text{DN}\ \dfrac{\neg\text{-I}, i\ \dfrac{\begin{matrix}[\neg A]^i \\ \vdots \\ \bot\end{matrix}}{\neg\neg A}}{A}\ .$$

There is no way we can get from $\neg A$ to A, from the same or fewer assumptions, without applying double negation elimination. Hence, the pair {¬-I, DN} fails to satisfy intrinsic harmony, and, *a fortiori*, strong intrinsic harmony. The pair also fails to satisfy GE harmony, since the GE schema induces ¬-E, not DN, as the harmonious E-rule for ¬-I. Finally, on our account of harmony as full invertibility, ¬-I and ¬-E respectively induce each other, but there is no room for DN. As Dummett puts it:

> Plainly, the classical rule is not in harmony with the introduction rule. (Dummett, 1991b, p. 291)

Prawitz also writes, in a similar spirit, that

> clearly [...] we know procedures that make all intuitionistic inference rules acceptable [...] but not any procedure that makes the rule of double negation elimination acceptable. (Prawitz, 1977, p. 39)

Two pairs of rules?

It might be objected that the problem can be solved by postulating that classical negation is governed by *four* pairs of pairwise harmonious rules: the standard intuitionistic rules of ¬-I and ¬-E, and the obviously harmonious rules of double negation introduction

$$\text{DN-I}\ \frac{A}{\neg\neg A}$$

and double negation elimination. There are at least two problems with this suggestion, however.

To begin with, one might wonder why the difference between intuitionistic and classical negation should only emerge when we are dealing with successive occurrences of the negation operator. If meaning is compositional, a difference in the meaning of an expression E should be already manifest in *single* occurrences of E. Tennant voices the concern:

> Why should we deal with two occurrences simultaneously? [...] Surely, the intuitionist maintains, whatever disagreement there may be about the very meaning of negation should be able to be brought into the open in the context of differing (schematizable) logical practice with regard to single occurrences of the logical operator concerned? (Tennant, 1997, p. 310)

Second, and more importantly, if the meaning of a logical constant can be determined by more than one set of I- and E-rules, then harmony alone does not suffice to rule out `tonk`-like connectives, as shown by the rules for ⊕ we introduced in § 5.4.3.

5.7.3 Classical *reductio*

The classical logician may perhaps insist that the classical rule of *classical reductio* can be justified on purely proof-theoretic grounds. Recall, the rule allows us to derive A if the supposition that $\neg A$ leads us to absurdity. In symbols:

$$\begin{array}{c} [\neg A]^i \\ \vdots \\ \text{CR}, i\ \dfrac{\bot}{A} \end{array}.$$

One may interpret CR as a rule for introducing A, that is, on the foregoing assumptions, as a rule constitutive of the meaning of A. Thus, Milne (1994, p. 58) suggests that CR be treated as a I-rule for A, and points out that CR is intrinsically harmonious with $\neg$-E, "but now understood with A, not $\neg A$, as the major premise". He argues that the "justification of the elimination rule is straightforward", as shown by the following reduction procedure:

$$\neg\text{-E}^8\,\dfrac{\text{CR}, i\,\dfrac{\begin{matrix}\Gamma_0, [\neg A]^i\\ \Pi_0\\ \bot\end{matrix}}{A} \qquad \begin{matrix}\Gamma_1\\ \Pi_1\\ A\end{matrix}}{\bot} \quad \rightsquigarrow_r \quad \begin{matrix}\Gamma_1\\ \Pi_1\\ \underbrace{\Gamma_0, \neg A}\\ \Pi_0\\ \bot\end{matrix}$$

where $\neg$-E* denotes the foregoing unorthodox interpretation of the standard rule of negation elimination. Indeed, Milne's rules are not only locally sound. They are also locally complete, as shown by the following expansion:

$$\begin{matrix}\Pi\\ A\end{matrix} \quad \rightsquigarrow_e \quad \text{CR}, i\,\dfrac{\neg\text{-E}^*\,\dfrac{\begin{matrix}\Pi\\ A\end{matrix} \qquad [\neg A]^i}{\bot}}{A}$$

Thus, Milne's rules satisfy strong intrinsic harmony. Is not this enough to meet the inferentialist's standards?

The main problem with the foregoing suggestion is that CR would seem to break with compositionality. If we interpret it as an I-rule for introducing atomic statements, it makes the meaning of A dependent on the meaning of a more complex expression, $\neg A$, whose meaning in turn depends on the meaning of A. Prawitz writes:

> this explanation of the meaning of atomic sentences seems incredible. [...] It breaks with the molecular and inductive character of intuitionistic (and Platonistic) theories of meaning since the canonical proofs of an atomic sentence A now depend on the canonical proofs of $\neg A$, which again depend on the canonical proofs of A. (Prawitz, 1977, p. 36)

Notice that Prawitz's point does not depend on the adoption of a *proof-conditional* conception of meaning and understanding. Insofar as the meaning of a complex statement depends on the meanings of its component expressions but not *vice versa* (in particular, insofar as the meaning of $\neg A$ depends on the meaning of A but not *vice versa*), the point would equally go through on the assumption that A's meaning is given by its *truth*-conditions.

One might start feeling the force of the conclusion Dummett invites us to draw:

> it almost seems that there is no way of attaining an understanding of the classical negation-operator if one does not have one already. That is a strong ground for suspicion that the supposed understanding is spurious. (Dummett, 1991b, p. 299)

On the other hand, Dummett writes,

> the meaning of the intuitionistic logical constants can be explained in a very direct way, without any apparatus of semantic theory, in terms of the use made of them in [our deductive] practice. (*Ibid.*)

To be sure, Dummett has not shown that classical negation is disharmonious on *every* admissible formalization of classical logic. Still, it is undeniable that, in standard natural deduction settings, classical negation is proof-theoretically suspicious. For one thing, there is an asymmetry between the classical law of double negation elimination and the intuitionistic rule of negation elimination: the latter, but not the former, can be justified on the assumption that the rule of $\neg$-I specifies the necessary and sufficient conditions for asserting negated statements. For another, CR cannot be treated as a meaning-constitutive without entering, as Prawitz observes, in an explanatory circle.

5.8 Conclusions

We have defined at least *three* viable notions of harmony: strong intrinsic harmony, GE harmony, and harmony as full invertibility. Strong intrinsic harmony presupposes that we can, so to speak, run through the various possible I- and E-rules for a certain logical operator $, and choose the ones that satisfy strong intrinsic harmony: we are not given a procedure for generating E-rules from arbitrary I-rules, and *vice versa*. GE harmony gives such a procedure, but, in general, it allows for some 'leakage' of information: harmonious GE rules do not in general give us back the canonical grounds specified by the corresponding I-rules. By contrast, harmony as full invertibility delivers a procedure for deriving harmonious E-rules from their corresponding I-rules, and *vice versa*. We also observed, though, that harmony as full invertibility invalidates the standard rules for disjunction and the existential quantifier, which means that the account requires the adoption of alternative rules for $\vee$ and $\exists$. On the other hand, as we have just seen, on standard natural deduction frameworks, each of the three accounts of harmony we have introduced sanctions the classical rule of Double Negation Elimination as inharmonious. The rule is not justified by the standard rule of negation introduction. How can the classical logician react? As far as I can see, there are four main options.

The *first* is to object that *intuitionistic logic* is itself already defective. We will consider one possible argument to this effect in § 6.5 below. The *second* is to reject the requirement of harmony, together with its main justifications. This is of course a coherent option. However, I will set it aside for present purposes, in keeping with the general strategy of the this thesis, which is to grant the revisionist her main metaphysical and semantic assumptions. What is more, it is worth reminding that rejecting harmony would require rejecting the Fundamental Assumption, since the latter entails the former. Rejecting the Fundamental Assumption is not as easy as it might seem, however. One would have to find a complex statement A such that (i) we can be in a position to assert A without it being possible, even in principle, to introduce A by means of *some* I-rule for its main logical operator. (Notice that quantification over possible I-rules here is here forced upon us, since the Fundamental Assumption, by itself, does not mention any specific I-rule for any of the logical operators.) The *third* possibility is to try to formalize classical logic by means of proof-theoretically acceptable rules, within a standard natural deduction framework. There are a few such 'conservative' proposals in the literature—some of which we have already considered (see e.g. Weir, 1986), and some of which we will not consider, for reasons of space (see e.g. Milne, 2002).[18] I will present my own conservative formalization of classical logic in § 7.4 below. Finally, the *fourth* option is to provide an alternative regimentation of logic within an altogether different logical framework—one on which the rules for classical negation are after all harmonious. We shall consider two such alternative frameworks in Chapter 7.

[18]But see Steinberger (2009a) for an excellent assessment of Milne's 2002 proposal.

Chapter 6

Inferentialism and separability

All parties agree that, on any decent account of harmony, the standard textbook formalizations of classical logic are not harmonious. But what if the problem lies with the textbook formalizations, and not with classical logic itself? Since classical logic can be formalized in many ways, the argument we have presented in the previous chapter does not show that one cannot provide a harmonious formalization of classical logic. It only shows that the *existing ones* are defective. The aim of this chapter is to introduce more proof-theoretic requirements—requirements that, in some cases, have been claimed to be be necessarily incompatible with *classical logic*—not just with its standard regimentations. In § 4.4.2, we drew a distinction between *local* proof-theoretic requirements and *global* ones. The former only apply to rules independently of the systems to which they may belong. The latter concern relations between rules and logical systems. It is now time to explore the relations between local and global constraints, as well as their possible motivations, and their alleged revisionary consequences. Our main focus will be on the twin global requirements of *separability* and *conservativeness*—respectively, and very roughly, that every provable rule or theorem must be provable by means of the rules for the logical operators occurring in it; and that, for any logical operator \$, the addition of the \$-rules to any well-behaved logical system S may not affect the logical relations among the expressions of S's language. It can be shown that, given minimal assumptions, classical logic does not satisfy either constraint. The totality of the correct uses of classical disjunction and classical implication can only be derived if the rules for classical negation are also present in the system. Their meanings, it would seem, are not fully determined by their I- and E-rules alone.

Classical logicians may object that the requirements of separability and conservativeness are too strong. They amount to assuming that, for any logical operator

\$, the \$-fragment be complete with respect to \$'s correct uses. However, this assumption fails for *higher-order* logics, i.e. logics in which quantifiers range not only over individuals, but also over properties or sets (and properties of properties, sets of sets, etc.). As a result, inferentialists who are willing to countenance higher-order logics appear to be hoist with their own petard: they find themselves in the uncomfortable position of dismissing classical logic on grounds that would also compel them to dismiss higher-order logics. What is more, classical logicians may insist that inference rules are not *categorical*, in the sense that they do not even determine meanings in the sense of determining their satisfaction clauses—a problem first noticed by Rudolf Carnap (1943), and recently revived by Panu Raatikainen (2008). I argue, however, that a careful examination of the arguments for imposing separability and conservativeness reveals that these requirements should not hold across the board—in particular, they should not hold for higher-order logics, although they may well do for first-order logics. Moreover, as I show, basic rules *do* determine their satisfaction clauses, both in an intuitionist framework, as we shall see at the end of this chapter (§ 6.5), and in a classical one, as we shall see in the next chapter (§ 7.4.4).

Our plan is as follows. Section 6.1 considers some orthodox and less orthodox views about what meaning-constitutive rules should look like, and briefly discusses their possible motivations, as well as some of their revisionary consequences. Section 6.2 introduces the requirements of conservativeness and separability, and closely examines the standard arguments for imposing them, as given by Michael Dummett, Kent Bendall, and Peter Milne. Section 6.3 discusses what I shall call the *Argument from Separability*: an argument to the effect that classical logic is necessarily proof-theoretically defective. I will defend two main claims: first, that the existing arguments for separability beg the question against the classical inferentialist, in that they presuppose the falsity of a broadly *holistic* account of our understanding of the logical vocabulary; and second, that, all the same, classical inferentialists had better be able to give a non-holistic, i.e. separable, formalization of classical logic. Sections 6.4-5 consider and address two possible classicist replies, purporting to show that the Determination Thesis—that basic rules determine meanings—is untenable. Section 6.6 offers some concluding remarks.

6.1 Meaning-constitutive rules

The requirement of harmony places no restriction on the form of the meaning-constitutive rules. Yet, it easy to see that, just as there are pathological combinations of I- and E-rules, as Arthur Prior first showed, there are also I- (E-rules) that are pathological independently of whether they are harmonious with their corresponding E-rules (I-rules). Consider Stephen Read's zero place connective bullet (see Read, 2000):

Example 6. *Bullet:*

$$
{\scriptstyle \bullet\text{-}I,\, i}\ \dfrac{\begin{array}{c}[\bullet]^i\\ \vdots\\ \bot\end{array}}{\bullet}
\qquad\qquad
{\scriptstyle \bullet\text{-}E}\ \dfrac{\bullet \qquad \bullet}{\bot}\,.
$$

Unlike the rules for `tonk`, the rules for bullet are harmonious. They satisfy strong intrinsic harmony, as the following conversions show:

$$
{\scriptstyle \bullet\text{-E}}\ \dfrac{{\scriptstyle \bullet\text{-I},\, n}\ \dfrac{\begin{array}{c}[\bullet]^n\\ \Pi_1\\ \bot\end{array}}{\bullet} \qquad \begin{array}{c}\Pi_2\\ \bullet\end{array}}{\bot}
\quad \rightsquigarrow_r \quad
\begin{array}{c}\Pi_2\\ \bullet\\ \Pi_1\\ \bot\end{array}
\qquad\qquad
\begin{array}{c}\Pi\\ \bullet\end{array}
\quad \rightsquigarrow_e \quad
{\scriptstyle \bullet\text{-I},\, n}\ \dfrac{{\scriptstyle \bullet\text{-E}}\ \dfrac{\begin{array}{c}\Pi\\ \bullet\end{array} \qquad [\bullet]^n}{\bot}}{\bullet}\,.
$$

They satisfy GE harmony, as one can easily verify by applying the GE schema to •-I (the standard rule of •-E is obtained by setting C equal to $\bot$):

$$
{\scriptstyle \bullet\text{-E}_{GE},\, i}\ \dfrac{\bullet \qquad \bullet \qquad \begin{array}{c}[\bot]^i\\ \vdots\\ C\end{array}}{C}\,.
$$

Finally, they also satisfy full invertibility: one can derive from •-E precisely what was required for introducing bullet in the first place, viz. a derivation of $\bot$ from bullet.

However, •-I is already inconsistent: it tells us that $\bot$ may be asserted if and only if its assertion leads us to absurdity, i.e. if and only if it may *not* be asserted. As a result, bullet unsurprisingly yields a proof of absurdity:

$$
{\scriptstyle \bullet\text{-E}}\ \dfrac{{\scriptstyle \bullet\text{-I},\,1}\ \dfrac{{\scriptstyle \bullet\text{-E}}\ \dfrac{[\bullet]^1 \qquad [\bullet]^1}{\bot}}{\bullet} \qquad {\scriptstyle \bullet\text{-I},\,2}\ \dfrac{{\scriptstyle \bullet\text{-E}}\ \dfrac{[\bullet]^2 \qquad [\bullet]^2}{\bot}}{\bullet}}{\bot}
$$

Harmony does not guarantee consistency. If meaning-constitutive rules are to be consistent, this suggests that either the local proof-theoretic constraints need to be strengthened, perhaps by imposing some requirements on the form of the acceptable I-rules, or inferentialists need to resort to *global* constraints. We will consider the first option first.

6.1.1 Atomism, molecularism, and holism

Following Dummett (1991b, pp. 256-7), let us say that a constant $ *figures* in a rule *R* if it occurs in its schematic representation. We may then distinguish three different inferentialist approaches to the meaning of the logical constants:

- *Logical atomism*. The meanings of the logical constants are independent of each other. For each logical constant $, one can formulate I- and E-rules for $ such that (i) $ is the only logical expression figuring in their schematic representation, and (ii) $-I and $-E, or some subset thereof, fully determine $'s meaning.

- *Logical molecularism*. For each logical constant $, the I- and E-rules for $, or some subset thereof, fully determine $'s meaning. There may be relations of meaning-dependence among logical constants, although there are at least two constants whose meanings are independent of one another.

- *Logical holism*. The meanings of the logical constants are *all* interdependent: for each constant $, the meaning of $ is determined by the totality of the rules of the system to which $ belongs.

These different approaches warrant a couple of remarks.

First, logical molecularism may come in at least two flavours: a weak and a strong one. *Weak molecularism* allows for local forms of holism: there are at least two constants $ and ⋆ whose meanings are fully determined by the union of their meaning-constitutive rules, irrespective of whether ⋆ figures in the rules for $ and *vice versa*. By contrast, *strong molcularism* only allows for *asymmetric* relations of meaning-dependence: $'s meaning can depend on ⋆'s meaning only provided that $ does not itself figure in the ⋆ rules.

Second, although both logical atomism and logical molecularism are incompatible with *global holism*, the view that the meanings of all the expressions of language are interdependent, logical holism is compatible with global molecularism, the view that meanings are in general independent of each other, although some expressions, e.g. colour terms and, perhaps, the logical constants, form packages that can only be acquired *en bloc*.

6.1.2 Purity, simplicity, single-endedness, and sheerness

Let us now introduce some definitions concerning *inference rules*. Following Dummett (1991b, pp. 256-7), we shall call a rule *single-ended* "if it is either an introduction rule but not an elimination rule, or an elimination rule but not an introduction rule". The following rule, for instance, is not single-ended, in that it is both an elimination rule for negation and an introduction rule for disjunction:

$$\frac{\neg(\neg A \wedge \neg B)}{A \vee B}.$$

We shall call a rule *pure* if only one logical constant figures in it. Thus, the rule of Double Negation Elimination

$$\frac{\neg\neg A}{A}\,\text{DN}$$

is pure, but the standard intuitionist rules for negation

$$\neg\text{-I}, i\,\frac{\begin{matrix}[A]^i \\ \vdots \\ \bot\end{matrix}}{\neg A} \qquad \neg\text{-E}\,\frac{A \quad \neg A}{\bot}$$

and the rule of classical *reductio*

$$\text{CR}, i\,\frac{\begin{matrix}[\neg A]^i \\ \vdots \\ \bot\end{matrix}}{A}$$

are impure—at least if absurdity is taken to be a propositional constant.[1] We shall say that a rule is *simple* if any constant that figures in it only occurs as the main logical operator of the sentence in which it figures (see also Milne, 2002, p. 507). So, for instance, intuitionistic negation is simple, but classical *reductio* and Double Negation Elimination are not. Finally, we shall also say that a rule is *sheer* if either it is an I-rule, but the constant it introduces does not figure in its assumptions and in its hypotheses for discharge, or it is an E-rule, but the constant it eliminates does not figure in its conclusions or in its hypotheses for discharge. Thus, the standard rules of $\wedge$-I and $\wedge$-E are both sheer, but the following rule of negation introduction is not:

[1]This is not a trivial assumption, as we shall see in § 6.5.5 below.

$$\dfrac{\begin{matrix}[A]^i \\ \vdots \\ B\end{matrix} \qquad \begin{matrix}[A]^i \\ \vdots \\ \neg B\end{matrix}}{\neg A}\ \neg\text{-I}^*, i$$

Here $\neg$ occurs both in the conclusion and in one of the premises of the rule.

6.1.3 Rule-purism

Now to the key question: what is, in general, the form of an admissible inference rule? Everybody agrees that a rule R is an introduction rule for a constant $ only if its conclusion has $ as its principal operator. Likewise, everybody agrees that a rule R is an elimination rule for $ only if its main premise is required to have $ as its principal operator. However, inferentialists typically do not take either of these necessary conditions also to express a *sufficient* condition for being, respectively, admissible I- and E-rules (see e.g. Dummett, 1991b, p. 256). So which rules can be meaning-constitutive, and why? Let us first introduce one final definition (Prawitz, 1965, p. 16).

Definition 6. *(Subformula) The notion of a subformula is inductively defined by the following clauses:*

(1) A is a subformula of A;

(2) A is a subformula of $\neg A$;

(3) If $B \wedge C$, $B \vee C$, or $B \rightarrow C$ is a subformula of A, then so are B and C;

(4) If $\forall xFx$ or $\exists xFx$ is a subformula of A, then so is $F[x/t]$, for all t free for x in F.

We can now introduce Dag Prawitz's definition of an I-rule:

> An introduction rule for a logical constant $ allows the inference *to* a formula A that has $ as principal sign from formulas that are subformulas of A. (Prawitz, 1965, p. 32)

On Prawitz's view, I-rules give necessary and sufficient conditions for introducing complex formulae that are "stated in terms of subformulas of these formulas" (*Ibid.*). It follows from Prawitz's definition that, in the terminology we have introduced in § 6.1.2, admissible I-rules must be pure, sheer, simple, and single-ended. Let us dub the view that admissible I-rules should satisfy these requirements and admissible E-rules should be in harmony with the corresponding I-rules *rule-purism*.

Definition 7. *(Rule-purism) A pair of I- and E-rules for a logical operator \$ is admissible if and only if (i) \$-I and \$-E are harmonious, and (ii) \$-I satisfies purity, sheerness, simplicity, and single-endedness.*

Contemporary rule-purists

Rule-purists typically subscribe to *logical atomism*, the view that the meanings of the logical constants are all independent of one another. Here is Neil Tennant:

> One [should] be able to master various fragments of the language in isolation, or one at a time. It should not matter in what order one learns (acquires grasp of) the logical operators. It should not matter if indeed some operators are not yet within one's grasp. All that matters is that one's grasp of any operator should be total simply on the basis of schematic rules governing inferences involving it. (Tennant, 1997, p. 315)
>
> the analytic project must take the operators one-by-one. The basic rules that determine logical competence must specify the unique contribution that each operator can make to the meanings of complex sentences in which it occurs [...]. This is the requirement of separability. (Tennant, 1997, p. 315)

Heinrich Wansing declares, in a similar vein:

> if one wants to avoid a (partially) holistic account of the meaning of the logical operations, the meaning assignment should not make the meaning of an operation f dependent on the meaning of other connectives. The [...] rules for f should give a purely structural account of f's meaning in the sense that they should not exhibit any connective other than f. This property may be called *separation*. (Wansing, 2000, p. 10)

A terminological quibble. Wansing's and Tennant's terminological choices are somewhat infelicitous. What they mean by 'separability', or 'separation', is clearly Dummett's notion of purity. In the standard usage, 'separability' rather refers to a global property of logical systems—one that will be introduced and discussed in detail in § 6.2 below. Tennant further requires that meaning-constitutive rules also be, in Dummett's terminology, *simple*, *sheer*, and *single-ended*:

> introduction and elimination rules are, and should be, formulated in such a way that the only occurrence of a logical operator mentioned in them is precisely the dominant occurrence within the conclusion of the introduction rule or the dominant occurrence within the major premise of the elimination rule. (Tennant, 1997, p. 315)

Two questions present themselves. What are the consequences of rule-purism? And how, if at all, can it be motivated? Let us begin with the first question.

Rule-purism, bullet, and classical logic

It is easy to see that Read's $\bullet$ is unacceptable by the rule-purist's standards. Its I-rule is not sheer: $\bullet$ occurs both as the conclusion and as the discharged hypothesis of $\bullet$-I. However, given standard assumptions on how logic is to be formalized, it is not difficult to see that classical logic is itself at odds with rule-purism. The law of Double Negation Elimination is pure, but not simple, since negation occurs other than as the main logical operator of the sentence in which it figures. As for the classical laws of classical *reductio*, the Law of Excluded Middle, and Classical Dilemma, Tennant correctly points out that

> each of them falls foul of one of our requirements so far. Classical *reductio*, though it confines itself to a single occurrence of the negation operator, and is schematic elsewhere, nevertheless has the occurrence in the wrong place: neither in the conclusion, nor in a major premiss [...]. The rule of dilemma is objectionable for the same reason. Finally, the law of excluded middle sins by joining negation and disjunction inseparably. Such a marriage is bound to be unstable, given that each of them is going to have to consort separately with other operators in order to produce valid arguments. (Tennant, 1997, p. 317)

If meaning-constitutive rules must be pure, simple, sheer, and single-ended, then neither CR nor Dilemma can be meaning-constitutive, since they are neither sheer nor single-ended. Similarly, LEM is also objectionable because it is impure: it "forces [negation and disjunction] into a shoddy marriage of convenience" (Tennant, 1997, p. 317).

Order does not matter

But what is the inferentialist's *justification* for requiring that proof-theoretically acceptable formalizations of logic only contain I-rules that are pure, simple, and single-ended? Why should inferentialists want to avoid a partially holistic account of the meaning of the logical operations, as Wansing puts it? Neither Prawitz nor Wansing and Tennant offer much by way of argument. Prawitz and Wansing do not even attempt to motivate their requirement. Tennant first limits himself to saying that "the rules have to be thus focused [...] otherwise they are not isolating sufficiently the logical operator whose meaning is in question" (Tennant, 1997, p. 315). But, of course, this is to *state* the requirement of purity, not to provide an argument for it. He then gestures at an argument for purity from the acquisition of logical concepts. In Tennant's words:

> It follows from separability that one would be able to master various fragments of the language in isolation, or one at a time. It should not matter in what order one learns (acquires grasp of) the logical operators. It should not matter if indeed some operators are not yet within one's grasp. All that matters is that one's grasp of any operator should be total simply on the basis of schematic rules governing inferences involving it. (Tennant, 1997, p. 315)

The idea is that *order does not matter*: it does not matter whether, say, one learns disjunction or conjunction first; both possibilities should be left open. One wonders, though, why should that be. To say that order does not matter is merely to say that the meanings of the logical constants are independent of one another: precisely what was meant to be shown! We will consider something close to Tennant's reasoning in due course. For the time being, let us consider whether the inferentialist's purism can be somehow relaxed.

6.1.4 Dummett on I- and E-rules

In *The Logical Basis of Metaphysics*, Dummett too expresses sympathy for the view that the meanings of the logical expressions are not interdependent. He writes:

> to understand $\ulcorner A \vee B \urcorner$, one need not understand $\ulcorner A \wedge B \urcorner$ or $\ulcorner A \rightarrow B \urcorner$. (Dummett, 1991b, p. 223)

Dummett's point is that meanings of the logical constants are independent of one another: in his view, it is just plain that, say, one can understand negation without understanding disjunction, and *vice versa*. But again, Dummett presents us with an *intuition* that the meanings of the logical constants can be learned independently: he does not offer an *argument* for the view.

Dummett's liberalism

Quite surprisingly, some 35 pages later after having said that the logical constants do not satisfy the generality constraint, Dummett dismisses the idea that, following Gentzen, logical inferentialists should restrict their rules "to those that are pure, simple, and single-ended". He first attacks the requirement of purity:

> An impure \$-introduction rule will make the understanding of \$ depend on the prior understanding of the other logical constants figuring in the rule. Certainly we do not want such a relation of dependence to be cyclic; but there would be nothing in principle objectionable if we could so order the logical constants that the understanding of each depended only on the the understanding of those preceding it in the ordering. (Dummett, 1991b, p. 257)

He then criticizes the demand for simplicity:

> Given such an ordering, we could not demand that each rule be simple, either. The introduction rules for \$ might individually provide for the derivation of sentences of different forms with \$ as principal operator, according to the other logical constants occurring in them: together they would provide for the derivation of *any* sentence with \$ as principal operator. (*Ibid.*)

Dummett argues that order *may* matter: it is at least conceivable that there be relations of meaning-dependence among logical constants, provided that they be well-grounded. Furthermore, Dummett plausibly suggests that the requirement of simplicity may be excessive too: there may be more than one set of I-rules for each constant \$, and some of these I-rules may specify how to introduce \$ as principal operator in sentences in which some other constant $\star$ figure. For instance, one may give additional I-rules for asserting disjunctions of the form $A \vee \neg B$, and claim that the meaning of $\vee$ is fully determined by the $\vee$- *and* the $\neg$-rules. In the terminology we introduced in § 6.1.1, Dummett dismisses logical atomism, and concedes that what we have called strong molecularism may be a viable option.

The complexity condition

Even more surprisingly, Dummett abandons the suggestion that logical constants should be ordered by a relation of meaning-dependence, on the grounds that "the principle of compositionality in no way demands this" and that "all that is essentially presupposed for the understanding of a complex sentence is the understanding of the subsentences" (Dummett, 1991b, p. 258). Accordingly, Dummett offers the following, final characterization of an I-rule aiming at being meaning-constitutive:

> The minimal demand we should make on an introduction rule intended to be self-justifying is that its form be such as to guarantee that, in any application of it, the conclusion be of higher logical complexity than any of the premisses and than any discharged hypothesis. We may call this the 'complexity condition'. In practice, it is evident that there will be no loss of generality if we require the rule to be single-ended, since, for a premiss with the same principal operator as the conclusion, we may substitute the hypotheses from which that premiss could be derived by the relevant introduction rule. We may accordingly recognize as an introduction rule a single-ended rule satisfying the complexity condition. (Dummett, 1991b, p. 258)

Dummett's definitive view about I-rules essentially amounts to demanding that they be *only* I-rules, and that they satisfy the complexity condition: *in each of their applications*, their conclusion must be of higher complexity than any of the premises and than any discharged hypothesis.

Definition 8. *(Complexity condition)*
An I-rule R satisfies the complexity condition *if,* in any application *of R, the conclusion is of higher logical complexity than any of the premisses and than any discharged hypothesis.*

Given the complexity condition, $\bullet$-I is clearly inadmissible: its conclusion, $\bullet$, is no more complex than $\perp$ and $\bullet$ itself.

Compositionality

It is less clear, though, what the inferentialist's reasons for imposing the complexity condition can be. Dummett takes his condition to be motivated by a principle of compositionality for understanding, that "all that is essentially presupposed for the understanding of a complex sentence is the understanding of the subsentences". However, consider the following pair of harmonious rules for $\rightarrow$:

Example 7. *Material implication:*

$$\rightarrow\text{-I}^*,i\ \dfrac{\begin{matrix}[A, \neg B]^i \\ \vdots \\ \bot\end{matrix}}{A \rightarrow B} \qquad \rightarrow\text{-E}^*\ \dfrac{A \rightarrow B \quad A \quad \neg B}{\bot}.$$

Just like •-I, →-I* violates the complexity condition. But now, suppose someone took *these* rules to be constitutive of the meaning of →. If Dummett's reasoning were correct, this assumption should be inconsistent with the principle of compositionality. It is hard to see why it should be so, however. On the foregoing assumptions, one's understanding of $A \rightarrow B$ would depend, by compositionality, on one's understanding of A, →, and B. In turn, one's understanding of → would depend on the logical operators figuring in the modified →-rules, viz. ¬ and ⊥. Yet, the converse relation does not hold, since → does not itself occur in the rules for ¬ and ⊥. In Dummett's own words, it follows that the "relation of dependence between expressions and sentence-forms" is "asymmetric". But this is all, according to Dummett himself, "the principle of compositionality essentially requires" (Dummett, 1991b, p. 223).

The complexity condition*

There is worse, however. Dummett takes his condition to apply to all the possible *instances* of a rule. But this demand seems exorbitant: if our language contains terms operators with free variables, for instance, then some applications of

$$\forall\text{-I}\ \frac{F[t/x]}{\forall x F(x)}$$

will violate Dummett's condition. At most, it would seem, one could argue that the *schematic* formulation of a rule be such that its premises and dischargeable hypotheses be subformulae of its conclusions. So what is an admissible I-rule?

Consider again •-I. This rule is aimed at giving the meaning of • by laying down its assertibility-conditions, viz. that, if one can derive absurdity from •, one can discharge • and assert •. This definition is clearly circular, however. For "if one can derive absurdity from •, then one discharge • and assert •" effectively means that, if every proof of • can be transformed into a proof of ⊥, then one can assert •. That is, our supposed definition of what counts as a canonical proof of • *already presupposes that we already know what counts as a proof of* •. A minimal requirement on the admissibility of I-rules, then, is that they be non-circular: knowledge of the premises of a \$-I rule, i.e. knowledge of the canonical grounds of \$-statements, may not presuppose that we already know how to introduce \$-statements. Thus, Read's • is defective precisely for the same reasons as *classical reductio*, viewed as an I-rule for atoms, is.

Inferentialists may ensure this non-circularity requirement by weakening Dummett's original formulation of the complexity condition in a natural way. Instead of requiring that *all the possible applications* of an admissible I-rule satisfy the complexity condition, they may require that *the schematic form* of an admissible I-rule should be such that its conclusion is of higher complexity than any of the premises and than any discharged hypothesis. Call this the *complexity condition**.

Definition 9. *(Complexity condition*)*
An I-rule R satisfies the complexity condition* *if,* the schematic formulation *of R is such that the conclusion is of higher logical complexity than any of the premisses and than any dischargeable hypothesis.*

This weakened, but seemingly more reasonable, requirement is still strong enough to dismiss •-I, as it should be. On the other hand, as we shall see in § 6.4.2, it allows us to keep the rules for the higher-order quantifiers.

The difficult question, to be sure, is whether the complexity condition* can be plausibly motivated. But I will leave the answer to this question, if it can be given at all, for another occasion.

6.2 Global proof-theoretic constraints

If the foregoing considerations are correct, the inferentialist arguments for rule-purism and strong molecularism are defective. The same can be said of Dummett's argument from compositionality for his suggested account of an admissible I-rule. Does this mean that the intuitionist's reservations (see *supra*, § 6.1.3) about such rules as LEM and Dilemma may after all be unmotivated? Not necessarily. To begin with, LEM and Dilemma are neither I- nor E-rules. Hence, they are not justified by the harmony considerations we introduced in § 5.1. Secondly, as we shall see, both LEM and Dilemma are, at least *prima facie*, incompatible with the *global* inferentialist requirements of *separability* and *conservativeness*—requirements, however, that are often thought to be integral to the inferentialist approach to logic. Inferentialists take meaning-constitutive rules to determine correct use—this is the first of the two interpretations of the Determination Thesis we made explicit in § 4.1.2. As we have seen in § 5.1, they typically assume that I-rules are complete in the weak sense specified by the Fundamental Assumption, viz. that the complex statements we are in a position to assert can in principle always be introduced by means of an argument whose last step is taken into accordance with one of the I-rules for their main logical operator. However, as we have already anticipated, sometimes inferentialists make an even stronger assumption, to the effect that, for any logical operator \$, \$'s meaning-constitutive rules allow us to derive *all* of \$'s inferential uses. This thought leads to the twin global requirements of conservativeness and separability, and to what I shall call *orthodox inferentialism*. Let us have a closer look.

6.2.1 Separability and conservativeness

First, some formal preliminaries. For ease of exposition, we shall work within a single-conclusion natural deduction calculus in *sequent* style. Sequent calculi systems were first introduced by Gerhard Gentzen in his doctoral dissertation.[2] They combine features of both axiomatic and natural deduction systems. Like axiomatic systems, they have axioms and they are not assumption-based. Unlike axiomatic systems, they have both introduction and elimination rules. More specifically, proofs are manipulations of *sequents* of the form $\Gamma \vdash A$, where Γ is a finite, possibly empty multiset of formulae (where a multiset is an aggregate that is insensitive to order, like sets, but is sensitive to repetitions, like lists). A *sequent* is an ordered pair whose first element, the *antecedent*, is a list of either formulae or multisets and whose second member, the *succedent*, is a formula.[3] The antecedent of a sequent lists the assumptions on which the formula in the succedent depends. Intuitively, a sequent of the form $\Gamma, \Delta \vdash A$ says that if everything in Γ and Δ is true, then A is also true. Γ and Δ are usually called *contexts*, A, B, etc. are known as *active formulae*.

Resorting to a natural deduction calculus in sequent style allows us to make a more perspicuous distinction between the *structural rules*, i.e. rules in which no logical operator figures, and the *operational rules*, i.e. rules governing the use of the logical operators—a distinction that is otherwise somewhat blurred in standard natural deduction systems. Thus, the system allows for a perspicuous formulation of the rules of weakening and contraction, as follows:

$$\text{Weakening}\ \frac{\Gamma \vdash C}{\Gamma, A \vdash C} \qquad \text{Contraction}\ \frac{\Gamma, A, A \vdash C}{\Gamma, A \vdash C}\,.$$

The operational rules are just like the standard natural deduction ones, except that they are written in a sequent format, with contexts made explicit. For instance, the rules for conjunction are as follows:

[2] See Gentzen (1934). See also §§ 7.4.1-2 below.

[3] We shall consider *multiple-conclusions* sequent calculi in § 7.1.1 below.

$$\wedge\text{-I}\ \frac{\Gamma \vdash A \quad \Gamma \vdash B}{\Gamma \vdash A \wedge B} \qquad \wedge\text{-E}\ \frac{\Gamma \vdash A \wedge B}{\Gamma \vdash A} \quad \frac{\Gamma \vdash A \wedge B}{\Gamma \vdash B}.$$

We shall then say that a sequent $X_1, \ldots, X_n \vdash Y$ *holds* on an interpretation $\mathcal{M}$ if it is not the case that each member of $X_1, \ldots, X_n$ is true on $\mathcal{M}$ and Y is not true on $\mathcal{M}$, and that it *fails* otherwise. We shall also say that a sequent is *valid* if it holds in every interpretation, and that a rule is valid if, for all interpretations, its second member holds only if its first member does. We shall call a system S *complete* if and only if every valid sequent is derivable in S. Conversely, we shall call S *sound* if and only if every derivable sequent is valid. We shall then say that S is *rule complete* if and only if every valid rule is derivable, and that S is *rule sound* if and only of every derivable rule is valid. Rule completeness (soundness) entails plain completeness (soundness).

Separability and conservativeness: weak and strong

Following (Bendall, 1978, p. 250), we can now define the core notions of weak and strong *separability* as follows:

Definition 10. *(Weak separability) A system S is* weakly separable *if every provable sequent $X_1, \ldots, X_n \vdash Y$ in the system has a proof that only involves either structural rules or rules for the logical operators that figure in $X_1, \ldots, X_n \vdash Y$.*

Definition 11. *(Strong separability) A system S is* strongly separable *if every provable* rule *R is provable in S by means of either structural rules or rules for the logical operators that figure in R.*

It follows from the foregoing definitions that any calculus that is both complete and strongly separable will also be *locally complete*: for each logical operator \$, the rules for \$ will be strong enough to prove all the provable \$-rules—and, since rule completeness entails completeness *tout court*, all the provable \$-sequents.

Now to the requirement of *conservativeness*. In his reply to Prior's 1960 article, Nuel Belnap (1962) famously suggested that admissible logical constants should yield a *conservative extension* of the systems to which it may be added. Weak and strong conservativeness may respectively be defined as follows:

Definition 12. *(Weak conservativeness) Let S and S' be two systems with language L and L' respectively, where $S \subseteq S'$ and $L \subseteq L'$. Then, S' is* weakly conservative *over S if, for all $A \in L$ and for all sets Γ of L-sentences, $\Gamma \vdash_{S'} A$ only if $\Gamma \vdash_S A$.*

Definition 13. *(Strong conservativeness) Let S and S' be two systems with language L and L' respectively, where $S \subseteq S'$ and $L \subseteq L'$. Then, S'* is strongly conservative *over S if, for all $A, B \in L$ and for all sets Γ, Δ of L-sentences,* $\dfrac{\Gamma \vdash_{S'} A}{\Delta \vdash_{S'} B}$ *only if* $\dfrac{\Gamma \vdash_{S} A}{\Delta \vdash_{S} B}$.

Simply put, the addition of new rules introducing new vocabulary is weakly (strongly) conservative if and only if all the new sequents (rules) provable in the extended system involve new vocabulary. In Belnap's own words:

> We may now state the demand for the consistency of the definition of the new connective, *plonk*, as follows: the extension must be *conservative*; i.e., although the extension may well have new deducibility-statements, these new statements will all involve *plonk*. (Belnap, 1962, p. 132)

Dummett and Ian Hacking follow suit:

> I shall presently qualify the suggestion that [...] operational rules can be regarded as definitions. I claim here only that if we are to pursue that idea, we shall require that the definitions be conservative. (Hacking, 1979, pp. 237-8)

> The best hope for a more precise characterisation of the notion of harmony lies in the adaptation of the logicians' concept of conservative extension. (Dummett, 1991b, pp. 217-18)

Because of its intrinsically global character, Dummett dubs the requirement of (presumably, strong) conservativeness *total harmony*. Prior's `tonk` is spectacularly non-conservative, and hence disharmonious, in this sense: provided the old system could prove at least one theorem, the rules for `tonk` now allow us to prove *any* sentence in the old language!

Separability, conservativeness, and local constraints

We now prove that separability and conservativeness are equivalent requirements. In our next step, we begin to investigate the relationships between separability and conservativeness, on the one hand, and some key local proof-theoretic constraints, on the other.

To begin with, it is not difficult to see that a system S is weakly separable if and only if, for each logical operator \$, the rules for \$ yield a conservative extension of its structural base.

Theorem 1. *A system S is weakly (strongly) separable if and only if, for each logical operator \$, the rules for \$ yield a weakly (strongly) conservative extension of its structural base.*

> *Proof*: For the left-to-right direction, suppose S is separable, and let S^- be the system obtained by subtracting from S the rules for some logical operator $\star$. Now suppose that, in S, there is a sequent $X_1, \ldots, X_n \vdash Y$ such that (i) only \$ figures in $X_1, \ldots, X_n \vdash Y$ and (ii) $X_1, \ldots, X_n \vdash Y$ can only be proved in S by means of the $\star$-rules. That is, suppose that the $\star$-rules yield a non-conservative extension of S^-. Then, it is easy to see that our assumption that S was separable has been violated, since, $X_1, \ldots, X_n \vdash Y$ cannot be proved by means of the \$-rules alone, contrary to what separability demands. Hence, separability implies conservativeness. For the converse direction, suppose that a system S is the result of adding rules for a finite number of logical operators to a structural base **B**, so that, for every two logical operators \$ and $\star$, the rules for \$ and $\star$ respectively yield a conservative extension of (i) **B** and (ii) of **B** $\cup\{\star\text{-I}, \star\text{-E}\}$ and **B** $\cup\{\$\text{-I}, \$\text{-E}\}$. Now suppose that S is not separable, i.e. suppose that there is at least one sequent $X_1, \ldots, X_n \vdash Y$ such that (i) $X_1, \ldots, X_n \vdash Y$ is provable in S, but (ii) $X_1, \ldots, X_n \vdash Y$ is not provable by means of the rules for the logical operators figuring in it. Again, this contradicts our assumption that S was conservative over **B**. Hence, conservativeness entails separability. Likewise, it is not difficult to verify that a system S is strongly separable if and only if, for each of its logical operators \$, the rules for \$ yield a strongly conservative extension of S. The foregoing proofs easily carry over—one only needs to substitute talk of sequents with talk of rules.
> ■

It also seems plausible to conjecture that rule-purism, the view that admissible I-rules are pure, simple, sheer, and single-ended, and E-rules must be in harmony with the corresponding I-rules, implies separability. The converse implication, on the other hand, is more problematic. Consider the following 'rules':

$$\odot\text{-I}\,\frac{A \odot B}{A \odot B} \qquad \odot\text{-E}\,\frac{A \odot B}{A \odot B}.$$

These rules can be harmlessly added to any separable system. However, on the plausible assumption that these rules are pure but not single-ended, they provide a counterexample to the claim that separability implies rule-purism.

As for the relations between harmony and separability, let us consider first whether strong separability implies harmony. To this end, consider a standard formalization of intuitionistic logic, minus one half of the rule of conjunction elimination—call this formalization $\mathbf{Nip}^-$. Then, $\mathbf{Nip}^-$ is both weakly and strongly separable, but some of its rules, to wit, the rules for $\wedge$, are not harmonious. Separability, both weak and strong, does not imply harmony. The converse implication does not in general hold either, as we shall see in § 6.2.2.

Orthodox and ultra-orthodox inferentialism

Before we turn to the question whether harmony implies separability and conservativeness, let us first introduce the global analogue of rule-purism—what I shall call *orthodox inferentialism*. Following Tennant (1997, p. 294) and Shapiro (1998, p. 611), let us say that an inference $\Gamma \vdash A$ is *strictly analytic* if A can be derived from Γ in a separable system, i.e. by means of a proof in which only operational rules for the logical operators occurring in A or Γ are used, and that it is *loosely analytic* if it can be derived by means of rules which are themselves strictly analytic. (Notice that meaning-constitutive rules trivially qualify as strictly analytic.) Then, I will call inferentialists who think that $\Gamma \vdash A$ is logically valid if and only if there is a strictly analytic derivation of A from Γ *orthodox inferentialists*.

If admissible I-rules are required to satisfy Dummett's complexity condition, single-endedness, separability and conservativeness entail that admissible systems must satisfy the *subformula property*:

Definition 14. *(Subformula property) A system S has the* subformula property *if, whenever $\Gamma \vdash_S A$, then there is a proof of A from Γ every line of which is either a subsentence of A or a subsentence of one of the sentences in Γ.*[4]

Accordingly we may say that an inference $\Gamma \vdash A$ is *ultra strictly analytic* if A can be derived from Γ by means of a derivation satisfying the subformula property. And we may call inferentialists who think that $\Gamma \vdash A$ is logically valid if and only if there is an ultra strictly analytic derivation of A from Γ *ultra orthodox inferentialists*. (In what follows, I will exclusively focus on orthodox inferentialism.)

6.2.2 Some conjectures

It is tempting to think that, insofar as harmony requires intrinsic harmony (i.e. reducibility), harmony implies separability—and hence conservativeness. The

[4]Notice that the subformula property entails separability and conservativeness, but the converse direction does not hold.

reason is simple: reducibility guarantees that elimination rules do not allow us to prove anything that we could not have already proved by means of the corresponding introductions. Hence, how could harmonious rules ever be non-conservative?

Dummett's conjecture

In keeping with the foregoing considerations, Dummett conjectures in *The Logical Basis of Metaphysics* that "intrinsic harmony implies total harmony in a context where stability [i.e. any adequate conception of harmony] prevails" (Dummett, 1991b, p. 290). In other words: any adequate account of harmony entails conservativeness. Harmonious I- and E-rules should always yield conservative extensions of the systems to which they are added. Let us term this *Dummett's conjecture.*

As Prawitz (1994) first pointed out, Dummett's conjecture is false.[5] Indeed, we have *already* presented a counterexample to it: Read's •. Here are two, related, counterexamples. First, consider the following I- and E-rules for set abstraction, first introduced by Prawitz (1965, p. 94):

$$\in\text{-I}\,\frac{\phi[t/x]}{t \in \{x : \phi(x)\}} \qquad \in\text{-E}\,\frac{t \in \{x : \phi(x)\}}{\phi[t/x]}$$

These rules are perfectly harmonious, if anything is. Yet, as we know, they yield inconsistency, and, in non-paraconsistent logics, triviality. Russell's Paradox can be derived by letting r be the term $\{x : x \notin x\}$ (see Prawitz, 1965, p. 95). Frege's infamous and yet harmonious Basic Law V is another case in point:

$$\text{BL-V}\,\frac{\forall x(Fx \leftrightarrow Gx)}{\forall F \forall G(\{x : F(x)\} = \{x : Gx\}} \qquad \frac{\forall F \forall G(\{x : F(x)\} = \{x : Gx\}}{\forall x(Fx \leftrightarrow Gx)}$$

Basic Law V famously yields, too, a version of Russell's Paradox. For better or worse, harmony alone implies neither separability nor conservativeness.

A more plausible conjecture

Perhaps Dummett may insist that harmony implies conservativeness on the further assumption that reduction-steps reduce the degree of complexity of local peaks, where the *degree* of a formula A is defined as the number of occurrences of logical operators in A, except $\bot$ (Prawitz, 1965, p. 16). For consider the following •-reduction:

[5] We shall consider Prawitz's own counterexamples in § 6.4 below, in the context of our discussion of higher-order logics.

Example 8. *$\bullet$-reduction:*

$$\dfrac{\dfrac{\begin{matrix}[\bullet]^n\\ \Pi_1\\ \bot\end{matrix}}{\bullet}\,{\scriptstyle \bullet\text{-}I,\, n} \qquad \begin{matrix}\Pi_2\\ \bullet\end{matrix}}{\bot}\,{\scriptstyle \bullet\text{-}E} \qquad \leadsto_r \qquad \begin{matrix}\Pi_2\\ \bullet\\ \Pi_1\\ \bot\end{matrix}$$

Now let the degree of a derivation be determined by the degree of its most complex formula. Then, unlike the reductions for the standard (intuitionistic) logical operators, whose rules satisfy rule-purism, here the local peak has been 'levelled', but the reduction has not reduced its degree.

But why should reducibility imply that reductions lower the degree of derivations? To assume that harmony must imply reducibility in *this* sense is tantamount to incorporating the complexity condition in the requirement of harmony itself. It is best, however, to keep these two requirements apart. A more reasonable conjecture, therefore, and perhaps a more charitable reading of Dummett, is that harmony *and the—nota bene, unstarred—complexity condition* jointly entail conservativeness. This conjecture is very likely to be true, as we shall see, but, insofar as we lack compelling arguments for Dummett's complexity condition, it does not provide a reason for requiring separability and conservativeness.[6] It is now time for us to examine some of the inferentialist's reasons for requiring separability and conservativeness.

6.3 The Argument from Separability

In the first part of this section, we shall introduce two arguments for separability and conservativeness, and isolate their common core. In our next step, we will begin to explore the revisionary consequences of the requirements of separability and conservativeness—what I shall call the Argument from Separability. The argument rests on assumptions that are strictly stronger than the ones required for the Argument from Harmony presented in § 5.5. Unsurprisingly, it also yields a much stronger conclusion.

6.3.1 The Argument from Analyticity

We said in Chapter 4 that logical inferentialists typically claim that logically valid arguments are valid in virtue of the meanings of the logical vocabulary. In short:

[6]One may also conjecture that harmony and the complexity condition* entail separability and conservativeness. But they do not, as we shall see in § 6.4.2.

logic is *analytic*. Now suppose that

(i) valid inferences are valid in virtue of the meaning of the logical expressions occurring in them,

and that

(ii) the correct use of a logical operator $ is fully determined by its I- and E-rules, or some subset thereof.

Then, all the inferences in which *only* $ figures must be derivable by means of the rules for $. If logic is analytic, it would seem, logical systems must be separable. Thus, Milne writes:

> Conservativeness is an extremely natural requirement from the proof-theoretic perspective. Granted (i) that logically valid inferences are valid in virtue of the meanings of the logical operators occurring in them and (ii) that the meaning of a logical operator is given by (some subset of) its introduction and elimination rules, it follows that we ought never to be in the position of declaring an inference valid that nonetheless cannot be derived without application of rules governing an operator not occurring in the inference. (Milne, 2002, p. 521)

The requirement is indeed "extremely natural", given Milne's assumptions. These assumptions warrant a couple of remarks, however.

To begin with, Milne's claim that valid arguments are "valid in virtue of the meanings of the logical operators occurring in them" amounts to requiring that I- and E-rules are complete in a very strong sense: $-statements must be provable by means of a proof *each of whose steps* is taken into accordance with some $-rule—not just the *last* step, as the Fundamental Assumption demands.[7] Second, Milne's contention that valid inferences are valid in virtue of the meanings of the logical operators *occurring in them* requires nothing less than the rejection of logical holism, the view that the meanings of the logical constants are all interdependent.[8] We will return to this point in § 6.3.4 below. For the time being, let us move on to consider a second argument for separability and conservativeness.

[7] Recall Milne's words, which we already quoted in § 4.1.1: "[every correct] use of the constant in question is [...] in some sense to be specified, derivable and/or justified on the basis of the putatively meaning-conferring rule or rules" (Milne, 1994, pp. 49-50).

[8] We shall introduce some examples of arguments that are not valid in virtue of the rules for the logical operators occurring in them in § 6.3.5 and 6.4 below.

6.3.2 Belnap's Argument for Structural Conservativeness

In his response to Prior's `tonk`, Belnap (1962) offers the following argument for conservativeness. Consider a language L, containing only atomic sentences, and let $\vdash$ be the deducibility relation induced by (i) the inferential relations among the atoms in L, and (ii) *the structural rules* of the system. Call this initial base system **B**. Finally, assume that $\vdash$ expresses *all* the inferential relations among atoms. In Belnap's words, "this little system [...] express[es] all and only the universally valid statements and rules expressible in the given notation: it completely determines the context" (Belnap, 1962, p. 132). Then, Belnap requires that the introduction of logical vocabulary may not affect the inferential relations among atoms, the justification for this being "precisely our antecedent assumption that we already had all the universally valid deducibility-statements not involving any special connectives" (Belnap, 1962, p. 132). For every purported logical operator \$, the result of adding the \$-rules to **B** should yield a conservative extension of **B**. Call this *structural conservativeness*.

Definition 15. *(Structural conservativeness) A set of I- and E-rules for a logical operator \$ is structurally conservative if and only if it yields a conservative extension of our base system **B**.*

Notice that structural conservativeness is a much weaker requirement than conservativeness *tout court*: unlike the latter, it leaves it open that the rules for an operator $\star$ may not be conservative over the system obtained by adding to **B** the rules for some other operator \$, while both the $\star$- and the \$-rules are individually conservative over **B**.

It may be objected that we will never be in a position to capture all the inferential relations among atoms, once and for all, in a single deductive system. But this objection would miss the point. The real motivation for requiring conservativeness over **B** is that logic should not be creative: logical rules *alone* should not allow us to prove atomic statements that we could not otherwise have proved. On these assumptions, inferentialists are in a position to rule out both Prior's `tonk` and Read's $\bullet$. These operators respectively validate $A \vdash B$ and $\vdash \bot$, thereby violating our assumption that either **B** contained *all* the inferential relations among atoms, or logic alone should not be creative.

6.3.3 Bendall's generalization of Belnap's argument

In Bendall (1978), Kent Bendall argues that Belnap's requirement of conservativeness generalizes. His main assumptions are that (i) basic inference rules are complete with respect to the logical expressions they introduce and eliminate, and (ii) the meanings of the logical operators are independent of each another.

Local completeness

Bendall writes:

> on the same grounds [...] it seems plausible to generalize this requirement [structural conservativeness] along the following lines. Suppose a set of single-operator rules introduced as defining a logical operator is conservative relative to specified structural rules, and that a further set of single-operator rules is proposed by way of introducing a second logical operator. Then it would seem plausible, for all the same reasons, to require that this new set be conservative relative to the combination of the structural rules and the previously admitted logical rules. But the order in which these two sets of logical rules are introduced should not matter. Hence the first set should also be required to be conservative relative to the combination of the structural rules and the second set. The obvious continuation of this line of thought leads to the requirement that the calculus determined by the rules at each extension be separable. (Bendall, 1978, p. 255)

Suppose we add to **B** the rules for a logical operator \$, and that these rules are conservative over **B**. Suppose also that the \$-rules are *locally complete*. That is, they are complete with respect to \$'s correct uses: they yield all the "universally valid" inferences in which \$ figures. Then, Bendall argues, "for all the same reasons" the rules of a *second* logical operator $\star$ should themselves be conservative over $\mathbf{B} \cup \{\text{\$-I}, \text{\$-E}\}$.

Order does not matter

However, Bendall adds, order should not matter: one might have as well added $\star$ first, and then \$. It follows that the \$-rules should be conservative not only over **B**, but also over the rules *for each of the logical operators*. Bendall persuasively argues that the argument generalizes to the requirement of strong separability:

> But when we add a logical operator and corresponding logical rules, why should we require only that no new sequents not involving that operator should become provable? Shouldn't we require further that no new sequent rules not involving that operator should become derivable? It seems that the same considerations that call for the requirement of [separability]—i.e., that facts about entailment pertaining to certain sentences depend only on universal properties of the entailment relation and the meanings of the logical operators occurring in these sentences, as determined by entailment-theoretic definitions—call for strong [separability] as well [...]. (*Ibid.*)

If the \$-rules are locally complete, and if order does not matter, inferentialists have a compelling argument for requiring weak separability and strong separability. That is, if \$'s I- and E-rules *alone* determine \$'s meaning, and if whatever determines the meaning of an expression determines the *totality* of its possible uses, the introduction in the language of a new logical operator $\star$ may not validate new inferential relations among \$-sentences, on pain of altering \$'s meaning. Thus Dummett:

> when [...] a logical constant [...] is introduced into the language, the rules for its use should determine its meaning, but its introduction should not be allowed to affect the meanings of sentences already in the language. If, by its means, it becomes possible for the first time to derive certain such sentences from other such sentences, then either their meanings *have* changed, or those meanings were not, after all, fully determined by the use made of them. (Dummett, 1991b, p. 220)

6.3.4 Order, purity, and inferential packages

I will argue in § 6.4.2 that Milne's and Bendall's arguments are, at least in some contexts, unsound: one cannot always assume that I- and E-rules are locally complete. For the time being, though, let us briefly focus on Bendall's anti-holistic assumption that order does not matter.

Bendall assumes that meaning-constitutive rules are *pure*—this is what his reference to "set[s] of single-operator rules" amounts to. Yet, as we have seen in § 6.1.3, the arguments for requiring purity are weak, and potentially question-begging. In absence of compelling reasons for requiring purity, therefore, we may take Bendall's assumption that order does not matter to apply to rules that are not necessarily pure. That is, we may assume that the meanings of some logical constants are ordered by a well-founded relation of meaning-dependence—for instance, the rules for a constant \$ may involve a second constant $\star$, provided that the rules for $\star$ do not themselves involve \$, nor any other constant figuring in the rules for \$. Then, the meaning-constitutive rules for \$ will be the rules for \$ *and* the rules for $\star$, and Bendall's argument will require that the rules for \$ yield a conservative and separable extension of $S \cup \{\star\}$. Call the meaning-constitutive rules for a constant \$ \$'s *inferential package*. We may then take Bendall's assumption that order does not matter to apply to inferential packages: inferential packages as a whole should yield conservative extensions of our base system.

Even dropping the requirement of purity, though, Bendall's and Milne's arguments still require the falsity of logical holism. Insofar as we require the relation of meaning-dependence to be well-ordered, we are ruling out the possiblity that there be a single inferential package which simultaneously defines *all* the logical constants.

To be sure, the idea that the meanings of the logical constants are all interdependent does not seem very plausible—or, at least, it does not seem plausible to the present author. The logical holist needs to claim that logical expressions such as $\exists$ and $\wedge$ are just like 'father' and 'son': their meanings are interdependent, and one cannot learn one independently of the other. But, although this seems plausible in the case of expressions like 'father' and 'son', it seems far less plausible in the case of $\exists$ and $\wedge$. It would seem that, *as a matter of fact*, our understanding of the existential quantifier and our understanding of conjunction are independent of one another.

This, however, is by no means an *argument* against logical holism: we are merely opposing to logical holism the *intuition* that it is after all false. We cannot but conclude, then, that Milne's and Bendall's considerations in favour of separability do not lead us very far. For all the inferentialist has told us, the question whether separability holds or not crucially depends on our choice the meaning-constitutive rules for the logical constants—precisely the sort of *conclusion* we were expecting from an inferentialist argument for logical revision. Can inferentialists do better?

6.3.5 Separability and understanding

Perhaps inferentialists should frame their argument for separability as an argument from *understanding*, rather than meaning itself. Here is how such an argument might go. Suppose we can establish that, say, the rules that are constitutive of our understanding of $\rightarrow$ are indeed, *pace* McGee and Field, $\rightarrow$, $\rightarrow$-I and $\rightarrow$-E. Then, one might argue, these rules form a very bad inferential package for the material conditional, since they do not allow us to assert Peirce's Law

$$\text{(PL)}\ ((A \rightarrow B) \rightarrow A) \rightarrow A$$

and they do not allow us to derive Peirce's Rule

$$\begin{array}{c} [A \rightarrow B]^i \\ \vdots \\ \text{Peirce's Rule, } i\ \dfrac{A}{A}, \end{array}$$

both of which are nevertheless valid with respect to the standard classical semantics. In other words, $\rightarrow$-I and $\rightarrow$-E are not locally complete, i.e. they do not allow us to prove the entire $\rightarrow$-fragment of classical propositional logic (henceforth, **CPL**). But since **CPL** *is* complete, this means that there is a gap between $\rightarrow$-I and $\rightarrow$-E, on the one hand, and $\rightarrow$'s correct use, on the other—that is, on the foregoing assumptions, there is a gap between our understanding of $\rightarrow$ and its correct use. Yet, how could there be a gap between our understanding of a *logical* expression and its correct use, if (i) understanding is equated with deductive competence, and (ii) we know that standard formalizations of **CPL** are sound and *complete* with respect to classical semantics, i.e. we know that they allow us to derive, for any constant \$, all and only the correct uses which essentially involve \$? At the very least, it would seem, the rules that are constitutive of our understanding of a constant \$ should account for *all* the deductive uses of \$. In a complete system, this is tantamount to requiring that the rules that are constitutive of our understanding of \$ be locally complete, i.e. that they allow us to derive all the rules which essentially involve \$.

In a nutshell, the argument may be put as follows. Understanding a logical expression \$ belonging to a complete logical system requires mastering *all* of \$'s deductive uses. Given our cognitive limitations, however, these deductive uses cannot all be individually learned: one can only be said to grasp them all if one possesses a finite method for producing them. Meaning-constitutive rules can be such a method, provided they allow speakers to derive, at least in principle, all the deductive uses of the logical expressions they define. Only in this sense grasp of the meaning-constitutive rules for a constant \$ can be equated with our understanding of \$. Hence the requirement of separability: if the rules for $\star$ are required for deriving some \$-rules or sequents, and if $\star$ does not figure in

$'s inferential package, then the rules for $ alone may not be constitutive of our understanding of $.

It should be clear, though, that this argument is no less problematic than Milne's and Bendall's own arguments. To be sure, if we take $'s I- and E-rules to be constitutive of our understanding of $, then inferentialists may require, in light of the foregoing considerations, that these rules allow us to prove the entire $-fragment of **CPL**. For instance, if we take $\to$-I and $\to$-E to be constitutive of our understanding of $\to$, then it *may* be plausible to require that these rules allow us to prove the entire $\to$-fragment of **CPL**, given that we know **CPL** to be complete. However, logical holists would never subscribe to the claim that our understanding of $\to$ is constituted by our grasp of $\to$-I and $\to$-E *alone*: they would rather stress that our understanding of $\to$, like our understanding of all the logical connectives, is given by *all* the rules of **CPL**. Once more, the argument from separability, however framed, requires *as an assumption* the falsity of its main target: logical holism.

6.3.6 Logical holism and the possibility of logical atomism

Logical inferentialists, I would like to suggest, may still be in a position to argue against logical holism. The line of argument I have in mind assumes that it is *possible* that subjects grasp logical operators independently of one another. For instance, one *could* understand $\vee$ without thereby understanding $\forall$. If we take this assumption on board, logical holists may be in trouble. The reason is that they are forced to *deny* this possibility. For if the meanings of the logical expressions are all interdependent, as the holist suggests, one could not understand $\vee$ without understanding $\forall$. On the holistic view, speakers who understand $\vee$ without understanding $\forall$ have a *different understanding* of $\vee$ than the ones who understand them both. And yet, for all we know, it may actually *be* that one can grasp a logical operator $ without thereby grasping some other logical operator $\star$. If this is true, however, given logical inferentialism, any formalization of logic that does not satisfy the requirements of separability and conservativeness does not correctly describe our actual logical practice. The upshot is that logical holism may require the falsity of an empirical claim—one, we may add, that is very likely to be true.

On the other hand, if it is thought that, for any two logical operators $ and $\star$, the meanings of $ and $\star$ could have been learned independently, then admissible formalizations of logic should respect separability and conservativeness, on pain of ruling out this seemingly plausible possibility.

It may be objected that it is also possible that the totality of the logical operators can only grasped *en bloc*, as the holist would have it. Hence, one might argue, the advocate of separability and conservativeness also runs the risk of misrepresenting our actual logical practice. However, even if this possibility can be coherently entertained, there is a crucial asymmetry here. To begin with, the intuitionist is, unlike the classicist, a revisionist—indeed, in our terminology, a radical one. She does not seek to *describe* our actual inferential practice: she urges a *revision* of our practice. Intuitionistic logic only requires that the meanings of the logical constants *could* be learned independently of one another, which is consistent with the possibility that our understanding of the logical constants is actually holistic, as classical logicians typically maintain. Secondly, whilst lack of separability is inconsistent with the idea that we could learn the meanings of the logical operators independently of one another, the separability of a logical system is perfectly consistent with the possibility that we actually grasp logical rules *en bloc*. Hence, while logical holism runs the risk of being empirically false, orthodox inferentialism does not—at least not on these grounds.

Logical holists may further object that the foregoing argument rests on the assumption that the logical operators defined in separable formalizations of logic are the same logical concepts defined in non-separable formalizations.[9] After all, the holist might insist, if the rules are different, the meanings are also different—hence, the assumption on which the argument relies is untenable. The problem with this rejoinder, however, is that there *is* a dimension of meaning along which the logical concepts are the same. If the logic is classical, for instance, the rules—even if different—determine the same truth-functions, and hence, in a sense, the same meanings. Different logical systems that are sound and complete with respect to the same semantics differ in the way they depict our *understanding* of the logical operators—their Fregean sense. But the operators themselves are the same.

6.3.7 The Argument from Separability

Standard natural deduction regimentations of classical logic do not sit very well with the idea that we could learn the meanings of the logical expressions independently of one another. Indeed, standard formalizations of classical logic are *incompatible* with the requirement of separability, as a relatively little-known theorem by Hughes Leblanc shows:

[9] Thanks to Dominic Gregory for having raised this potential concern.

Theorem 2. *(Leblanc, 1966, p. 35) If either* Double Negation Elimination *or* classical reductio *(or some equivalent rule) are taken to partly determine the meaning of classical negation, then no complete natural deduction formalization of classical logic is separable.*

Thus, Bendall writes:

> certain facts pointed out by Leblanc (1966) as "shortcomings of natural deduction" cause trouble (and otherwise it is not clear why they should be called "shortcomings"). Namely, Leblanc shows, in effect, that no classically complete [natural deduction formalization of classical logic] is separable. So even the weak form of the separation problem for such languages appears to be unsolvable. And hence, by our generalization of Belnap's requirement of conservativeness, it would seem to follow that there are familiar first-order logical operators which cannot be assigned their full classical meaning or force by an entailment-theoretic definition. (Bendall, 1978, p. 256)

For instance, consider a natural deduction formalization of classical logic, call it $\mathbf{Ncp}^{DN}$, obtained by adding the rule of Double Negation Elimination (DN) to a standard natural deduction formalization of intuitionistic logic. As Bendall points out, $\mathbf{Ncp}^{DN}$

> is not separable [...]. One symptom of this is the well-known fact that there are classically valid [rules] not involving $\neg$ which cannot be proved in $\mathbf{Ncp}^{DN}$ without using the rule ($\neg$-E)—namely, all and only those classically valid [rules] which do not involve negation and are not intuitionistically provable–of which $((A \rightarrow B) \rightarrow A) \rightarrow A$ and $(\forall x)(A \vee C) \rightarrow A \vee (\forall x)C$ are well-known examples. The trouble thus seems to be that the pair $\{(\neg\text{-I}), \text{DN}\}$ cannot be admitted as defining $\neg$ since it is not conservative relative to the remaining rules of $\mathbf{Ncp}^{DN}$. If one assumes that the intelim pairs for the other operators [...] are 'right', then on the basis of the preceding considerations one might argue that ($\neg$-I) or ($\neg$-E) must be weakened. But no such weakening can leave the system classically complete. (*Ibid.*)

Standard natural deduction formalizations of classical logic *cannot* be separable. And, insofar as speakers *could* come to grasp the logical operators independently of one another, this seems to be bad news for the classical logician. For now, if logical inferentialism holds, the classicist's contention that standard natural deduction formalizations of classical logic correctly describe our logical practice rests on the very strong assumption that the meanings of the logical constants—in particular, the meanings of $\rightarrow$, $\vee$, and $\neg$—could *not* be grasped independently of one another.

Let us call this the *Argument from Separability*: if negation is partially defined by DN (or by some equivalent rules, such as classical *reductio*), then classical logic does not even satisfy weak separability. This means, however, that the following four claims

(a) I- and E-rules are complete with respect to the correct uses of the logical vocabulary,

(b) valid inferences are strictly analytic,

(c) order does not matter, and

(d) classical logic is the correct logic

form an inconsistent set. The conclusion of this argument is stronger than the conclusion of the argument from harmony: it shows that, given certain assumptions, *no* formalization of classical logic satisfies certain proof-theoretic requirements, not just the existing ones. As we have already stressed, however, the argument also requires stronger assumptions. Specifically, it requires that I- and E-rules be complete in a sense that is strictly stronger than the one required by the Fundamental Assumption.

How can classicists react? If we grant the assumption that order should not matter, there are at least four main options. First, they may try to show, perhaps empirically, that logical holism is actually true. Second, they could reject Milne's and Bendall's assumption that I- and E-rules must be in general locally complete. Third, they may seek to meet the inferentialist's challenge head on, by providing a strongly separable formalization of classical logic. Fourth, they may try to show that intuitionistic logic is *itself* at odds with logical inferentialism. The third option will be our concern in Chapter 7. The first option, if viable at all, will be left for another occasion. A possible way of implementing the fourth option will be considered in § 6.5 below.

In the next section, we shall explore the prospects for the second option. As we have seen, Milne's and Bendall's assumption that I- and E-rules are locally complete, together with the rejection of logical holism, is the main motivation for requiring separability and conservativeness. Yet, one can show that the rules for the *higher-order quantifiers* are not locally complete, and that, as a result, separability and conservativeness fail for *higher-order* logics.

6.4 Conservativeness and higher-order concepts

We have seen that the inferentialist's main assumption for requiring conservativeness is the *completeness* of the systems to which the logical vocabulary is added, in the following sense: the rules for a logical operator \$ must allow us to derive *all* of \$'s intuitively correct inferential uses. This was, as we already noted at the end of § 4.1.1, the second possible interpretation of the first reading of the Determination Thesis, that basic rules determine meanings. But there is a hitch. Given sufficiently rich expressive resources, this cannot in general be required: there are expressions some of whose correct uses may not be validated by their introduction and elimination rules alone. For the logical inferentialist, this need not be a problem, insofar as incompleteness affects *non-logical* expressions. However, unfortunately for the inferentialist, incompleteness affects logical and non-logical expressions alike. Section 6.4.1 considers, and opposes, Stephen Read's and Stewart Shapiro's contention that the rules for the truth-predicate already create trouble for the logical inferentialist. Section 6.4.2 shows that the real problem for logical inferentialism is rather caused by higher-order logics, and argues that the worry disappears upon reflection. Higher-order logics do not give us reasons for relaxing the requirements of separability and conservativeness for *complete* logics, such as classical propositional logic.

6.4.1 Truth and conservativeness

In his review of Dummett's *Logical Basis of Metaphysics*, Prawitz writes:

> from Gödel's incompleteness theorem we know that the addition to arithmetic of higher-order concepts may lead to an enriched system that is not a conservative extension of the original one in spite of the fact that some of these concepts are governed by rules that must be said to satisfy the requirement of harmony. (Prawitz, 1994, p. 374)

Following up on this, Shapiro (1998, pp. 616-7) and Read (2000, p. 127) have argued that the truth-predicate is a case in point. Both take the truth-predicate to be governed by the following rules:

$$\text{T-I}\ \frac{A}{T\ulcorner A\urcorner} \qquad \text{T-E}\ \frac{T\ulcorner A\urcorner}{A}\,.$$

They then observe that, when we add a truth-predicate to Peano Arithmetic (PA, with suitable restrictions to ward off paradox), it is a routine exercise to show that a Gödel sentence of PA, call it G, can now be proved. For instance, one may argue within $\mathsf{PA} \cup \{\mathsf{T\text{-}I}, \mathsf{T\text{-}E}\}$ that (i) all of PA's axioms are true, and (ii) all of its inference rules are truth-preserving, and thereby infer its consistency. But this is enough to infer G itself, given that $\mathsf{Cons(PA)} \to G$ is itself a theorem of PA (where 'Cons(PA)' says that PA is consistent). Tennant (1997, pp. 293-4) suggests the following more direct proof. He first notices that G is of the form $\forall n G(n)$. He then observes that each instance $G(0)$, $G(s0)$, $G(ss0)$, etc. is provable in the metatheory. Hence, Tennant concludes, since each instance is true, the universal quantification must also be true. This reasoning can be represented by adding to PA a primitive truth-predicate: we infer from $\mathsf{PA} \vdash G(0)$, $\mathsf{PA} \vdash G(s0)$, $\mathsf{PA} \vdash G(ss0)$, etc. that each instance is *true*; and, by allowing instances of the primitive truth-predicate to to appear in the induction axiom, we conclude by mathematical induction that $\forall n G(n)$ must be also true. The addition of the harmonious rules for the truth-predicate, Shapiro and Read maintain, has yielded, *pace* Dummett, a non-conservative extension. Orthodox inferentialism is false. Or is it?

This example is not yet decisive. Shapiro argues that, "on the [inferentialist] view, the predicate T qualifies as logical", since it "is governed by an introduction rule (T-I) and an elimination rule (T-E)" and "one can argue that for present purposes at least, the rules fully constitute the meaning of T" (Shapiro, 1998, p. 618). But this is problematic. The truth-predicate is not governed by its introduction and elimination rules *alone*. If it is to be strong enough to allow us to run a soundness proof, its introduction and elimination rules must be supplemented with non-harmonious compositional axioms, such as

($\wedge_T$) For any formula two formulas A, B and any function s assigning objects to their free variables, $\ulcorner A \wedge B \urcorner$ is true relative to s if and only if $\ulcorner A \urcorner$ is true relative to s and $\ulcorner B \urcorner$ is true relative to s.

Although the truth-rules allow us to prove *each instance* of these axioms, they do not allow us to prove the axioms themselves. Let TPA be PA supplemented with the T-rules, restricted to purely arithmetical statements. Decisively, it can be shown that, because of the *compactness* of TPA, TPA is *conservative* over PA.[10]

[10]*Proof* (Halbach, 2005, § 3.1): Suppose for *reductio* that PA together with the T-rules, and hence all the T-sentences, proved an arithmetical sentence A not provable in PA. Then, by the *Compactness Theorem*, it would follow that a finite subtheory (with finitely many T-sentences) proves A. However, the finite subtheory can be translated in PA by interpreting T by an appropriate partial truth predicate; the arithmetical vocabulary is not affected by this interpretation. But then, A is already provable in PA, contrary to what we had assumed. ■ The Compactness Theorem states that if a set Γ of sentences is consistent, then every finite subset of Γ is also consistent. It follows that, if A is a logical consequence of a set of sentences Γ, then A is a logical consequence of some finite subset Δ of Γ.

Hence, they do not allow us to prove G.[11]

All the same, Shapiro correctly observes that, once we have added a truth-predicate to PA, our arithmetical sentence G becomes provable, but *only* by making a detour through non-arithmetical vocabulary:

> Although the sentence G consists only of arithmetic terminology, to establish G we must invoke something other than the meanings of the arithmetic terminology. In a sense, we have to change the subject [...]. (Shapiro, 1998, p. 615)

In our terminology, G is not strictly analytical. Hence, *if* we identify (at least one dimension of) its meaning with the totality of its correct uses, its meaning is not fully determined by the meaning of its compound expressions.[12] In this sense, Shapiro claims, we have a change of subject; more precisely, a change of meaning. *Logical* inferentialists may insist that, whether this change of meaning is a problem or not, the issue need not concern *them*, given that G is an *arithmetical* truth. But the problem surfaces again in higher-order logics.

6.4.2 Higher-order logics

In what follows, I will be assuming that inferentialists are not only willing to countenance first-order quantifiers, i.e. quantifiers whose bound variable ranges over objects, but also *higher-order* quantifiers, i.e. quantifiers whose bound variable ranges over properties or sets, properties of properties, sets of sets. etc. I cannot defend this assumption here, but let me nevertheless mention two considerations in its favour. First, higher-order logics appear to be indispensable for carrying out a number of philosophical programs—such as, for instance, the neo-logicist approach to mathematics (see e.g. Hale and Wright, 2001). And, to my mind, one's approach to logic should not determine in advance whether such programs are correct or not. Second, there are strong reasons for thinking that higher-order logics are proper logics.[13] Insofar as these reasons are compelling, inferentialists had better be able to account for first- and higher-order logics alike.[14]

---

[11] Many thanks to Völker Halbach for helpful correspondence on this point. See also Steinberger (2009a, p. 93). However, see Field (2006) for an argument to the conclusion that properly interpreted schemata allow us, contrary to the received view, to prove compositional axioms such as ($\wedge_T$) from T-I and T-E (and logic) alone. Considerations of space prevent me from assessing Field's proposal.

[12] We will return to this point in § 6.4.5 below.

[13] For a sustained defense of higher-order logics *from a proof-theoretic perspective*, see Rossberg (2006). The *locus classicus* for a defense of higher-order logics in general is, of course, Shapiro (1991).

[14] It may be objected that these reasons are unavailable to *inferentialists*, because, from an inferentialist perspective, higher-order logics are not really higher-order. In a nutshell, the problem is

Higher-order quantifiers

Consider the following standard natural deduction rules for the second-order quantifiers:

Example 9. *The second-order universal quantifier:*

$$\forall^2\text{-}I\,\frac{\Phi[F^n/X^n]}{\forall X^n\Phi(X^n)} \qquad \forall^2\text{-}E\,\frac{\forall X^n\Phi(X^n)}{\Phi(T^n)}.$$

The usual restrictions apply.[15] These rules are clearly harmonious: the elimination rule allows us to infer from $\forall X^n(\Phi)X^n$ precisely what was required to introduce it in the first place. However, on the assumption that rules are *open-ended*, i.e. that they hold for all possible extensions of the language, they do not respect Dummett's complexity condition. Unlike the first-order rules, where, if $\Phi[t/x]$ is a subformula of $\forall x\Phi x$, then $\Phi[t/x]$ is logically less complex than $\forall x\Phi x$, $\Phi[F^n/X^n]$ is not guaranteed to be logically less complex than $\forall X^n\Phi(X^n)$, even if it is technically a subformula of $\forall X^n\Phi(X^n)$ (see Leivant, 1994, pp. 24-5). The reason is that F^n can be a predicate of unbound complexity, i.e. of potentially higher-complexity than $\forall X^n\Phi(X^n)$.[16] On the other hand, the I-rules for the second-order quantifiers (and n^{th}-order I-rules more generally) satisfy our more liberal *complexity condition**. Even if, in some of their applications, their premises are logically more complex than their conclusion, their *schematic formulation* is such that their conclusion *is* logically more complex than any of their premises.

As a result, even if higher-order logics in general satisfy some subformula property, this property does not guarantee that "the logical complexity of formulas in a normal proof is [...] bounded by the complexity of the derived formulae" (Leivant, 1994, p. 29). For instance, as we shall see in a moment, the rule of $\forall^2$-I can be instantiated to a *third-order* formula, thus violating Dummett's complexity condition, that a legitimate I-rule should be such that *in any of its applications* the conclusion be of higher complexity of any of the assumption and of any discharged hypothesis. This means, as we shall see, that conservativeness and separability

that higher-order quantifiers can receive different semantic interpretations, depending on what we take to be the domain of quantification of higher-order variables. The problem arises, then, as to how we can distinguish these interpretations, from a merely proof-theoretic perspective. I argue in Appendix D that inferentialists have the resources to address this objection.

[15]Thus, in $\forall^2$-I, F^n is a n place predicate letter which does not occur in any of the assumptions on which $\Phi[F/X^n]$ depends, and, in $\forall^2$-E, T^n is an open sentence with n argument places such that every variable bound in $\Phi(T)$ is already bound in T. See Rossberg (2006). Notice also that $\forall^2$-E conveys the Comprehension Principle, that $\exists R\forall x_1,\ldots,x_k(R(x_1,\ldots,x_k) \leftrightarrow \Phi)$, where $k \leq 0$, R is a k-ary relation-variable, and Φ is a second-order formula in which R does not occur free. See Leivant (1994, pp. 23-4).

[16]Hence, T^n in $\forall^2$-E can also be instantiated by formulae of unbound complexity.

fail: there are higher-order sentences that can only be proved via a detour through the rules for logical operators not figuring in them.

Non-conservativeness

This result can be constructively proved.[17] Let PA_2 be the conjunction of the axioms of second-order PA:

$$\neg\exists x 0 = s(x)$$
$$\forall x \forall y (s(x) = s(y) \rightarrow x = y)$$
$$\forall X[(X0 \wedge \forall x(Xs(x))) \rightarrow \forall x Xx]$$

We then prove that for every $n > 2 \in \omega$, $n + 1^{th}$-order logic is non-conservative over n^{th}-order logic.

Theorem 3. *For every $n > 2 \in \omega$, $n + 1^{th}$-order logic is non-conservative over n^{th}-order logic.*

> Let G_2 be a Gödel sentence for second-order PA, and consider the conditional $\mathsf{PA}_2 \rightarrow G_2$. On pain of inconsistency, this conditional is unprovable in second-order PA if G_2 is. But then, so must be its universal closure
>
> $$\forall f \forall x(\mathsf{PA}_2{}^* \rightarrow G_2^*),$$
>
> where φ^* is $\varphi[f/s, x/0]$. Although this universally quantified sentence contains only second-order vocabulary, it can be proved in third-order logic. One can define, in third-order logic, a truth-predicate for second order PA (see Leivant, 1994, § 3.7), and mimic the informal reasoning that allows us to prove G_2. Then, $\mathsf{PA}_2 \rightarrow G_2$ follows by a simple step of arrow introduction. Since this proof rests on no arithmetical assumptions, we may conclude $\forall f \forall x(\mathsf{PA}_2{}^* \rightarrow G_2^*)$: third-order logic is non-conservative over second-order logic. The result generalizes: for any $n \in \omega$, $n + 1^{th}$-order logic is non-conservative over n^{th}-order logic.
> ■

It may be objected that the truth predicate definable in third-order logic is not sufficient, on its own, to prove $\forall f \forall x(\mathsf{PA}_2{}^* \rightarrow G_2^*)$, just as the rules for T are not sufficient, on their own, to prove the Gödel sentence for PA. Thus, Florian

[17] I am here following Rossberg (2006) and Wright (2007a).

Steinberger (2009a, p. 93, fn. 15) conjectures that "that case of higher-order quantification can be dealt with along similar lines as the truth predicate". However, does not attempt to verify this conjecture.

Unfortunately for the orthodox inferentialist, Steinbeger's conjecture is incorrect. For one thing, Halbach's proof that the T rules yield a conservative extension of PA cannot be adapted to the case at hand, since one of Halbach's key assumptions, compactness, only holds for *first order* theories. For another, the truth-predicate definable in third-order logic is a *real* truth-predicate, in the sense that one can show *in third order logic* that it satisfies the Tarskian compositional axioms—see Leivant (1994, § 3.7). The truth-predicate itself is not harmonious, but it is nevertheless definable in a logic that can be harmoniously formalized.[18]

Remarks

Some observations are in order. To begin with, the foregoing result applies to intuitionistic and classical higher-order logics alike: the proof of Theorem 4 does not rely on any exclusively classical assumption. The question whether higher-order logics are proof-theoretically acceptable is thus orthogonal to the question whether classical logic is proof-theoretically acceptable. The problem already arises for *intuitionist* inferentialists who are willing to avail themselves of higher-order resources.

Second, the lack of conservativeness does not affect the Fundamental Assumption. For instance, the proof of $\forall f \forall x(\mathsf{PA}_2{}^* \to G_2^*)$ ends with a step of $\forall^2$-I, as the Fundamental Assumption demands. Thus, one of the inferentialist's main motivations for requiring harmony for higher-order rules is not undermined (see *supra*, § 5.1). I-rules can still be "in a certain sense complete", even though one may have to enrich the language in order for this to be the case. As for Dummett's conjecture that harmony entails conservativeness, it is, once more, disproved: harmonious operators, even non-pathological ones, can be non-conservative. On the other hand, the more plausible conjecture that harmony *and the—nota bene, unstarred—complexity condition* jointly entail conservativeness is confirmed: the rules for the higher-order quantifiers satisfy the requirement of harmony, but they are not conservative, and they do not satisfy the complexity condition.

[18]This claim will be fully vindicated in Chapter 7, when we will introduce some harmonious and separable formalizations of classical logic.

Milne's and Bendall's arguments: what went wrong?

So what went wrong in Milne's and Bendall's arguments for separability and conservativeness? Milne's *Argument from Analyticity* assumed that (i) meanings (correct uses) are fully determined by the I- and E-rules and that (ii) valid inferences are analytic. Bendall's argument assumed that (iii) I- and E-rules are complete and that (iv) order does not matter. But even granting (ii) and (iv), inferentialists have no reason to assume, in general, (i) and (iii). The foregoing non-conservativeness results show that, for every $n \in \omega$, the $\forall^n$-fragment is incomplete. This means, however, inferentialists have no reason to require that, for every $n \in \omega$, the rules for the $n + 1^{th}$-order quantifiers be *conservative* over the rules of the n^{th}-order quantifiers. On the other hand, they may still require that the rules of the logical operators be conservative over $\mathbf{B}$, i.e. they may still require *structural conservativeness*. Moreover, and crucially for our present purposes, they may still require that the rules for logics that we know, or that we have reasons for thinking, to be complete, satisfy separability and conservativeness. Higher-order logics, therefore, need not undermine the Argument from Separability we presented in § 6.3.7. It remains to be seen, though, whether inferentialists can make sense of higher-order logics. Are higher-order logics really higher-order, from a proof-theoretic perspective? And can inferentialists account for the meaning, and for our understanding, of the higher-order quantifiers, if their meaning-constitutive rules are incomplete? Finally, can higher-order logics be made consistent with the inferentialists claim that valid logical inferences are *analytic*? I argue that inferentialists can give positive answers to these questions in Appendix D. For the time being, we shall turn to yet another possible objection to the Determination Thesis, that I- and E-rules determine the meanings of the logical operators.

6.5 Inferentialism and the Categoricity Problem

We have already mentioned in § 4.1.3 that there are reasons for thinking that basic rules *fail* to determine the truth-conditions of certain logical operators—to wit, negation, disjunction, and implication. It is now time to explore this worry in some detail. In a recent paper, Panu Raatikainen argues that logical inferentialism is undermined by some "very little known" considerations by Carnap (1943) to the effect that "in a definite sense, it is not true that the standard rules of inference" themselves suffice to "determine the meanings of [the] logical constants" (Raatikainen, 2008, p. 283). In a nutshell, Carnap showed that the rules allow for non-normal interpretations of negation and disjunction. Raatikainen concludes that "no ordinary formalization of logic [...] is sufficient to 'fully formalize' all the essential properties of the logical constants" (2008, p. 283). I suggest that this is a mistake. *Pace* Raatikainen, intuitionists like Dummett and Prawitz need not worry about Carnap's problem.[19] A little appendix presents a little-known result by James W. Garson, to the effect that, given certain assumptions, basic rules determine the referent of the logical operators—i.e. they are categorical—only if they are conservative.

6.5.1 Carnap's problem and Raatikainen's diagnosis

Consider the language of classical propositional logic (henceforth, **CPL**), call it L, with its set of well-formed formulae WFF. Let (1) be the standard semantics for **CPL**, where V_0 is the set of of admissible assignments of Boolean values to propositional letters, and V is the set of valuations induced by the recursive clauses for the connectives. Let (2) be a semantics just like (1), but whose set of admissible valuations is $V \cup \{v^*\}$, where, for every $A \in WFF$, $v^*(A) = 1$. It is easily shown that (1) and (2) yield the same consequence relation, that is, $\Gamma \models_V A$ iff $\Gamma \models_{V\cup\{v^*\}} A$.[20] For assume $\Gamma \models_V A$. Since $v^*(A) = 1$ for any $A \in WFF$, v^* provides no counterexample. Hence, $\Gamma \models_{V\cup\{v^*\}} A$. Now assume $\Gamma \not\models_V A$. Then, there exists a valuation $v \in V$ such that $v(B) = 1$ for every $B \in \Gamma$, and $v(A) = 0$. Since $v \in V \cup \{v^*\}$, any countermodel in V is in the extended set. Therefore, $\Gamma \not\models_{V^*} A$, where $V^* := V \cup \{v^*\}$. It follows that any adequate formalization $\vdash_{\mathbf{CPL}}$ of **CPL** is sound and complete with respect to $\models_V$ if and only if it is sound and complete with respect to $\models_{V^*}$. Yet on v^* the satisfaction clause for negation fails massively: there is a valuation $v \in V \cup \{v^*\}$, namely v^*, such that $v(A) = v(\neg A) = 1$. Similarly, it is possible to define a valuation $v^@$ such that it can be shown that $\Gamma \models_V A$ iff $\Gamma \models_{V\cup\{v^@\}} A$, where $v^@(A \vee \neg A) = 1$ and $v^@(A) = v(\neg A) = 0$.[21] But surely, a disjunction can't be true, if both of its disjuncts are false.

On the assumption that it is part of the meaning of negation and disjunction that, respectively, A is true (false) if and only if $\neg A$ is false (true), and that a true disjunction must have a true disjunct, there is a precise sense in which "the standard rules [of CPL] fail to capture an important aspect of the intended meaning[s]

[19]The contents of §§ 6.5.1-2 and of part of § 6.5.3 are drawn from Murzi and Hjortland (2009) and Murzi (2010a).

[20]The subscripts indicate the set of admissible valuations quantified over in the model-theoretic consequence relations.

[21]See Carnap (1943, Chapter C) and Smiley (1996, pp. 7–8).

of [negation and disjunction]" (Rumfitt, 1997, p. 224): for all the rules tell us, A and $\neg A$ may have the same truth-value, and a true disjunction may have no true disjunct.

One might object that the problem only arises because we are allowed to quantify over non-normal valuations and that these valuations are inadmissible, in some sense to be specified. This reply misses the point, however. Carnap's v^* and $v^@$ are only inadmissible in that they violate the recursive satisfaction clauses for negation and disjunction:

(NEG) $v(\neg A) = 1$ iff $v(A) = 0$;

(DISJ) $v(A \vee B) = 1$ iff either $v(A) = 1$ or $v(B) = 1$.

But, if meanings are to be determined by the inference rules, and if meanings are truth-conditions, logical inferentialists can't legitimately appeal to NEG and DISJ, on pain of invoking a previous knowledge of the meanings they are trying to capture.

Raatikainen considers three different replies to the problem. First, he writes,

> a radical formalist may just deny the very meaningfulness [...] of the notions of truth and falsehood [...] and insist that his use-theoretical approach is a genuine alternative to the truth-conditional approach and that it would beg the question to appeal to [the standard recursion clauses for negation and disjunction] against it. (Raatikainen, 2008, p. 285)

He goes on to argue that no "contemporary adherent of [logical inferentialism] accepts such a radical formalism, certainly not intuitionists such as Dummett, Prawitz and their followers" (p. 285).

Second, he sketches a possible deflationist response, according to which all we need to know about truth and falsity is exhausted by the equivalences:

(T1) A is true $\Leftrightarrow A$;

(T2) A is false $\Leftrightarrow \neg A$.[22]

A little logic suffices for deriving, from these two equivalences alone, the desired truth-conditional properties of negation and disjunction—see Raatikainen (2008, p. 285). But, Raatikainen argues, the problem has just been temporarily removed, since supplementing a natural deduction proof-system for **CPL** with T1 and T2

[22] Of course, given a sufficiently strong background logic, something would have to be done to ward off paradoxes.

doesn't prevent overlaps between truth and falsity, i.e. that there be a glutty (relational) valuation v^g which, for every $A \in WFF$, $v^g<A, 1>$ and $v^g<A, 0>$.

Finally, Raatikainen briefly considers what he takes to be "the view of Dummett, Prawitz and their followers", namely "that there is a sort of match between the proof-theoretical meaning-giving rules of inference and semantical notions of truth and falsity (possibly understood [...] in terms of provability)" (p. 285). He claims that "for this kind of view, Carnap's problem seems to pose a real challenge" (*Ibid.*).

6.5.2 The intuitionist's response

Raatikainen's dismissal of Dummett's and Prawitz's view is too quick. One does not need to be a radical formalist to "deny the very meaningfulness of the notions of truth and falsehood" in play in Carnap's argument. That the notions of truth-in-a-model and falsity-in-a-model are not relevant for determining the meaning of the logical connectives is precisely one of the key elements of Dummett and Prawitz's critique of realist notions of truth. Thus Prawitz:

> Michael Dummett is one of the earliest and strongest critics of the idea that meaning could be fruitfully be approached via model theory, the objection being that the concept of meaning arrived at by model theory is not easily connected with our speech behaviour so as to elucidate the phenomenon of language. (Prawitz, 2006, p. 507)

One might object that Carnap's argument may be run within some intuitionistically acceptable model theory, such as, say, Kripke's semantics for intuitionistic logic, or the Beth trees.[23] But this would not do. Dummett not only rejects classical model-theory. He also argues at length against Kripke semantics and Beth trees as a means of specifying the meanings of the intuitionistic connectives. He writes:

> [Beth trees] are not to be thought of as giving the full picture of the way in which the intuitionistic logical constants are given meaning: that can only be done directly in terms of the notion of a construction and of a construction's being recognized as a proof of a statement. (Dummett, 2000, p. 287)

Within an intuitionistic framework, truth is identified with the existence of a proof: the notion of a proof for atomic sentences is taken as primitive; proofs for

[23]See Dummett (2000, pp. 137-42; 186-203).

statements involving logical connectives are assumed to be reducible to *canonical* proofs—roughly, proofs whose last step is an introduction rule. The content determined by the inference rules is given by the so-called BHK clauses, specifying the proof-conditions for complex statements. The clauses for '$\vee$', '$\rightarrow$' and '$\bot$' are as follows ($\neg A$ is defined as $A \rightarrow \bot$):

(DISJ$_I$) A proof of $A \vee B$ is given by presenting either a proof of A or a proof of B.

(IF$_I$) A proof of a $A \rightarrow B$ is a construction that allows us to convert any proof of A into a proof of B.

(BOT$_I$) $\bot$ has no proof.

In this framework, Carnap's original problem doesn't arise. Recall, the argument targeted the claim that the standard inference rules of **CPL** determine the truth-conditions of complex statements. But there are two crucial differences here: the inference rules are those of *intuitionistic logic*, and the notion of truth has been replaced by that of *proof*. The right question to ask, then, is whether there can be a Carnap-like problem for BHK semantics, i.e. whether the intuitionistic rules determine the proof-conditional contents expressed by the BHK clauses.

6.5.3 A Carnap-like problem for BHK semantics?

Presumably, a proof-theoretic version of Carnap's valuation v^* is a possible situation where every sentence of the language has a proof, and a proof-theoretic version of $v^@$ is a possible situation where $A \vee \neg A$ is provable, but A and $\neg A$ aren't.[24] On this assumption, it would look like a variant of Carnap's problem could surface again. For, it would seem, the existence of a possible situation in which both A and $\neg A$ are provable doesn't affect the validity the intuitionistic rules: the rules are still valid, in the sense that the provability of their premises still guarantees the provability of their conclusions.[25] Similarly for disjunction: the provability or otherwise of A and $\neg A$ does not seem to affect the validity of the inference from A, or $\neg A$, to $A \vee \neg A$, nor does it seems to affect the inference from $A \vee B$ to whatever follows from both A and B.

[24]I am using the term 'proof' in a rather broad sense: I mean by 'proof' whatever notion intuitionists are willing to take as the key semantic concept of their meaning theory.

[25]The rules are also still valid in Dummett and Prawitz's sense—roughly, an argument ending with an introduction rule is valid provided that its subarguments are valid; an argument whose last step is an elimination rule may be accepted provided that it can be reduced to introduction form—see Prawitz (1973), Prawitz (2006), Dummett (1991b, pp. 252-6), and Appendix E below.

This alleged problem falters on closer inspection, however. From an inferentialist standpoint, negation is standardly not defined by its introduction and elimination rules alone. Given that both rules essentially mention absurdity, one should also consider the introduction and elimination rules for $\bot$. Which are they? There are two main inferentialist accounts. On the one hand, Dag Prawitz suggests that $\bot$ be defined by the *empty* introduction rule. That is, in his view, there is *no* canonical way of introducing $\bot$. He writes:

> the introduction rule for $\bot$ is empty, i.e. it is the rule that says that there is no introduction whose conclusion is $\bot$. (Prawitz, 2005, p. 685)

The rule can be shown to be in harmony with *ex falso quodlibet*:

$$(EFQ)\ \frac{\bot}{A},$$

where A is atomic.[26]

On the other hand, Dummett has claimed that $\bot$ should rather be defined by the following infinitary rule of $\bot$-introduction

$$(\bot I_D)\ \frac{P_1 \qquad P_2 \qquad P_3\ \dots}{\bot},$$

where the P_n are all the atoms of the language, which Dummett takes to be jointly inconsistent (see Dummett, 1991b, pp. 295-6). The idea is to specify canonical grounds for $\bot$ that can never obtain: no rich enough language will allow for a possibility in which all atoms, including basic contraries such as "This table is all red" and "This table is all white", can be proved. The rule is evidently harmonious with EFQ: one can derive from an assertion of $\bot$ precisely what was required for asserting $\bot$ in the first place. Armed with these definitions, let us now ask ourselves what Prawitz's and Dummett's rules for $\bot$ tell us.

Now recall the Carnap argument for negation, that for all its I- and E-rules tell us, A and $\neg A$ could both be provable. This argument is too quick. For any situation in which both A and $\neg A$ are provable is a situation in which there is a proof of both A and $A \to \bot$, from which we can conclude that there is a proof of $\bot$. But, on the foregoing assumptions, this cannot be. If introduction rules determine canonical grounds, Prawitz's *empty* rule of $\bot$-introduction says that there are no canonical grounds for $\bot$. If the Fundamental Assumption holds, though, it follows from this that there can't be non-canonical grounds for $\bot$ either. That is, in an

[26]See Prawitz (1973, p. 243), Read (2000, p. 139), and Negri and von Plato (2001, p. 8). The restriction on atomic formulae is a mere matter of convenience. An induction proof allows one to infer *any* well-formed formula from $\bot$.

intuitionistic framework, a proof-theoretic analogue of Carnap's valuation v^*, viz. a possible situation in which every sentence has a correct argument, *is* ruled out by the rules for negation.[27] The problem does not arise for intuitionistic disjunction either, for similar reasons. We are now asked to consider the existence of a possible situation where there is a correct argument for $A \vee \neg A$, while there are no correct arguments for A and $\neg A$. But this cannot be. For one thing, intuitionistic logic has the *disjunction property*: if $A \vee B$ is provable, so must be either A or B.[28] For another, by the Fundamental Assumption, if there is a correct argument for $A \vee B$, then there is an argument for it ending with one step of disjunction introduction, which means that either there is a correct argument for A, or there is a correct argument for B.[29]

6.5.4 Incurvati and Smith's objections

The same result follows on Dummett's account of $\bot$, although this is more controversial. Dummett's account has been often criticized on the grounds that we have no guarantee that all the atoms of the language form an inconsistent set. Thus, for instance, Michael Hand writes that

> [Dummett's] rules cannot even prevent $\bot$ from meaning something that might be true: the rules do not preclude an assignment that assigns truth to all atoms including $\bot$. (Hand, 1999, p. 190)

If this is correct, Dummett's introduction rule for $\bot$ is of no help to the inferentialist. Luca Incurvati and Peter Smith have recently made the point:

> [Dummett's rule] is compatible with a situation in which there is a proof of P and a proof of $\neg P$: it will just be a situation in which there is a proof of P_1, and a proof of P_2, and so on. (Incurvati and Smith, 2010, p. 6)

This argument, however, rests on a misunderstanding of Dummett's rule. As Neil Tennant observes, "logic has to allow for for languages whose sets of atomic sentences may or may not be jointly consistent" (Tennant, 1999, p. 215). That is, logic does not, and should not, know whether the set of atoms forms an

[27]See also Murzi and Hjortland (2009). Notice that this is not to say that one cannot *derive* $\bot$ in intuitionistic logic, but rather that any such derivation isn't canonical; see also *supra*, fn. 7.

[28]See Troelstra and Van Dalen (1988, p. 139).

[29]See Troelstra and Van Dalen (1988, p. 139). It is easy to see that analogues of Raatikainen's glutty valuation are ruled out too.

inconsistent set. But this means, Tennant points out, that Dummett's introduction rule for $\perp$

> has to be understood as potentially open ended [...]: namely, [...] it should hold whatever extension of the language might be undertaken. And we must allow that some of those extensions could involve the inconsistency of all the sets of atoms. Now this does not just mean that, in order to derive $\perp$ in the existing language, it suffices to derive each atomic sentence of the language. Rather, it means that in order to derive $\perp$ one has to be in a position to derive *any* atomic sentence of *any extension* of the language. (Tennant, 1999, p. 215)

On Tennant's interpretation, Dummett's rule is fully *schematic*: it applies to *all possible extensions of the language*. It tells us that the conditions for introducing $\perp$ will never be met, not even in a situation in which all atoms are assertible. For in any such situation, we will not be in a position to assert any atom of any possible extension of the language.

One might wonder whether a rule formulated in a language L can really be about sentences *outside* that language. However, it seems to me that it is implicit in our understanding of a schema that, if it is valid, it must apply to all the possible extensions of the language. For instance, we certainly do not need to check whether, say, $\wedge$-I is still valid, when we introduce a new expression in the language (see McGee, 1997, p. 58).

It follows that, even if one accepts Dummett's rules for $\perp$, a situation in which every atom is provable need not be a situation in which both A and $\neg A$ are also provable. More needs to be done to show that Carnap's problem poses "a real challenge" for the kind of view advocated by Dummett, Prawitz, and their followers.

It might still be objected that the foregoing defence entirely rests on the assumption that the rules for $\perp$ tell us that $\perp$ is always false, in both intuitionistic and classical settings. This assumption may be challenged on at least two counts. First, one might retort that rules by themselves don't *say* anything: for instance, the empty rule of $\perp$ introduction does not *say* that there are no canonical proofs of $\perp$, nor does it *tell us* that there are no necessary and sufficient conditions for asserting $\perp$. Second, it might be argued that $\perp$ has no content, and can't therefore be susceptible of being true or false.

Luca Incurvati and Peter Smith have voiced the first concern in a recent reply to Murzi and Hjortland (2009):

> now one might grant that if the null (non-existent) rule of $\bot$-introduction says that there is no canonical proof of $\bot$, then the rule is incompatible with a situation in which there is a proof of P and a proof of $\neg P$. But the crucial question is precisely whether the simple non-existence of an inference rule can convey so much. (Incurvati and Smith, 2010, p. 5)

Smith and Incurvati substantiate their worry by pointing to an argument by Tennant (1999, p. 216) to the effect that, in contrast with all the proof-theoretic justifications of the elimination rules of the standard intuitionistic connectives, the proof-theoretic justification of EFQ must itself rely on EFQ in the metalanguage.

This is puzzling, however. For even if proof-theoretic justifications of EFQ are bound to be circular (a claim for which neither Incurvati and Smith nor Tennant provide a proof), this fact is orthogonal to the question whether Prawitz's empty introduction rule effectively tells us that there are no canonical grounds for introducing $\bot$. An analogy might help clarifying this point. Consider our connective $\circ$:

$$\circ\text{-I}\,\frac{A \qquad B}{A \circ B} \qquad \circ\text{-E}\,\frac{A \circ B}{A}\,.$$

The fact that the elimination rule for $\circ$ is not in harmony with the corresponding introduction rule does not seem to prevent $\circ$'s introduction rule from telling us that $A \circ B$ must be true, if A and B are also true. At any rate, it is unclear whether rule-circularity, i.e. the use of a rule R in a justification of R, is itself problematic. As Dummett points out,

> [it] is not the ordinary gross circularity that consists of including the conclusion to be reached among the initial premises of the argument [...] but only that at least one of the inferential steps in the argument must be taken in accordance with the law. (Dummett, 1991b, p. 202)

The point is a familiar one: unlike grossly circular arguments, rule-circular arguments can nevertheless be interesting, since they can fail. For instance, the rules for Prior's `tonk` cannot be proof-theoretically justified in the way harmonious introduction and elimination rules are, even if `tonk` is admitted in our metalanguage (see e.g. Tennant, 2005b).

6.5.5 Absurdity as a logical punctuation sign

Perhaps more convincingly, one might question the assumption that $\bot$ has a content in the first place. For what does $\bot$ *mean*, more exactly? Tennant (1997) and

Ian Rumfitt (2000) have recently suggested that $\perp$ is not a propositional constant, and should rather be interpreted as a logical punctuation sign. Here is Tennant:

> an occurrence of '$\perp$' is appropriate only within a proof [...] as a kind of structural punctuation mark. It tells us where a story being spun out gets tied up in a particular kind of knot—the knot of a patent absurdity, or self contradiction. (Tennant, 1999, p. 204)

Similarly, Rumfitt writes that '$\perp$' "marks the point where the supposition [...] has been shown to lead to a logical dead end, and is thus discharged, prior to an assertion of its negation" (Rumfitt, 2000, pp. 793-4). Tennant's main argument for interpreting $\perp$ as a punctuation sign is that it can't be identified with any specific asburdity, on pain of making the meaning of $\neg$ "provincial" to the discourse to which that absurdity belongs. But, Tennant writes, "absurdity is much more cosmopolitan a notion than the the discourse-specific model would make it" (Tennant, 1999, p. 203). Rumfitt concurs that $\perp$ cannot be identified with any specific absurdity, on the grounds that logic does not know that, say, '$0 = 1$' is actually false: for all logic knows, '$0 = 1$' could be true—say in a model in which both '0' and '1' denote 1 (see Rumfitt, 2000, p. 793).

While I think these arguments can be ultimately convincing, I also do not think that to treat $\perp$ as a logical punctuation sign can help Incurvati and Smith in the present context. To begin with, even if $\perp$ is interpreted as a logical punctuation sign, it is still the case that, whenever $\perp$ follows from an application of a valid rule, the premises of this rule cannot all be true, if the rule is to be truth-preserving. Thus, even on Tennant's and Rumfitt's interpretation of $\perp$, the rule of negation elimination still guarantees that A and $\neg A$ can't both be true, and our new rule of disjunction elimination still guarantees that disjunctions with only false disjuncts must be false. Second, Tennant's and Rumfitt's main arguments for treating $\perp$ as a punctuation sign are off target in the present context. For, after all, inferentialist accounts of $\perp$ *do* respect their own requirement that $\perp$ should not be identified with any specific language-dependent absurdity: the point of these accounts is precisely to provide an inferential and *language-independent* definition of absurdity.

To recapitulate: I agree with Raatikainen that "Carnap's forgotten result" (Raatikainen, 2008, p. 6) deserves attention. However, it does not seem that the problem raises a challenge for intuitionists like Dummett and Prawitz, even when the argument is run within a proof-theoretic framework. Intuitionists can block the argument by identifying truth with provability, and by defining the notion of a canonical proof by proof-theoretical means. Incurvati and Smith's objection that the foregoing defense rests on an ill-conceived account of absurdity, Prawitz's account, falters on closer inspection. The rules for negation rule out the proof-theoretic analogous of Carnap's non-normal valuations, irrespective of how absurdity is accounted for, or at least so I have argued. It remains to be seen, however, whether *classical logicians* can adequately solve Carnap's categoricity problem. We shall deal with this issue in Chapter 7, when we will consider the more general question whether logical inferentialism is compatible with classical logic.

6.6 Conclusions

In the first part of this chapter, we have introduced several local proof-theoretic constraints on rules, and we have examined their relations with the global constraints of separability and conservativeness. In the second part, we have presented one more proof-theoretic argument for classical logic: the Argument from Separability. I have argued that the argument rests on two main assumptions: that logical holism—the view that the meanings of the logical vocabulary are all interdependent—is false, and that basic rules are complete. Both assumptions are problematic, as we have seen. The first obviously begs the question against the logical holist. The second clashes with the incompleteness of higher-order logics. I have also argued, however, that neither problem shows that classical logicians with inferentialist leanings can justifiably ignore the inferentialist's challenge. To begin with, pending further arguments for logical holism, classical logicians are not in a position to rule out the molecularist view of the meaning of the logical vocabulary advocated by orthodox inferentialists. Insofar as we *could* understand logical expressions independently of one another, it would seem that any adequate formalization of logic should allow for this possibility. Secondly, inferentialists may still claim that I-rules are complete in the weak sense specified by the Fundamental Assumption, and decide on a case-by-case basis whether rules should also be complete in the sense required by Milne's and Bendall's arguments. As we have seen, the Fundamental Assumption is weak enough to allow for the non-conservativeness of higher-order logics, but strong enough to justify the requirement of harmony for higher-order logics. It follows that insofar as we have reasons for thinking that a certain proof-theoretic relation of logical consequence $\vdash$ is *complete*, inferentialists should provide a separable axiomatization of $\vdash$. On the plausible assumption that classical logic *is* complete, this suggests that classical logicians must face the inferentialist's challenge, so to speak, head on.

Appendix: categoricity and conservativeness

In a relatively little-known paper, James W. Garson proves a relatively little-known result, to the effect that categorical systems, i.e. systems whose basic rules allow us to derive the satisfaction clauses of the logical operators, are separable. But, as we have seen, higher-order logics are not separable. Does it follow that the rules for the higher-order quantifiers cannot determine their satisfaction clauses? It does not. The aim of this appendix is to present Garson's result, and to show that it need not worry inferentialists who are willing to countenance higher-order logics.

V-validity and natural semantics

Garson assumes that inference rules preserve *validity*, as opposed to truth. He writes:

> preservation of truth is a bad choice for for understanding a rule's meaning, because it incorporates a covert prejudice against non-standard truth-conditions. It focuses on the behavior of individual valuations, so it automatically eliminates clauses that depend on truth behaviour over a whole set of valuations (possible worlds). Furthermore, there are important rules (such as Necessitation [...]) which do not preserve truth [...]. The more general way to characterize what a set of rules expresses employs preservation of *validity*. Since validity of a rule is only defined for a *set* of valuations, it follows that semantical conditions should be properties of sets of valuations as well. (Garson, 2001, p. 117)

This is not the place to assess these claims. Hence, I will take them on board for the argument's sake, without further ado.

Next, some definitions. Let **V** be a set of valuations. We can then define the following notions:[30]

Definition 16. *(V-validity) An argument* $\Gamma \vdash A$ *is* V-valid *if and only if, for every valuation* $v \in V$*, if, for every* $\gamma \in \Gamma$*,* $v(\gamma) = 1$*, then* $v(A) = 1$.

Definition 17. *(C-validity) An argument* Π *is* V-valid, *for some condition C on sets of valuations, if and only if* Π *is V-valid for every set of valuations that obeys C.*

[30]Here I will be following Garson's own excellent presentation. Garson (See 2001, pp. 118 and ff.).

Let us say that a set of valuations **V** is a *model*. Then, **V** is a model of a proof-system S if and only if S's rules preserve **V**-validity.

Definition 18. *(Standard model) A model* **V** *is* standard *for a logical operator \$ if and only if* **V** *satisfies \$'s intended truth-conditions.*

Definition 19. *(Semantics) A* semantics *for proof-system S is a condition on models (i.e. sets of valuations) specifying how the logical vocabulary is to be interpreted—i.e., for every logical operator \$ in S's language, it provides a recursive definition of \$'s truth-conditions.*

Thus,

$$(\wedge)\ \ v(A \wedge B) = 1 \text{ if and only if } v(A) = 1 \text{ and } v(B) = 1$$

$$(\vee)\ \ v(A \vee B) = 1 \text{ if and only if either } v(A) = 1 \text{ or } v(B) = 1$$

is a semantics for a proof-system whose only logical operators are $\wedge$ and $\vee$.

Now, as Garson points out, every proof-system S expresses a condition on *some* set of valuations **V**, viz. that **V** is a model of S: the S-rules preserve **V**-validity. The problem, however, is that not every condition corresponds to a semantics—this was, in essence, Carnap's point. Let us say that, for every logical operator \$, the condition C expressed by the set of the \$-rules and \$'s valuation clause (\$) are *equivalent* if, for every model (i.e. set of valuations) **V**, **V** obeys C if and only if it obeys (\$). At very long last, we can now introduce Garson's definitions of a *natural semantics* and of a *natural system*:

Definition 20. *(Natural semantics for systems with one logical operator) (\$) is a* natural semantics *for a proof-system S if and only if the condition expressed by S is equivalent to (\$).*

Definition 21. *(Natural semantics for systems with more than one logical operator) Let S be a system obtained by adding to a standard structural base* **B** *the rules for each logical operator \$. Then, a* natural semantics *for S is any sound semantics obtained by conjoining the natural semantics for each logical operator \$.*

Definition 22. *(Natural system) S is a* natural system *if and only if it has a natural semantics.*

We are now almost ready to prove Garson's result that every natural system is conservative.

Garson's result

Let $||S||$ be the semantics for S, and let us further assume that $S \supset \mathbf{B}$ is sound and complete with respect to its natural semantics $||S||$. That is:

$$\Gamma \vdash_S A \text{ if and only if } \Gamma \vdash A \text{ is } ||S||\text{-valid.}$$

Then, Garson proves that every natural system is conservative—i.e. for every logical operator \$, the addition of the \$-rules to the rules for each of the remaining logical operators yields a conservative extension.

Theorem 4. *(Garson) Every complete natural system is conservative.*

> Let S be a natural system, and let S' be the rules for some sublanguage of S. We must show that S is a conservative extension of S', i.e. that S does not increase the stock of theorems generated by S. So assume $\Gamma \vdash A$ is an argument provable in S that contains only connectives of S'. We must show that it is provable in S'. Since S is natural, we know that S is sound for a semantics $||S||$ which consists of a recursive truth clause (\$) for each of its connectives \$. By the soundness of S, $\Gamma \vdash A$ is $||S||$- valid. The validity of $\Gamma \vdash A$ depends only upon the connectives it contains, and so $\Gamma \vdash A$ is $||S||$-valid iff $\Gamma \vdash A$ is $||S'||$-valid, where $||S'||$ is the result of deleting clauses from $||S||$ that do not mention connectives in S'. Since the completeness of S insures completeness of S', it follows that $\Gamma \vdash A$ is provable in S'. (Garson, 2001, p. 131)

This is a *very* interesting result. To the best of my knowledge, it establishes the only known link between the notions of conservativeness and categoricity—i.e. between the two senses of the Determination Thesis we made explicit in §§ 4.1.2-3. It should be clear, however, that Garson's theorem cannot serve as a premise of an argument against the inferentialist claim that the rules of the higher-order quantifiers determine their satisfaction clauses—or, in Garson's terminology, their recursive truth-clauses. For all that follows from Garson's theorem and the non-conservativeness of higher-order logics is that *either* the rules for the higher-order quantifiers fail to determine their satisfaction clauses, *or* higher-order logics are incomplete. But since higher-order logics *are* incomplete, Garson's result gives no reason to also believe the first disjunct of this true disjunction.

Chapter 7

Classical inferentialism

Can classical logic be regimented in a proof-theoretically acceptable way? Leblanc's result shows that, insofar as classical logicians (i) adopt a standard framework for formalizing logic, and (ii) take classical negation to be partially defined by either Double Negation Elimination or classical *reductio*, the answer to this question cannot but be negative. Classical inferentialists, however, can drop either assumption, and thereby provide harmonious, separable, and conservative formalizations of classical logic. In the first part of this chapter, I introduce and critically assess two non-standard formalizations of classical logic: the so-called *multiple conclusions* formalizations, on the one hand, in which arguments are allowed to have more than one conclusion; and the so-called *bilateral* formalizations, on the other, where the meanings of the logical expressions are given by specifying both assertibility and *deniability* conditions for complex statements.[1] As we will see, both multiple conclusions and bilateralist frameworks satisfy not only categoricity, but also harmony and separability. It remains to be seen, though, whether their non-standard features are ultimately acceptable. In the second part of the chapter, I present a novel formalization of classical propositional logic, in a *standard*, i.e. single-conclusion and assertion-based, natural deduction framework. I prove a normalization theorem for this proposed formalization which, unlike Prawitz's original theorem, entails the key proof-theoretic property of separability. Moreover, I show that, given certain semantic assumptions, natural deduction rules can be seen to determine the satisfaction clauses of the classical connectives. The unifying theme of the chapter is that we *can* be in harmony with classical logic, although harmony can only be achieved by—so to speak—smuggling in

[1] For multiple-conclusion formalizations of classical logic, see Boričić (1985) and Read (2000). For an excellent study of multiple-conclusion logics, see Shoesmith and Smiley (1978). For bilateral formalizations of logic, see Smiley (1996), Rumfitt (1997), Rumfitt (2000), and Humberstone (2000).

classicality in the structural assumptions of the logic.

The structure of the chapter is as follows. Section 7.1 and Section 7.2 respectively introduce, and critically expound, multiple-conclusions and bilateral formalizations of classical logic. Section 7.3 discusses some objections to bilateralism, both old and new. Section 7.4 presents our proposed unilateral (i.e. non-bilateralist) and single-conclusion formalization of classical logic. Section 7.5 offers some concluding remarks.

7.1 Multiple conclusions

It is now time to investigate whether classical inferentialists can address the inferentialist's challenge head on. The question, as we shall see, hinges on the *structural features* of admissible systems, i.e. features which do not directly concern the rules for one or another logical operator, but which rather apply to logical systems *as a whole*. The first structural feature we shall focus on is whether logical arguments should allow multiple *conclusions*, in addition to multiple premises. In everyday contexts, we typically argue from one *set* of premises to a *single* conclusion. For instance, we may infer that *there is a chance that there will be disruptions in the Eurostar service next Sunday* from our assumptions that *it will be very cold on Sunday* and that *for every t, if it is very cold at t, there is a chance that there will be disruptions in the Eurostar service at t*. We do not give logical arguments for *lists* of sentences. Or do we?

7.1.1 Sequent calculi

In his doctoral dissertation, Gerhard Gentzen famously introduced *two* different logical calculi: natural deduction and sequent calculus. Gentzen himself explains why he had to do so (see Gentzen, 1969, pp 68-9). Although he was able to prove a normalization theorem for his natural deduction formalization of intuitionistic logic (see von Plato, 2008), he could not prove normalization for natural deduction formalizations of classical logic.[2] This is why he introduced sequent calculus formalizations of both intuitionistic and classical logic, which he respectively called LJ and LK. He could then prove *Cut Elimination*, his *Hauptstaz*, the sequent calculus analogue of a normalization theorem, for both these logics. He writes:

> In order to prove the Hauptsatz in a convenient form, I had to provide

[2]The result was first proved in Prawitz (1965).

> a logical calculus especially suited to that purpose. (Gentzen, 1969, pp. 68-9).

Gentzen's 'main theorem', the Hauptsatz, states that the rule of cut

$$\frac{\Gamma \vdash A, \Delta \qquad A, \Gamma' \vdash \Delta'}{\Gamma, \Gamma' \vdash \Delta\Delta'} \; Cut$$

is always dispensable: as he puts it, "every LJ- or LK-derivation can be transformed into another LJ- or LK-derivation with the same end-sequent, in which no cuts occur" (Gentzen, 1969, p. 88). Gentzen thus glosses his own theorem:

> The *Hauptsatz* says that every purely logical proof can be reduced to a definite, though not unique, normal form. Perhaps we may express the essential properties of such a normal form by saying: it is not roundabout. No concepts enter into the proof other than those contained in its final result, and their use was therefore essential to the achievement of that result. (Gentzen, 1934, p. 69)

Gentzen's sequent calculus is much more than a convenient tool for proving the cut elimination theorem. It is a very useful tool for proof search. It is by far the most convenient tool for studying so-called sub-structural logics. And, closer to our present concerns, it may afford an inferentialistically acceptable formalization of classical logic, or at least so classical inferentialists such as Ian Hacking (1979) have long been arguing. Let us have closer look.

7.1.2 Sequent calculi and classical logic

We have seen that standard natural deduction formalizations of classical logic are either non-harmonious, or non separable, or both. But what about *sequent calculi* formalizations of classical logic? In Gentzen's sequent calculi, the rules for classical negation are as follows:

$$R\neg_c \frac{\Delta, A \vdash \Gamma}{\Delta \vdash \Gamma, \neg A} \qquad\qquad L\neg_c \frac{\Delta \vdash \Gamma, A}{\Delta, \neg A \vdash \Gamma}$$

where a *multiple-conclusions* sequent $\Gamma \vdash \Delta$ intuitively says that if everything in Γ is true, then at least one formula in Δ is also true. As for the rules for intuitionistic negation, they are *just like the classical rules*, except that the cardinality of the succedents is restricted to at most one formula:

$$R\neg_i \frac{\Delta, A \vdash}{\Delta \vdash \neg A} \qquad\qquad L\neg_i \frac{\Delta \vdash A}{\Delta, \neg A \vdash}$$

Both sets rules yield a conservative extension of the negation-free fragment of classical logic, and both rules satisfy Gentzen's cut elimination theorem. Thus, Roy Cook has recently argued that

> the sequent calculus rules, with multiple formulae allowed to the right of the inference arrow, provide a harmonious codification of classical logic. The requirement that logical rules be harmonious and/or conservativeness does not, therefore, weigh more in favour of intuitionistic logic as opposed to its classical rival. (Cook, 2005, p. 391)

But there is more.

First, it is well-known that sequent calculus formalizations of classical logic satisfy separability, both weak and strong. Here is, for instance, a derivation of Peirce's Law satisfying the separability requirement:

Example 10. *Peirce's Law:*

$$R\to\,\dfrac{L\to\,\dfrac{R\to\,\dfrac{\text{Weakening}\,\dfrac{A \vdash A}{A \vdash B, A}}{\vdash A \to B, A} \qquad A \vdash A}{A \to B) \to A \vdash A}}{((A \to B) \to A) \to A}$$

Second, as Carnap himself first pointed out, the categoricity problem does not arise within a multiple conclusions framework. Let us say that a sequent $\Gamma \vdash \Delta$ is *verified* by v if, whenever for every $\gamma \in \Gamma$, $v(\gamma) = 1$, for some $\delta \in \Delta$, $v(\delta) = 1$. Then, a sequent is *valid* if it is verified by every valuation; it is *invalid* if it is not valid. Now consider the following two sequents, both of which are provable in multiple conclusions formalizations of **CPL**:

(NC) $A, \neg A \vdash$;

($\vee$-E) $A \vee \neg A \vdash A, \neg A$.

If inferentialists accept NC and $\vee$-E, neither v^* nor $v^@$ are admissible: the former does not verify NC; the latter does not verify $\vee$-E.[3]

In short, *sequent calculi* formalizations of classical logic appear to tick all boxes: they are not obviously disharmonious, they satisfy the requirement of separability, and they the determine the meanings, i.e. the truth-conditions, of the classical logical operators. If intuitionists wish to argue that classical negation is proof-theoretically defective, they need to show that there is something amiss either with sequent calculi in general, or with the classical rules $R\neg_c$ and $L\neg_c$.

[3] See Shoesmith and Smiley (1978) for a proof that multiple conclusions formalizations of classical logic are indeed categorical.

It may be objected that Cook's conclusion only follows on the assumption that inferentialists can define a suitable notion of harmony for sequent calculi—such as, for instance, the satisfiability of a cut elimination theorem. However, it is well known since the publication of Boričić (1985) that multiple conclusions formalizations of classical logic *are also available within a natural deduction framework* (see also Read, 2000). Not only do these formalizations satisfy the requirements of separability and categoricity: they can also be shown to be strongly intrinsically harmonious, GE harmonious, and fully invertible.

Perhaps more to the point, intuitionist inferentialists may object that standard multiple-conclusions formalizations of classical logic do not respect the Fundamental Assumption, since the standard proof of the Law of Excluded Middle ends with a step of Contraction, not disjunction introduction:

Example 11. *The Law of Excluded Middle:*

$$\dfrac{\dfrac{\dfrac{\dfrac{[A]^{(1)}}{A \vee \neg A}\,\vee\text{-}I}{A \vee \neg A, \neg A}\,\neg\text{-}I, 1}{A \vee \neg A, A \vee \neg A}\,\vee\text{-}I}{A \vee \neg A}\,Contr$$

However, the objection assumes that the Fundamental Assumption may not be weakened to the claim that whenever we can assert a complex statement A, we could have introduced A by means of an argument whose last step taken in accordance with an operational rule is taken in accordance with one of the I-rules (or right rules) for its main logical operator. Moreover, it presupposes that the I-rules for $\vee$ are the correct rules for disjunction:[4] an assumption that we have already questioned in § 5.4.5 (on the grounds that they presuppose an understanding of disjunction), and to which we shall return in § 7.4.1 below.

Some inferentialists might perhaps insist that sequent calculi essentially represent deductive relations between *sets* of sentences, which in turn might suggest that they do not strictly speaking provide rules for correctly *using* the logical vocabulary. Ian Rumfitt makes the point:

> [sequent calculus] is of little help in the quest for specifications of sense for the connectives that encapsulate their classical deductive use. In the statements
>
> > {"If it is raining then it is not snowing", "It is raining"} entails {"It is not snowing"}

[4]The multiple conclusions I-rule for $\vee$ are the same as the standard ones. As for the E-rule, it allows us to infer A, B from $A \vee B$.

> and
>
> The empty set entails {"It is raining","It is not raining"}
>
> the sentences [...] are mentioned rather than used. We, however, are exploring the idea that a connective's sense consists in the way in which it is correctly *used* in deductions. (Rumfitt, 2000, p. 795)

Rumfitt's point is that, in *sequent calculi*, sentences are mentioned rather than used. Hence, he submits, *sequent calculi* do not vindicate the inferentialist thought that inference rules characterize the *use* of the logical vocabulary. Whatever the merits of this objection, though, the point is dialectically irrelevant. As we have just seen, there are *natural deduction* multiple conclusions formalizations of classical logic. Hence, multiple-conclusions logicians are not forced to resort to *sequent calculi* in order to meet the inferentialist's challenge.[5]

The foregoing considerations suggest that the crucial question when assessing multiple-conclusions logics is the natural one, viz. whether inferentialists can *make sense* of multiple conclusions. It is to this issue that we shall now turn.

7.1.3 Multiple conclusions and actual reasoning

More plausibly, single-conclusion logicians may question whether multiple-conclusions logics adequately represent our actual logical practice. Indeed, it is rather doubtful that we actually reason from sets of premises to *sets* of conclusions. Thus, Tennant writes that the multiple-conclusions logician's point is ultimately "not well-taken", because "sequents are supposed to represent acceptable arguments" but "in normal practice arguments take one from premisses to a single conclusion" (Tennant, 1997, p. 320). Rumfitt has recently developed the point. After having observed that "the rarity, to the point of extinction, of naturally occurring multiple-conclusion arguments has always been the reason why mainstream logicians have dismissed multiple-conclusion logic as little more than a curiosity", he goes on to argue that " attempts by enthusiasts to alleviate the embarrassment here have often ended up compounding it" (Rumfitt, 2008, p. 79). Rumfitt quotes the following passage from Shoesmith and Smiley:

> Perhaps the nearest one comes to [multiple-conclusions] is in proof by cases, where one argues "suppose $A_1 \ldots$ then B; ...; suppose $A_m \ldots$

[5]See Steinberger (2009a, pp. 192 and ff.) for an argument to the effect that, if Rumfitt's argument is correct, then both natural deduction and sequent calculi can be accused of merely mentioning formulae.

> then B; but $A_1 \vee \cdots \vee A_m$, so B". A diagrammatic representation of this argument exhibits the downward branching which we shall see is typical of formalised multiple-conclusion proofs [...]. But the ordinary proof by cases is at best a degenerate form of multiple-conclusion argument, for the different conclusions are all the same (in our example they are all instances of the same formula B) (Shoesmith and Smiley, 1978, p. 4-5)

Shoesmith and Smiley are here attempting to find instances of multiple-conclusions reasoning in our actual deductive practice. Proof by cases, or $\vee$-E, they argue, is one such example: we notice that C follows from both A and B, and we conclude C, C from $A \vee B$, discharging A and B. But, Shoesmith and Smiley themselves observe, this is a 'degenerate' form of multiple conclusion reasoning, since the multiple-conclusions are *just one*! Rumfitt sarcastically comments:

> I do not know how the word 'multiple' is used in Cambridge, but in the rest of the English-speaking world it is understood to mean 'more than one'. (Rumfitt, 2008, p. 79)

This is fair enough: certainly proof by cases provides no justification for taking multiple-conclusions logic seriously. However, the question whether multiple-conclusion logics faithfully represent our logical practice depends on one's *interpretation* of a sequent—a point that, as we shall see in a moment, Rumfitt (2008, pp. 79-80) himself concedes.

On the *standard interpretation* of multiple conclusions sequents (and rules, for that matter), commas on the left and commas on the right respectively read 'and' and 'or', and a sequent $\Gamma \vdash \Delta$ intuitively says that, if *all* the conclusions are true, then at least *one* conclusion must also be true.

The standard interpretation, though, clearly raises a *prima facie* issue of circularity: if commas are interpreted as meaning, intuitively, what $\wedge$ and $\vee$ mean, it would seem that we are now relying on a previous understanding of $\wedge$ and $\vee$ in a context in which we are precisely trying to *explain* their meaning! Or are we?

Commas in the antecedent are not too difficult to make sense of. When we read a book, we certainly do not take its author to be committed to the *disjunction* of all the sentences it contains. We rather take her to be committed to the truth of each of its sentences. A satisfactory interpretation of the commas in the succedents, on the other hand, is much harder to come by.[6] We cannot take them to intuitively

[6] See also *supra*, § 5.4.5.

express disjunction, on pain of assuming a prior understanding of what we are trying to account for (see e.g. Dummett, 1991b, p. 187), and, perhaps even more importantly, on pain of turning multiple-conclusions into *single*-conclusions.

Vann McGee forcefully raises the second point. Suppose I wish to reject the conclusion of a multiple-conclusions argument—say, an argument to the effect that "Brown is a bully, Berlusconi is innocent". How to make sense of my rejection? On the standard interpretation, I have rejected a disjunction, viz. *that either Brown is a bully or Berlusconi is innocent*. But this strongly suggests that the conclusion of a multiple-conclusions argument really is a *single* conclusion. McGee writes:

> Once we allow multiple conclusions, in what sense can we be said to accept the conclusion set, when we don't accept any of its members? The only sense I can make of this is that we accept the conclusion set by accepting the disjunction of its elements, and that's a matter of replacing a multiple-conclusion inference by a single-conclusion inference. (McGee, 2004, p. 286)

So how to interpret multiple conclusions sequents?

7.1.4 Restall's interpretation

Greg Restall (2005) offers the following thought: a multiple conclusions sequent may be read as saying that one cannot, at the same time, *assert* its premises and *deny*, or *reject*, its conclusions: if $\Gamma \vdash \Delta$, "then it is incoherent to assert all of Γ and deny all of Δ" (Restall, 2005, p. 10).[7] Restall claims that his suggested interpretation does not require a prior understanding of disjunction, but assumes that denial be treated as a primitive speech act. The idea is that the rules of a classical sequent calculus codify our commitments in terms of assertion and denial. Consider, for instance, the sequent calculi rules for $\wedge$:

Example 12. *(L$\wedge$) and (R$\wedge$)*

$$\frac{\Gamma, A_i \vdash \Delta}{\Gamma, A_0 \wedge A_1 \vdash \Delta}\,(\text{L}\wedge)\ (i = 0,1) \qquad\qquad \frac{\Gamma \vdash A, \Delta \qquad \Sigma \vdash B, \Pi}{\Gamma, \Sigma \vdash A \wedge B, \Delta, \Pi}\,(\text{R}\wedge).$$

The left rule tells us that, if one may not assert A_i and deny everything in Δ, then one may not assert $A_0 \wedge A_1$ and deny everything in Δ. Likewise, the right rule tells

[7]Following Priest (2006a, p. 103 and ff.), I take assertion and denial to be a *speech acts* and I take acceptance and rejection to be a *mental attitudes*. In what follows, I will mainly focus on denial, for the sake of simplicity, although many writers, e.g. Smiley (1996) and Rumfitt (2008), talk of rejection instead.

that if we cannot assert everything in Γ and deny A and everything in Δ and if we cannot assert everything in Σ and deny B and everything in Π, then we cannot assert everything in Γ and Σ and deny $A \wedge B$ and everything in Δ and Π.

Rumfitt concedes that "seen from this angle [...] there seems to be no good reason to privilege multiple acceptances over multiple rejections", and submits that this is "the best case one can make for multiple-conclusion logic" (Rumfitt, 2008, pp. 79-80). I agree. He adds, however, that "here [...] the best is not good enough" (p. 80).

His main objection to Restall's interpretation of multiple-conclusions is a *normative* one: Restall's proposed interpretation, he argues, "does not capture anything like the full force of single-conclusion consequence" (Rumfitt, 2008, p. 80). The problem is that, on Restall's interpretation of a sequent, a speaker is not obliged to *accept* the conclusion of an argument $\Gamma \vdash A, B$, even if she accepts everything in Γ and she knows that A, B follows from Γ. On Restall's interpretation, the sequent rules merely tells us that, in such circumstances, the speaker may not deny, or reject, A, B. As Rumfitt puts it:

> A thinker who accepts all the statements in a set X, who knows that a set Y is a multiple-conclusion consequence of set X, but who refuses to accept any statement in Y need not be making any mistake. (Rumfitt, 2008, p. 80)

The multiple-conclusions logician may insist that the speaker "will be making a mistake if [s]he refuses to accept the claim that *some* member of Y is true" (*Ibid.*). However, Rumfitt objects,

> that point is grist to the mill of sceptics about multiple-conclusion logic. Yet again, they will say, we can only understand an instance of multiple-conclusion consequence as an instance of single-conclusion consequence in which the conclusion is a disjunctive or existentially quantified claim. (Rumfitt, 2008, p. 80)

Rumfitt's point seems to be this. Suppose I know that Y follows from X, and I know everything in X. Then, the multiple-conclusions logician may stress that I will make a mistake as to the facts if I do not accept the claim *that some member of Y is true*. This claim, however, is a *single-conclusion* consequence of X. Hence, we have done nothing to make sense of the normative force of *multiple-conclusions*.

The crux is that multiple-conclusions sequents do not wear on their face which conclusion ought to be asserted. Thus, the problem arises as to what rational subjects must accept, when they accept the premises of a multiple-conclusions argument but they refrain from accepting any of the conclusions. On both the standard and Restall's interpretation, multiple-conclusions logicians must accept either the *disjunction* of the conclusions or that *some* conclusion holds. In either case, as McGee puts it, "we have done nothing to make sense of the normative force of *multiple-conclusions*" (McGee, 2004, p. 286). Let us see, then, whether classical logic can meet the requirements of harmony and separability without having to resort to multiple-conclusions arguments.

7.2 Bilateralism

Standard approaches to sense are driven by the idea that there is just one fundamental speech act: assertion. Logical inferentialism, as we have presenting it so far, is no exception: the sense of the logical connectives is supposed to be given by the rules for *asserting* complex statements. Yet, it has been argued, there are reasons for thinking that *denial* might need to be taken as a primitive too. And, if these reasons are good, one might also argue that logical inferentialists should define the meanings of the logical expressions by means of rules for asserting *and* for denying complex statements. Let *unilateral inferentialism*, or *unilateralism* for short, be the view that the sense of a logical constant is determined by the assertibility conditions of the complex sentences in which it may occur, and let *bilateral inferentialism*, or *bilateralism* for short, be the view that the sense of a logical constant is determined not only by the conditions for correctly asserting the complex sentences in which it may occur, but also by the conditions for correctly *denying* such statements.[8] Given bilateralism, as we shall see, classical logic can be regimented in a proof-theoretically acceptable way, in a *single-conclusion* framework. We shall proceed as follows. Section 7.2.1 discusses some arguments in favour of bilateralism. Section 7.2.2 introduces a bilateralist formalization of classical logic, improving on Timothy Smiley's and Ian Rumfitt's original presentations.[9] Section 7.2.3 critically discusses the merits, and the limits, of the bilateral approach to logic. The approach affords, in the opinion of the present author, the second best available proof-theoretic presentation of classical logic. It remains to be seen,

[8]The expressions 'unilateralism' and 'bilateralism' were first introduced in this context by Rumfitt (2000); my use of the terms is somewhat more restricted than his, however. In keeping with *logical* inferentialism, I only take assertibility and deniability conditions to the determine the sense of the logical expressions. Thus, my understanding of the term remains neutral on the question whether they also fully determine the sense of *sentences*, as full-blooded bilateralists contend.

[9]See Smiley (1996) and Rumfitt (2000).

however, whether the bilateralist's assumptions can be ultimately defended.

7.2.1 Frege's razor and Restall's arguments

Assertion and denial may be represented by means of yes-or-no questions.[10] Suppose we answer "No" to the question: "Is it sunny today?" Then, we have just denied that it is sunny today. Had our answer been "Yes", we would have asserted that it is sunny today. This much is a platitude. What is not a platitude is whether the speech acts of assertion and denial, and the corresponding cognitive states of acceptance and rejection, may both be treated as primitive. Frege famously answered this question in the negative. He first assumed that the denial of A must always be equivalent to the assertion of $\neg A$, and then wondered why we should have *three* primitives instead of just two: "if we *can* make do with one way of judging, then we *must*" (Frege, 1977, p. 48). But can we make do with just one way of judging, as Frege contends? Several authors have defended a negative answer to this question; see for instance Price (1990), Smiley (1996) and Rumfitt (1997). Here we shall focus on a line of argument recently advanced by Greg Restall.[11] Restall's reasoning, I shall argue, may be turned into a powerful argument for adopting a bilateralist approach to logic.

Restall's starting point is the empirical observation that denial appears to be "acquisitionally prior" to negation. Restall writes:

> At face value, it seems that the ability to assert and to deny, to say yes or no to simple questions, arrives earlier than any ability the child has to form sentences featuring negation as an operator. [...] If this is the case, the denial of A, in the mouth of a child, is perhaps best not analysed as the assertion of $\neg A$. So, we might say that denial may be *acquisitionally prior* to negation. One can acquire the ability to deny before the ability to form negations. (Restall, 2005, p. 2)

Now, this observation *per se* is no argument that denial should find a place in any adequate formalization of logic. It simply does not follow that, just because the denial of A should not be analyzed as the assertion of $\neg A$, one should give rules for asserting *and denying* complex statements. After all, $\neg A$ and A's denial are classically equivalent (see e.g. Rumfitt, 2000, p. 818). And, as Frege asked, if can make do with two primitives, why should we use three? Dummett makes a

[10] See Frege (1977), Smiley (1996) and Rumfitt (2000).

[11] A second argument by Ian Rumfitt can be found in Rumfitt (2000, p. 818) and Rumfitt (2002, p. 314).

similar point. If deniability conditions are not to be "idle wheels", he writes, then they "must play some role in fixing the content of an assertion made by means of the sentence" (Dummett, 1976, p. 118). But what role could that be, if we already know that A may be correctly denied when, and only when, $\neg A$ may be correctly asserted ?

It seems to me, however, that this line of argument overlooks a crucial aspect of the classical inferentialist's conception of what a good formalization of logic is. Dummett's objection assumes that there must be a *semantic point* in taking the content of complex statements to be jointly determined by their assertibility and deniability conditions. But this assumption is unjustified. The fact that one can make do without denial does not imply that a formalization of logic in which the denial of A is defined as the assertion of $\neg A$ correctly represents our actual inferential practice. Thus, pending further argument that denial is not acquisitionally prior to negation, it is open to argue that the reason why we have not been able to give a proof-theoretically acceptable formalization of classical logic is that we have blindly followed Frege's *pragmatic* argument for unilateralism. Yet, the classical logician may object, this was a methodological mistake. The classical logician's aim is to *describe* our actual logical practice, on the assumption that it is indeed classical. But, if denial really *is* a primitive speech act, we should not *define* it in terms of assertion and negation, on pain of *distorting* the practice we are trying to describe. Pragmatic considerations such as Frege's and Dummett's implicit appeal to simplicity are beside the point in the present context. If denial really is a primitive, and if the meaning of the logical constant is determined by its correct use, then a faithful description of our logical uses should include both the conditions for asserting complex statements, *and* the conditions for denying them—or so classical inferentialists may argue. Let us see, then, where this assumption leads us.

7.2.2 Bilateralism and classical logic

Drawing on Smiley (1996, p. 5) and Bendall (1979), Ian Rumfitt (2000) presents a *bilateral* formalization of **CPL**, where '$+$' and '$-$' are nonembeddable force signs, and '$+A$' and '$-A$' are signed formulae for any $A \in WFF$, indicating "A? Yes" and "A? No" respectively.

Coordination principles

The system, call it **NBcp**, has the standard structural rules, Reflexivity, Weakening and Cut, together with two 'coordination principles': the following form of *reductio*,

(RED*) From $\alpha \vdash \bot$, infer $\vdash \alpha^*$,

and the following form of the Law of Non-Contradiction,

(LNC*) From α, α^*, infer $\bot$,

where lower case greek letters range over signed formulae, and α^* is the result of reversing α's sign. More perspicuously, coordination rules may be represented thus:

$$\dfrac{\begin{matrix}[+A]\\ \vdots \\ \bot\end{matrix}}{-A}\,\mathsf{RED}^*_{int} \qquad \dfrac{\begin{matrix}[-A]\\ \vdots \\ \bot\end{matrix}}{+A}\,\mathsf{RED}^*_{cl} \qquad \dfrac{+A \quad -A}{\bot}\,\mathsf{LNC}^*$$

Notice the affinity of RED^*_{int} and RED^*_{cl} with intuitionistic and classical *reductio* respectively. If an assertion of A leads to a contradiction, we may discharge $+A$ and deny A, just as in the case of negation introduction we may discharge A and infer $\neg A$ if the assumption that A has lead us to a contradiction. Similarly, if the denial of A leads to a contradiction, we may discharge $-A$ and assert A, just as in the case of the classical rule of *reductio* we may infer A if the assumption that $\neg A$ lead us to a contradiction.[12] Notice, also, that these rules do not define the logical behaviour of any logical constant in particular: they govern inferential relations between *force signs*, which, for the bilateralist, must be sharply distinguished from the logical operators. These rules are assumed to characterize basic properties of the relation of logical consequence: in the logician's jargon, they are *structural rules*.

Unlike Rumfitt, Smiley adopts the following signed version of classical and intuitionistic dilemma, which (i) does not resort to absurdity and (ii) is entailed by Rumfitt's own coordination principles—following Rumfitt, let us call it *Smilean reductio*:

[12]Well, it may be objected that there *is* a difference between these two rules. We know how to discharge an assumption, but do we know how to discharge a *denial*, i.e. a speech-act? I'll come back to this problem in § 7.3.3.

$$\frac{\begin{matrix}[\alpha]^n \\ \vdots \\ \beta\end{matrix} \qquad \begin{matrix}[\alpha]^m \\ \vdots \\ \beta^*\end{matrix}}{\alpha^*}\ \text{SR}, n, m$$

This shows that bilateral systems need not resort to an absurdity constant in the language: we have got enough expressive power with denial alone. Rumfitt's rules, on the other hand, essentially involve absurdity. So how can *he* interpret $\perp$? Following Tennant (1999), Rumfitt suggests that $\perp$ be treated as a logical punctuation sign, i.e. it does not have content, and thus does not need to be interpreted.[13]

Operational rules

Now to the operational rules of the systems. The rules for asserting conjunctions, disjunctions, and implications are exactly like the standard ones, except that formulae are prefixed with an assertion sign. The rules for denying conjunctions are as follows:

$$\frac{-A}{-(A \wedge B)} \quad \frac{-B}{-(A \wedge B)}\ \wedge\text{-I}^- \qquad \frac{-(A \wedge B) \qquad \begin{matrix}[-A]^n \\ \vdots \\ \phi\end{matrix} \qquad \begin{matrix}[-B]^n \\ \vdots \\ \phi\end{matrix}}{\phi}\ \wedge\text{-E}^-, n$$

Together with the rules for denying disjunctions

$$\frac{-A \qquad -B}{-(A \vee B)}\ \vee\text{-I}^- \qquad \frac{-(A \vee B)}{-A} \quad \frac{-(A \vee B)}{-B}\ \vee\text{-E}^-$$

they highlight the duality of conjunction and disjunction. The rules for denying implications, on the other hand, reveal the materiality of the classical conditional:

$$\frac{+A \qquad -B}{-(A \to B)}\ \to\text{-I}^- \qquad \frac{-(A \to B)}{+A} \quad \frac{-(A \to B)}{-B}\ \to\text{-E}^-$$

Finally, the rules for negation are radically new:

$$\frac{-A}{+(\neg A)}\ \neg\text{-I}^+ \quad \frac{+(\neg A)}{-A}\ \neg\text{-E}^+ \quad \frac{+A}{-(\neg A)}\ \neg\text{-I}^- \quad \frac{-(\neg A)}{+A}\ \neg\text{-E}^-$$

The foregoing rules give us a sound and complete formalization of **CPL**.[14]

---

[13]See *supra*, § 6.5.5. Alternatively, bilateralists may take $\perp$ to be inferentially defined by the following rules:

$$\frac{-A \qquad +A}{\perp} \qquad \frac{\perp}{+A} \quad \frac{\perp}{-A}$$

There are necessary and sufficient conditions for asserting $\perp$, but they can never be met, on the assumption that we can never be in a position to assert and deny the same statement.

[14]See Smiley (1996).

Remarks

A few remarks are in order. First of all, **NBcp**'s operational rules, including the rules for negation, satisfy both strong intrinsic harmony, and, if rewritten in a GE format, GE harmony. On the other hand, Rumfitt's rules for asserting disjunctions and for denying conjunctions are not fully invertible.[15] In the case of negation, what is needed for asserting $\neg A$ is precisely what we may derive from an assertion of $\neg A$; likewise, what is needed for denying $\neg A$ is precisely what we may derive from $\neg A$'s denial (see Rumfitt, 2000, p. 806).[16] What is more, *intuitionistic* negation turns out to be non-harmonious on a bilateral approach to sense. As Rumfitt puts it:

> within a bilateral framework, one will wish to know why the intuitionistic logician has no general account to offer of the consequences of rejecting a negated sentence or formula. (Rumfitt, 2000, p. 806)

Second, **NBcp** satisfies separability, both weak and strong—see Bendall (1978), Bendall (1979), and (Rumfitt, 2000, pp. 808-9). Third, the extra expressive power obtained by adding force signs for denial allows for a solution of the Carnap Problem, or at least so bilateralists argue (see Rumfitt, 2000, pp. 807-8).

Let us define a set of *correctness-valuations* C for signed formulae such that every member is induced by the truth-valuations in V by the following correctness clauses:[17]

(C1) $v_c(+A) = \mathbf{1}$ iff $v(A) = 1$;

(C2) $v_c(-A) = \mathbf{1}$ iff $v(A) = 0$.

One may correctly assert (deny) A just in case A is true (false). Validity for signed formulae may be defined thus:

(VAL) $\Gamma \models \alpha$ is valid just in case, for every correctness-valuation $v_c \in C$, whenever $v_c(\beta) = \mathbf{1}$ for every $\beta \in \Gamma$, $v_c(\alpha) = \mathbf{1}$.

[15]They can nevertheless be substituted with fully invertible rules, as we shall see in a moment.

[16]Rumfitt observes that LNC* provides a justification of the Law of Double Negation Elimination. For suppose that the principle holds for *atoms*. Then, Rumfitt writes, "we shall need to be able to show that [the specifications of the sense of the connectives] entail the coordination principle $+A, -A \vdash \bot$ for each well-formed formula A, given the information that the atoms are so coordinated" (Rumfitt, 2000, p. 816). In the case of negation, given $+A, -A \vdash \bot$, one would need to show that $+(\neg A), -(\neg A) \vdash \bot$. Given $+$-$\neg E$, $+(\neg A) \vdash -A$ may be easily derived. Yet, without assuming $(-$-$\neg E)$, there is no way one can get from $-(\neg A)$ to $+A$.

[17]See Humberstone (2000, p. 345). Boldface '**1**' and '**0**' should be read as 'correct' and 'incorrect' respectively.

Now consider $\neg$-E$^+$). If v^* were admissible, this rule would fail to preserve correctness: given $v^*(A) = v^*(\neg A) = 1$, there must be a $v_c \in C$ such that $v_c(+(\neg A))$ $= \mathbf{1}$, but $v_c(-A) = \mathbf{0}$. Similarly for disjunction: if $v^@$ were admissible, $\vee$-I$^-$ would not be correctness-preserving. If C1 and C2 are in place, Carnap's problem seems solved: one cannot add Carnap's deviant valuations without affecting the validity of the inference rules. Notice, however, that this solution requires that C1 and C2 hold. That is, denial bilateralists need to assume that denial 'means' what it is supposed to mean—an assumption which has been challenged, and to which we shall return in § 7.3.2 and § 7.3.3 below. Moreover, and *crucially*, bilateralists need to assume that the *relata* of the relation of logical consequence are items of the form *speech act* + *content*.

What Rumfitt's bilateralist offers, then, is a sound and complete formalization of **CPL** satisfying each of the following properties:

(i) the system satisfies both weak and strong separability;

(ii) the operational rules are all harmonious, in two of the three senses defined in Chapter 5;

(iii) the system allows for a solution of Carnap's Categoricity Problem—albeit a controversial one, as we shall see in §§ 7.3.3-4.

7.2.3 Minor adjustments

There are, however, some outstanding issues. As we have already observed, Rumfitt's system does not satisfy full invertibility. Moreover, it also fails to satisfy the Fundamental Assumption, at least in Dummett's and Prawitz's original formulation. The bilateralist proof of the Law of Excluded Middle does not end by a step of disjunction introduction, anymore than its multiple-conclusions counterpart does (see *supra* § 7.1.2):

Example 13. *The Law of Excluded Middle in* **NBcp**:

$$\dfrac{\dfrac{-(A \vee \neg A)}{-A}\,\vee\text{-E}^- \qquad \dfrac{\dfrac{-(A \vee \neg A)}{-\neg A}\,\vee\text{-E}^-}{+A}\,\neg\text{-E}^-}{+A \vee \neg A}\,\text{SR}$$

Bilateralists, though, may solve both issues at once by substituting Rumfitt's positive rules for disjunction with the following rules—rules whose non-signed analogues I will attempt to briefly justify in § 7.4.1:

$$+\lor\text{-I}^*, i\ \dfrac{\begin{matrix}[-A, -B]^i \\ \vdots \\ \bot\end{matrix}}{+(A \lor B)} \qquad\qquad +\lor\text{-E}^*\ \dfrac{+(A \lor B) \qquad -A \qquad -B}{\bot}$$

These rules are fully invertible: one is allowed to infer from $A \lor B$ precisely what was required to introduce it in the first place. Unsurprisingly, they also satisfy strong intrinsic harmony. The reduction step is as follows:

$$+\lor\text{-E}^*\ \dfrac{+\lor\text{-I}^*, i\ \dfrac{\begin{matrix}\Delta, [-A, -B]^i \\ \Sigma \\ \bot\end{matrix}}{+A \lor B} \qquad \begin{matrix}\Gamma_0 \\ \Pi_0 \\ -A\end{matrix} \qquad \begin{matrix}\Gamma_1 \\ \Pi_1 \\ -B\end{matrix}}{\bot} \quad \rightsquigarrow_r \quad \begin{matrix} \underbrace{\Delta \qquad \begin{matrix}\Gamma_0 \\ \Pi_0 \\ -A\end{matrix} \qquad \begin{matrix}\Gamma_1 \\ \Pi_1 \\ -B\end{matrix}} \\ \Sigma \\ \bot \end{matrix}$$

And here is the corresponding expansion:

$$\dfrac{\Pi}{+(A \lor B)} \quad \rightsquigarrow_e \quad \dfrac{\lor\text{-E}^*\ \dfrac{\dfrac{\Pi}{+(A \lor B)} \qquad \overline{-A, -B}}{\bot}}{+(A \lor B)}$$

Finally, the modified I-rule now allows us to prove LEM by means of a proof ending with a step of disjunction introduction—one just needs to assume $-A$ and $-\neg A$, and conclude $+(A \lor B)$ by one step of $+\lor$-I*.[18]

The foregoing rules for $\lor$ are derivable in **NBcp**.[19] Conversely, the standard

[18]Similar harmonious rules for asserting existential statements are also available:

$$+\exists\text{-I}, i\ \dfrac{\begin{matrix}[-F(t/x)]^i \\ \vdots \\ \bot\end{matrix}}{+\exists x Fx} \qquad\qquad +\exists\text{-E}\ \dfrac{\begin{matrix}\Gamma_0 \\ \vdots \\ +\exists x Fx\end{matrix} \qquad \begin{matrix}\Gamma_1 \\ \vdots \\ -F(a/x)\end{matrix}}{\bot}$$

where a may not occur free in $\exists x Fx$ and Γ_1. The reduction step is as follows:

$$+\exists\text{-E}\ \dfrac{+\exists\text{-I}, i\ \dfrac{\begin{matrix}\Gamma_0, [-F(t/x)]^i \\ \Pi_0 \\ \bot\end{matrix}}{+\exists x Fx} \qquad \begin{matrix}\Gamma_1 \\ \Pi_1 \\ -F(a/x)\end{matrix}}{\bot} \quad \rightsquigarrow \quad \begin{matrix}\Gamma_1 \\ \Pi_1 \\ -F(a/x) \\ \Gamma_0, -F(t/x) \\ \Pi_0 \\ \bot\end{matrix}$$

The corresponding expansion is also straightforward. Analogous rules for denying existential statements are obtained by reversing the signs.

[19]*Proof*: For the introduction rule, assume that $\bot$ follows from A, B. Then, $\bot$ follows from $-A, -B$, which allows us to infer $(\neg A \land \neg B) \rightarrow \bot$. But this entails $\neg(\neg A \land \neg B)$, from which $+A \lor B$ is classically (but not intuitionistically) derivable. For the elimination rule, assume $+A \lor B$, $-A$, $-B$. By a version of Disjunctive Syllogism, $+B$ follows from $+A \lor B$ and $-A$. Contradiction. By negation elimination, we may conclude $\bot$. ■

rules are derivable from the ones I am suggesting.[20] Dual fully invertible rules are also available for introducing denied conjunctions:

$$\dfrac{\begin{matrix}[+A, +B]^i \\ \vdots \\ \bot\end{matrix}}{-(A \wedge B)}\,{\scriptstyle -\wedge\text{-I}^*, i} \qquad \dfrac{-(A \wedge B) \quad +A \quad +B}{\bot}\,{\scriptstyle +\vee\text{-E}^*}$$

These rules are likewise provably equivalent to Rumfitt's rules.[21] In light of the foregoing interderivability results, our emended formalization of **CPL**—call it **NBcp**$^+$—is sound and complete with respect to **CPL** if and only if **NBcp** is.

7.3 Bilateralism, denial, and external negation

Bilateralist formalizations of logic have a number of virtues, as we have just seen. All the same, they raise several issues, to which we now turn. Section 7.3.1 discusses an objection by Dummett, to the effect bilateralists are committed to an incoherent conception of assertion. Section 7.3.2 offers considerations for thinking that denial just is an external *negation* and introduces an argument by Luca Incurvati and Peter Smith's to the effect that it is not. Sections 7.3.3 respond to Incurvati and Smith's objections.

7.3.1 Dummett's charge of incoherence

I begin with an objection by Dummett, to the effect that bilateralism entails an untenable view of assertion. If deniability conditions are not to be idle wheels, Dummett writes, they:

> must play some role in fixing the content of an assertion made by means of the sentence. It would then follow that a speaker might be neither right or wrong in making an assertion: not wrong, because it could be shown that the sentence could not be falsified; but not right either, because no way was known of verifying the sentence. This consequence would be fatal to the account, since an assertion is not

[20]*Proof:* For the I-rules, assume $+A$. Then assume $-A$ and $-B$. Contradiction. By $\vee$-I*, we may thereby infer $+A \vee B$ and discharge $-A$ and $-B$. Similarly for the proof from $+B$ to $+A \vee B$. For the E-rule, assume $+A \vee B$, $+A \vdash \delta$, and $+B \vdash \delta$. But given that, if $\alpha \vdash \beta$, then $\beta^* \vdash \alpha^*$, $\delta^* \vdash -A$ and $\delta^* \vdash -B$ follow. Notice, though, that $+A \vee B$, $-A$ and $-B$ entail $\bot$ by $+\vee$-E*), which in turn allows us to infer $-(A \vee B)$ from $-A$ and $-B$. By the transitivity of deduction, we get that δ^* entails $-(A \vee B)$. Since, however, if $\beta^* \vdash \alpha^*$, then $\alpha \vdash \beta$, we can conclude $+A \vee B \vdash \delta$. ■

[21]The proof is dual to the one we have just given in fn. 19.

> an act which admits of an intermediate outcome [. . .] it is a *reductio ad absurdum* of any theory of meaning if it entails that [it] is. (Dummett, 1976, p. 118)

Dummett further clarifies his terminology:

> we may say that the *speaker* is right if he is, at the time of the speaking, able to verify what he says, but that his *assertion* is correct if there is some means of verifying it, a knowledge of which by the speaker at the time of utterance would have made him right. The sense in which it is false to the nature of assertion to say that an assertion may be right nor wrong is that in which, in this terminology, the assertion itself is neither correct nor incorrect. (Dummett, 1976, p. 119)

The complaint seems to be that (i) bilateralism requires the existence of assertions that are neither correct nor correct, but (ii) it is false to the *nature* of assertion that there cannot be assertions that are neither correct nor incorrect. Dummett's argument seems problematic, however.

To begin, Dummett appears to be wrong in thinking that, if assertions can be neither correct nor incorrect, then deniability conditions are idle wheels. As we have already observed in § 7.2.1, classical logicians are out to *represent* our actual logical practice, and pragmatic considerations such as these ones appear to be beside the point in the present context.

But even setting this problem aside, Dummett's contention that there cannot be assertions that are neither correct nor incorrect is also suspect. On several views of vagueness, for instance, borderline instances of vague predicates may typically allow for assertions that are neither correct nor incorrect. And although Dummett explicitly states that he is setting vague statements aside, this does not quite alleviate the worry. Given that vague statements constitute the *vast* majority of the statements we actually make, it seems at least contentious to set vagueness aside, and claim that it is part of the *nature* of assertion that assertions do not admit of intermediate outcomes.

The next striking feature of Dummett's argument is Dummett's use of the term 'correct'. An assertion is correct, in Dummett's sense, if and only if there is a means of verifying it. That is, on Dummett's anti-realist assumption that a sentence is true if and only if there is a means of verifying it, "the truth of a sentence should be equated with its being objectively correct to assert it" (Dummett, 2002, p. 294). It follows that there cannot be assertions of which we know that that they are neither true nor false. Dummett writes:

> if our logic at all resembles intuitionistic logic, there is indeed no possibility of discovering, for any statement, that it can be neither verified nor falsified, since whatever would serve to show that it could be not be verified would *ipso facto* verify its negation. (Dummett, 1976, p. 119)[22]

If sound, this argument would be lethal for the bilateral view. If deniability conditions are needed only if there are assertions that are neither correct nor incorrect, or neither true nor false, in Dummett's sense, it follows by the above result that deniability conditions are not needed on a use-based approach to meaning.

Rumfitt (2000, p. 818) blocks the above argument by making the obvious move, i.e. by denying that an assertion is correct only if there is a means of verifying it. He writes (in a slightly different context):

> The oddity arises [only] if truth is equated with the correctness of assertion and falsity with the correctness of denial; and I accept neither of these equations as generally correct theses about truth and falsity. (Rumfitt, 2002, p. 313)

He then concludes that, from an inferentialist perspective, the bilateral view, and hence classical logic, is the right view to adopt whenever there are assertions that are neither correct not incorrect, while conceding that intuitionistic logic might well be the right logic for areas of discourse in which there are no assertions that are neither correct nor incorrect.[23]

Rumfitt's reply is perfectly legitimate. Moreover, it is worth recalling that this is a context in which the reformist is trying to put forward a *proof-theoretic* argument for the adoption of intuitionistic logic. On the face of it, Dummett's crucial assumption is a *metaphysical* claim concerning the relationship between truth and our epistemic capacities—a far cry from the proof-theoretic constraints we were starting from.

[22]We may represent Dummett's assumption that a statement is true only if it can be correctly asserted as follows:

$$\text{(VER)}\ \forall\varphi(\varphi \to C\varphi),$$

where '$C\varphi$' reads 'φ is correctly assertible'. Now suppose $\neg CP \wedge \neg C\neg P$. Then, $\neg CP$ follows. By contraposition of VER, $\neg P$ also follows. By VER, we may then infer $C\neg P$. Contradiction. We may then conclude $\neg(\neg CP \wedge \neg C\neg P)$, as required. As Dummett observes (Dummett, 1976, pp. 121-122), this does not intuitionistically entail that every assertion is either correct or incorrect.

[23]See Rumfitt (2000, Section VIII) and Rumfitt (2002, p. 313).

Let us now turn to a different objection—one according to which the bilateralist's solution of Carnap's categoricity problem is not viable after all.

7.3.2 Correctness valuations and external negation

We have seen that the system **NBcp**$^+$ is in many respects superior to the standard natural deduction formalizations of **CPL**. One might wonder how much has been achieved, however. Murzi and Hjortland (2009) raise the following objection. Consider a *correctness valuation* v_c^* such that $v_c^*(\alpha) = \mathbf{1}$, for every $\alpha \in WFF_{sign}$. Although on v_c^* both A and $\neg A$ are correctly assertible, each of $\mathcal{J}$'s rules is still valid, in the sense that the assertibility of the premises guarantees the assertibility of the conclusions. But, if it is correct to assert both A and $\neg A$, '$\neg$' can hardly be seen as a negation sign. Carnap's problem, it would seem, has now been been shifted to the next level.

It might be objected that v_c^* does not respect the correctness clauses

(C1) $v_c(+A) = \mathbf{1}$ iff $v(A) = 1$;

(C2) $v_c(-A) = \mathbf{1}$ iff $v(A) = 0$,

and that these principles are constitutive of assertion and denial. But this seems problematic. Syntactically, C2 and

(NEG) $v(\neg A) = 1$ iff $v(A) = 0$

are exactly alike. Yet, NEG was precisely the kind of semantic clause which inferentialists were not allowed to invoke, on pain of assuming a prior understanding of the connectives they want to define. Furthermore,

(RED*) From $\alpha \vdash \bot$, infer $\vdash \alpha^*$

and

(LNC*) From α, α^*, infer $\bot$

may be seen as classical rules governing '$-$', which, from an inferentialist perspective, may after all suggest that '$-$' just is a negation operator, and not a force sign. The only relevant difference between '$-$' and '$\neg$' is that the latter is embeddable, but not the former. However, a result by Kent Bendall (1979) shows that such a difference is expressively irrelevant. As he points out, '$-$' has sufficient expressive power to replace '$\neg$'. For let L$^\sim$ be the result of subtracting '$\neg$' from L, while adding a nonembeddable negation '$\sim$'. Then, it can be proved that, for any sentence in L with an embeddable classical negation, there is a unique logically equivalent sentence in L$^{\sim}$ that is either negation-free or of the form $\sim A$, where A is some negation-free sentence.[24] It follows that both '$-$' and '$\sim$' may be seen as a special kind of negation operators. But then, there are grounds for suspecting that the bilateralist is violating the rules of the game: if both '$-$' and '$\sim$' are *negations*, it is hard to see why C2 should be acceptable, if NEG is not.

A brief historical note. A similar worry had already been raised by Alonzo Church in his 1944 review of Carnap's *Formalization of Logic*. Carnap's recipe for ruling out 'non-normal' valuations for the classical connectives was to resort to what he called "junctives" and "disjunctives": essentially, a multiple-conclusions formalizations of logic, where commas in the antecedent of any given sequent are interpreted conjunctively, and commas in the succedent are interpreted disjunctively—just as '$-$' in a bilateral formalization of logic is intuitively interpreted as a classical external negation. But then, Church objected, Carnap's solution is not purely syntactical, as it presupposes that some structural expressions have a fixed interpretation. Church writes:

> In view of his requirement that disjunctives be interpreted in a particular way, Carnap's use of them is a concealed use of semantics; and in fact, if this arbitrary requirement is dropped, non-normal interpretations of his "full formalization" become possible. (Church, 1944, pp. 495-96)

If commas cease to mean what they Carnap takes them to mean, Carnap's non-normal interpretations are not ruled out. Similarly with denial, as we have seen: if denial does not mean what the bilateralist takes it to mean, non-normal interpretations are not ruled out either.

Incurvati and Smith (2010) have recently argued that, *pace* Murzi and Hjortland, it is a mistake to think that Carnap's problem has just been moved to the next level. They write:

> It is propositional *contents* that are the primary locus of evaluation, and it is in terms of such an evaluation that validity is being basically defined; and Smiley's rules ensure that negation – which is, remember, an operation on *contents* – behaves as we want. True, a positively signed sentence can then *derivatively* be said to be correct if it has a true propositional content, and a negatively signed sentence is derivatively

[24]See Bendall (1979, pp. 69-70).

> correct if it has a false propositional content, and so on. So we can give a derivative account of classical validity in terms of correctness-preservation. But, so defined, there just can't be a correctness-valuation which makes all signed sentences true together – for by our starting hypothesis we cannot simultaneously correctly assert and correctly reject the same content. So the alleged problem doesn't arise. (Incurvati and Smith, 2010, p. 9)

Furthermore, Incurvati and Smith contend that C1 and C2, and the "meaning" of the assertion and negation markers "are not up for revision, but [they are] part of the assumed background": "it is a given that we cannot simultaneously correctly use $+P$ and $-P$" (Incurvati and Smith, 2010, p. 10). The general point is that definitions and interpretations must come to an end somewhere. For the bilateralist, the ultimate bedrock is the 'meaning', i.e. the logic, of denial. *This* meaning is not up for revision. Quite the contrary, it has to be our starting point. It is the ultimate root of the meaning of classical negation, and of the categoricity of the meaning of the classical connectives.

However, even conceding that denial satisfies the classicist's coordination principles, there are reasons for thinking that Incurvati and Smith's insistence that '$-$' really expresses a *speech act*, and not an external negation, is ultimately unjustified. A closer look to the rules for denial reveals that denial may be incapable of serving the logical role bilateralists need it to serve, or at least so I shall argue.

7.3.3 Assuming denials

Let us consider again the crucial point of opposition between Murzi and Hjortland (2009) and Incurvati and Smith (2010). Murzi and Hjortland argued that '$-$' and '$\sim$', denial and external negation, look suspiciously similar. They both obey the same rules, with the proviso that neither '$-$' and '$\sim$' can ever be embedded. Moreover, they both satisfy satisfaction causes that also look very much alike. To this, Incurvati and Smith objected that the fact that '$-$' and '$\sim$' have the same logic does not imply that they are the same thing: the former is a force marker, which does not contribute to the content expressed by $-A$; the latter is a meaningful expression, which indeed contributes to $\sim A$'s content. They look similar on the surface, but they are quite different *things*.

Or are they? Classical bilateralists assume that denial figures in such rules as RED^*_{cl}, RED^*_{int}, *Smilean reductio*, and $+\vee$-I*. These *indirect* rules—i.e. rules involving discharge of assumptions—are needed if bilateralism is to give us classical logic at all. As Peter Gibbard (2002, p. 297, fn. 2) observes, without *Smilean reductio*, or some analogous principle, the operational rules of Rumfitt's original system give us a constructive logic with strong negation, but not classical logic.[25]

The question arises, though, as to how to make sense of rules such as RED^*_{cl}, RED^*_{int}, *Smilean reductio*, on the assumption that '$-$' really expresses *denial*, and not external negation. In outline, the difficulty is that if, as it seems plausible, (i) to assume A is *already* a speech act, and (ii) speech acts are not embeddable, one cannot assume a speech act, i.e. one cannot assume $+A$ and $-A$.

For the sake of simplicity, let us focus on RED^*_{cl}. How are we to interpret this rule? Here is a natural, but problematic, suggestion:

> "Assume that A has been denied. If you thereby reach a contradiction, you may discharge your assumption, and infer that A can be asserted."

The difficulty is this. While one can certainly assume *that A has been denied* (*that A has been asserted*), this is not the same as assuming $-A$ $(+A)$. What we have now assumed is an altogether *new content*, not something of the form speech-act + content. How can we make sense of rules such as RED^*_{cl}, without reducing $-A$ to 'not A'? Here is another suggestion:

> "Assume A's denial. If you thereby reach a contradiction, you may discharge your assumption, and conclude A's assertion."

This does not seem to work either, however. Presumably, "A's denial" denotes an *event*. But events are not the kinds of things we assume or discharge. Moreover, it is even more difficult to see how one can *conclude* A's assertion! We can conclude A, or we can conclude *that A has been asserted*. But, it would seem, events are not the kinds of things we can *conclude*.

Perhaps the following might do:

> "Assume A's *deniability*. If you thereby reach a contradiction, you may discharge your assumption, and *infer* A's *assertibility*."

But again, this seems problematic. For what is to assume A's deniability? What is to discharge A's assertibility? The expressions 'Assume A's deniability" and

[25]It is an interesting question which logic is obtained by dropping the co-ordination principles from $\mathcal{J}^*$—one, however, that I do not have space to explore here. The important point to notice for present purposes is that $+\vee$-I* alone, even in presence of the bilateralist rules for negation, does not give us classical logic (the standard positive rule of disjunction elimination, $+\vee$-I, can only be derived if either RED^*_{cl} or DN are in place. However, the former is a co-ordination principle, and does not hold in the system we are considering. As for DN, it cannot be derived from the bilateralist rules for negation alone).

"Discharge A's assertibility" just seem shorthand for, respectively, "Assume *that A can be denied*" and "Discharge the assumption *that A can be asserted*". The bilateralist might object that this is just to reiterate the old point that '$-$' and '$\sim$' look similar. But it is not. Compare with similar expressions like "Assume A's possibility" or "Discharge A's knowability". If these expressions make sense at all, they too are shorthand for, respectively, "Assume *that A is possible*" and "Discharge the assumption *that A is knowable*". There is nothing else these expressions could mean, if they mean anything at all.

Bilateralists may concede that we assume propositions or sentences, and not objects, but, at the same time, object that we may well *imagine* denials and assertions.[26] Thus, they may say, RED^{*}_{cl} is to be read:

> "Imagine A's denial/deniability. If you thereby reach a contradiction, you may stop imagining A's denial/deniability, and conclude A's assertibility".

This is hardly an improvement, however. For one thing, we are still asked to 'conclude A's assertibility'. For another, it would seem that the question whether we imagine $-A$ should be independent of whether $-A$ has been assumed or not. Finally, it is difficult to see how one could *discharge* $-A$, given that we were simply asked to *imagine* A's denial. Similarly for *trying*:[27]

> "Try denying A. Suppose a contradiction results from doing so. You may then stop trying to deny A, and assert A."

Here, too, it does not seem that "to stop trying to deny A" comes close to what the bilateralist really needs, viz. that A may now be *discharged*. For suppose I try to deny A. For instance, I respond "No!" to the question whether A. Suppose, too, that A's denial entails absurdity. How are we now to make sense of the instruction that I may now stop trying to deny A? After all, I might well answer: "I have already tried to deny A! How can I stop doing something that I already did?" If denials are events, it is difficult to see how discharging $-A$ can be rendered as ceasing to attempt to deny A.

Bilateralists may perhaps revert to talk of *pretense*:[28]

> "Pretend A's denial/deniability. If you thereby reach a contradiction, you may stop pretending A's denial/deniability, and conclude A's assertibility".

[26] Many thanks to Ole Hjortland for mentioning this possible interpretation.
[27] Many thanks to Bob Hale for suggesting this possible interpretation.
[28] Many thanks to Dominic Gregory for suggesting this possibility.

But again, this seems ungrammatical. Moreover, to pretend that P seems to imply that P is not true—and certainly it would be misleading to interpret the mere *assumption* that P in this way.

It seems fair to conclude that none of these attempts seems very promising. By contrast, rules such as RED^{*}_{cl} and RED^{*}_{int} make perfect sense, if '$-$' is interpreted as an *external negation*. One only has to interpret them as one would interpret, respectively, CR and $\neg$-I, with the only proviso that 'not' is not embeddable. We are left, it would seem, with a very strong suspicion that, in bilateralist formalizations of logic, denial *is* an external negation operator—it is not just that it appears to behave like one. Or are we?

The bilateralist might agree with what has been said so far. However, she might object that a fairly natural option has been left out, viz. to assume that there are *two modes of assumption*: a negative and a positive one. Recall, $+A$ and $-A$ are to be respectively interpreted as 'A? Yes!' and 'A? No!'. On this assumption, bilateralists might argue, the positive (negative) assumption of $+A$ ($-A$) can be naturally rendered as 'A? Suppose yes!' ('A? Suppose no!').[29]

There is a hitch, though. This response requires assertions and denials to have a question-answer form: questions and answers are not merely a way of representing assertions and denials in our formalization of logic. Yet, it would seem, there could be linguistic communities that are just like ours, except that *nobody ever asks questions*. (For generations, whoever asks questions is killed. Eventually, the very concept of a question is lost. Or maybe the members of these communities know everything they need to know. They do not need to ask questions, and they lack the concept of something they never do, nor need to do.) The members of these communities, we might imagine, have a perfect command of the English language. They assert and deny propositions, they give commands, they implore each other ...but, odd as this may seem, questions are not to be found among the speech acts they master. What should bilateralists say about them? How can the members of these communities assume assertions and denials? Bilateralists seem forced to say that either these communities are not possible, or their members cannot give meaning to their logical vocabulary—at the very least, they cannot give it a *classical* meaning. But this appears to be a bad consequence. For one thing, these communities seem perfectly possible. For another, being all-knowing or uncritical should not have consequences on whether $A \vee \neg A$ is a logical law or not!

[29]Thanks to Ian Rumfitt for supplying this telling interpretation of his own formalism.

7.4 Classical harmony

If correct at all, the foregoing considerations suggest the classical inferentialists must do better. The aim of this last section is to indicate a possible way for them to do so. There are at least five constraints that any proof-theoretically adequate formalization of classical logic C must respect:

(i) C must satisfy separability, both weak and strong;

(ii) C's rules must all be harmonious, in some sense of the term;

(iii) C must satisfy the Fundamental Assumption (and thus allow for a proof-theoretic definition of validity; see § 4.1.2 and Appendix D below);

(iv) C's rules must determine the meaning of the logical operators in the sense of determining their satisfaction clauses;

(v) C may not involve structural assumptions that cannot be plausibly made sense of.

In what follows, we will consider a formalization of **CPL** that satisfies each of (i)-(v). Section 7.4.1 introduces classical rules for disjunction—rules that, unlike the standard ones, satisfy the Fundamental Assumption, even if the logic is classical. Section 7.4.2 shows how CR can be interpreted as a structural rule and introduces a classical system **NHcp**. Section 7.4.3 proves a normalization theorem for **NHcp** that, unlike Prawitz's original theorem, entails separability and the subformula property. Section 7.4.4 argues that the foregoing rules allow us to derive the satisfaction clauses for negation and disjunction.

7.4.1 Classical disjunction and classical *reductio*

One often hears that the standard introduction rules for disjunction do not adequately reflect the way disjunctions are asserted in everyday practice, and that the meaning of 'or' in ordinary language is radically different from its meaning in logic. This complaint seems reasonable enough: we almost always assert $A \vee B$ on the grounds that A and B cannot both be false—not because we already know that one of the two disjuncts is true. As Scott Soames puts it:

> nearly always when we assert the disjunction of A and B in ordinary language, we do *not* so because we already know that A is true, or because we already know that B is true. Rather, we assert the disjunction because we have some reason for thinking that it is highly

unlikely, perhaps even impossible, that both A and B will fail to be true. (Soames, 2003, p. 207)

This suggests that inferentialists may adopt the following I-rule for disjunction instead, with the corresponding harmonious E-rule:

$$\vee\text{-I}^*, n\,\frac{\begin{matrix}[\neg A, \neg B]^n\\ \vdots\\ \bot\end{matrix}}{A \vee B} \qquad\qquad \vee\text{-E}^*\,\frac{A \vee B \quad\quad \neg A \quad\quad \neg B}{\bot}\,.$$

Like their signed analogues, the rules satisfy both (i) full invertibility and (ii) strong intrinsic harmony. Ad (i), it is sufficient to notice, once more, that one is allowed to infer from $A \vee B$ precisely what was required to introduce it in the first place. Ad (ii), the *reduction step* is as follows:

$$\vee\text{-E}^*\,\cfrac{\vee\text{-I}^*, i\,\cfrac{\begin{matrix}\Delta, [\neg A, \neg B]^i\\ \mathcal{D}_1\\ \bot\end{matrix}}{A \vee B} \quad \begin{matrix}\Gamma_0\\ \mathcal{D}_2\\ \neg A\end{matrix} \quad \begin{matrix}\Gamma_1\\ \mathcal{D}_3\\ \neg B\end{matrix}}{\bot} \quad \leadsto_r \quad \begin{matrix}\underbrace{\Delta \quad \begin{matrix}\Gamma_0\\ \mathcal{D}_2\\ \neg A\end{matrix} \quad \begin{matrix}\Gamma_1\\ \mathcal{D}_3\\ \neg B\end{matrix}}\\ \mathcal{D}_1\\ \bot\end{matrix}$$

A derivation of $\bot$ via the unnecessary detour originated by an application of $\vee$-I* immediately followed by an application of $\vee$-E* can always be reduced to a more direct derivation of $\bot$ from the same or fewer assumptions that does not resort to our disjunction rules. The corresponding *expansion* is also straightforward:

$$\frac{\Pi}{A \vee B} \quad \leadsto_e \quad \vee\text{-E}^*\,\cfrac{\cfrac{\Pi}{A \vee B} \quad \cfrac{}{\neg A, \neg B}}{\cfrac{\bot}{A \vee B}}\,.$$

The foregoing rules for disjunction are therefore harmonious in two of the three senses of harmony we introduced in Chapter 5: they satisfy both strong intrinsic harmony and full invertibility.

Higher-order rules

What about GE harmony? In order to answer this question, we now need to rewrite the E-rule in a GE format. Recall, GE rules tell us that whatever sentence C follows from the canonical grounds for A also follows from A itself. The canonical grounds for $A \vee B$, as specified by $\vee$-I*, are that, if we have derived $\bot$ from $\neg A, \neg B$, we may discharge our assumptions $\neg A, \neg B$ and introduce $A \vee B$. We thus get the following GE rule:

$$\vee\text{-E}^{*}_{GE},n\ \frac{A \vee B \qquad \begin{matrix}[\neg A, \neg B \Rightarrow C]^n \\ \vdots \\ C\end{matrix}}{C}\ .$$

The rule tells us that, if we can assert $A \vee B$, and if C follows from a derivation of $\perp$ from $\neg A, \neg B$, we may discharge $\neg A, \neg B \Rightarrow C$ and infer C. Unlike any of the rules we have encountered so far, this rule allows us to discharge *derivations*, as opposed to assumptions. It is, in Peter Schroeder-Heister's terminology, a *higher-order* rule (see Schroeder-Heister, 1984).

Now, even very common rules can be rewritten as higher-order rules. For instance, the standard GE E-rule of arrow elimination is often written as follows:

Example 14. *$\rightarrow$-E_{GE} (higher-order):*

$$\rightarrow\text{-E}^{ho}_{GE},n\ \frac{A \rightarrow B \qquad \begin{matrix}[A \Rightarrow B]^n \\ \vdots \\ C\end{matrix}}{C}\ .$$

Why should we introduce rules for discharging *derivations*, however? This question is best approached by first asking ourselves what it is to assume something in the context of a derivation. Let us begin with the assumption of *formulae*. Schroeder-Heister persuasively argues that to assume some formulae $\beta_1, \ldots, \beta_n$ is, technically, just to treat these formulae as *temporary axioms*:

> Assumptions in sentential calculi technically work like additional axioms. A formula α is derivable from formulas $\beta_1, \ldots, \beta_n$ in a calculus C if a is derivable in the calculus C' resulting from C by adding $\beta_1, \ldots, \beta_n$ as axioms. But whereas "genuine" axioms belong to the chosen framework and are usually assumed to be valid in some sense, assumptions bear an *ad hoc* character: they are considered only within the context of certain derivations. When deriving a from $\beta_1, \ldots, \beta_n$ we do not want to change our framework and to extend the calculus C; we are interested in the derivability relation between $\beta_1, \ldots, \beta_n$ and α with respect to C. This *ad hoc* character of assumptions, as compared with axioms, is made obvious in natural deduction systems: some of their inference rules allow one to *discharge* assumptions used in the derivations of the premises-that means, such assumptions are used only in specific subderivations for the purpose of establishing a certain formula in the superior derivation. (Schroeder-Heister, 1984, p. 1284)

But, if assumptions just are *ad hoc* axioms, one should also be free to use *ad hoc* rules in the context of a derivation. For why should be willing to temporarily expand our logical system exclusively with axioms, and not with *rules*? Thus Schroeder-Heister again:

> Instead of considering only *ad hoc* axioms (i.e. assumption formulas) we can also regard ad hoc inference rules, that is, inference rules [...] used as assumptions. Assumption rules technically work like additional basic rules: α is derivable from assumption formulas $\beta_1, \ldots, \beta_n$ and assumption rules $\rho_1, \ldots, \rho_m$, in C if a is derivable in C', where C' results from C by adding $\beta_1, \ldots, \beta_n$ as axioms and $\rho_1, \ldots, \rho_m$ as basic inference rules.(Schroeder-Heister, 1984, p. 1285)

If Schroeder-Heister's account of what it is to make an assumption is along the right lines, higher-order rules need not be regarded as new exotic animals in our proof-theoretic zoo. Quite the contrary: they stand, or fall, with the standard indirect rules involving discharge of assumptions. We shall return to higher-order rules shortly.

Classical disjunction

Having verified that our suggested rules for disjunction (or their GE counterpart) are indeed harmonious, let us now see what they can do for us.

To begin with, the Law of Excluded Middle is now provable on no assumptions from $\vee$-I*, as required by the Fundamental Assumption; one just needs to assume $\neg A$ and $\neg\neg A$:

$$\dfrac{\begin{matrix}[\neg A, \neg\neg A]^1 \\ \vdots \\ \bot\end{matrix}}{A \vee \neg A}\ {\scriptstyle \vee\text{-I}^*, 1}\ .$$

Secondly, given *classical reductio*, or some equivalent rule, the standard rules for disjunction and the new ones can be shown to be interderivable.

> *Proof*: For $\vee$-I*, assume A and derive $A \vee \neg A$ by disjunction introduction. Now assume $\neg(A \vee B)$ and derive $\neg A$ by negation elimination. Similarly, derive $\neg B$. By *classical reductio*, that, given a derivation of $\bot$ from $\neg A$, one may one infer A and discharge $\neg A$, $A \vee B$ follows. For $\vee$-E*, given two derivations of $\bot$ by negation elimination from, respectively, A and $\neg A$ and B and $\neg B$, one may infer $\bot$ from $A \vee B$ by

disjunction elimination, discharging A and B. For $\vee$-I, assume A. Then assume $\neg A$ and $\neg B$. By negation elimination we infer $\perp$, so by $\vee$-I* we may infer $A \vee B$, discharging $\neg B$ vacuously. Similarly for the proof from B to $A \vee B$. Finally, we derive $\vee$-E by means of the following derived rules, both of which are classically valid:

(CP$_1$) If $A \vdash B$, then $\neg B \vdash \neg A$;

(CP$_2$) If $\neg B \vdash \neg A$, then $A \vdash B$.

The proof is as follows; we ignore side assumptions for the sake of simplicity (see also Smiley, 1996, p. 5). Assume $A \vee B$, $A \vdash C$, and $B \vdash C$. By CP$_1$, $\neg C \vdash \neg A$ and $\neg C \vdash \neg B$ follow. Notice, though, that $A \vee B$, $\neg A$ and $\neg B$ entail $\perp$ by $\vee$-E*, which in turn allows us to infer $\neg(A \vee B)$ from $\neg A$ and $\neg B$ by negation introduction. By the transitivity of deduction, we get that $\neg C$ entails $\neg(A \vee B)$. By CP$_2$, we can conclude $A \vee B \vdash C$. ■

Classical disjunction and the Fundamental Assumption

We are now finally in a position to reconsider, in the light of our non-standard rules for disjunction, some of the objections to the Fundamental Assumption we introduced in § 5.1. One objection, as the reader may recall, was that the Fundamental Assumption begs the question against the classical logician, since, for all we know, a proof of LEM cannot end by a step of $\vee$-I, contrary to what the Fundamental Assumption requires. We have just seen, though, that this worry is misplaced: given alternative rules for $\vee$, LEM *can* be proved canonically. A second objection was that there are countless everyday uses of 'or' where we are entitled to assert disjunctions without knowing, nor being in a position to know, which of the disjuncts is true (in § 5.1, we called these the non-constructive uses of 'or'). But again, the foregoing rules circumvent this objection. Provided that we know that both disjuncts cannot both be false, we are now in a position to introduce disjunctions without having to know, even in principle, which of the disjuncts is true.

7.4.2 CR as a structural rule

Our revised rules for disjunction allow us to prove LEM, but not DN. Indeed, they do not even allow us to derive the standard rule of disjunction elimination. How, then, are we to obtain classical logic? Following Milne (1994), one might be tempted to supplement $\vee$-I* and $\vee$-E* with CR, on the grounds that the latter rule is in harmony with the standard rule of negation elimination, this time interpreted with A as its major premise. However, as we have seen in § 5.7.3, this introduces a circularity in the inferentialist's account of our understanding of atomic statements: it makes our understanding of A dependent on our understanding of $\neg A$, which already depends on our understanding of A. Moreover, in view of Leblanc's theorem that classical logic does not admit of a separable formalization if negation is taken to be partially defined by CR, this option would at best give us a harmonious formalization of classical logic—not a separable one.

My proposed 'trick', then, is to

(i) take Tennant's and Rumfitt's suggestion that $\perp$ is best treated as a logical punctuation sign seriously (see *supra* § 6.5.5)

and

(ii) accept Schroeder-Heister's invitation to regard higher-order rules as legitimate.

I do not have space to defend either assumption—my main aim here is merely to present one more possible way for the classicist to meet the inferentialist's demands for proof-theoretic order. It is worth reminding, however, that at least *some* logicians with non-classical leanings are willing to assume (i).[30] As for (ii), if we accept Schroeder-Heister's account of assumptions as temporary expansions of our logical systems, it is very difficult to see why we should be able to temporarily expand our systems with axioms, but not with rules. To be sure, one might reject Schroeder-Heister's account. Yet, it seems to me, this would be a desperate move. The account seems to be a *description*—a correct one—of what we actually do when we assume something for the argument's sake.

With these two assumptions on board, CR can be rewritten as a *structural* rule, as follows:

$$\mathrm{CR}^{ho}, n\ \dfrac{\begin{array}{c}[A \Rightarrow \perp]^n \\ \vdots \\ \perp\end{array}}{A}\ .$$

[30]See e.g. Tennant (1999) and Steinberger (2009a). Steinberger develops an account of *ex falso quodlibet* for intuitionist logicians who do not think $\perp$ has content. In a nutshell, the idea is to treat the rule as a case of weakening on the right: if one can infer nothing from $\Gamma, \perp$, then one can infer anything from Γ.

If one can derive a contradiction from the assumption that A itself leads to a contradiction, one can discharge that assumption and infer A. This is, to be sure, to smuggle in classicality under the structural carpet. But notice that also the multiple-conclusions and the bilateralist logician are required to do so. The former assumes that we can assert multiple-conclusions—disjunctions—neither of whose disjunct is assertible. The latter also assumes a version of classical *reductio*, RED^*_{cl}, among her coordination principles.

This may invite the criticism that

> it almost seems that there is no way of attaining an understanding of the classical negation-operator if one does not have one already. (Dummett, 1991b, p. 299)

However, this objection would be too quick. For one thing, one can know CR^{ho} without knowing what negation—and, *a fortiori*, classical negation—means (one would simply have to know CR^{ho} without knowing negation introduction and negation elimination). For another, the intuitionist's contention was that classical logic cannot be regimented in a proof-theoretically acceptable way—i.e. classical logic is bound to be inharmonious and inseparable. The formalization of classical logic I am about to introduce, if acceptable at all, shows that *this* accusation is misplaced.

If we are granted (i) and (ii), we can now rewrite our proposed *impure* rules for disjunction into the following admittedly awkward, but *pure*, rules:

$$\vee\text{-I}^*_p, n\ \dfrac{\begin{matrix}[A \Rightarrow \bot, B \Rightarrow \bot]^n \\ \vdots \\ \bot\end{matrix}}{A \vee B} \qquad\qquad \vee\text{-E}^*_p\ \dfrac{A \vee B \quad A \Rightarrow \bot \quad B \Rightarrow \bot}{\bot}$$

Together with these two rules, CR^{ho} and the standard I- and E-rules for conjunction, implication, and negation afford a harmonious, sound and complete formalization of **CPL** (there is no need for *ex falso quodlibet*, which is just a special case of CR^{ho}, if we are allowed vacuous discharge of assumptions). Call this formalization **NHcp**.

Definition 23. *Formulae of* **NHcp** *are built up from atoms and from the standard binary connectives* $\wedge$, $\vee$, $\rightarrow$, *and the unary connective* $\neg$. *Absurdity* ($\bot$) *is not an atom (recall, we are working on the assumption that it is a logical 'punctuation sign'). The rules for* $\wedge$, $\rightarrow$, *and* $\neg$ *are the standard ones:* $\wedge$-*I*, $\wedge$-*E*, $\rightarrow$-*I*, $\rightarrow$-*E*, $\neg$-*I*, $\neg$-*E*. *The rules for* $\vee$ *are non-standard:* $\vee$-*I** *and* $\vee$-*E**. *There is a structural rule:* CR^{ho}.

As we shall see in what follows, **NHcp** is not only harmonious: it also satisfies the more demanding requirements of separability and conservativeness.

7.4.3 CPL normalized

We will now prove that every deduction in **NHcp** converts into a normal deduction, where, for present purposes, a normal deduction can be defined as follows:

Definition 24. *(Normal deduction) A normal deduction is a deduction which contains no* maximum formulae.

We have already introduced the reduction-step for our proposed rules for disjunction. The remaining conversions steps are standard (see e.g. Prawitz, 1965, Chapter 2). In order to prove our theorem, we first need to prove that we can restrict applications of CR^{ho} to the case where its conclusion is atomic. This is a routine exercise in the case of $\wedge$, $\rightarrow$, and $\neg$. Given the standard rules for disjunction, on the other hand, it is not possible to break an inference by CR^{ho} whose conclusion is $A \vee B$ down into inferences by the same rules whose conclusions are A or B. Things change once our new rules for disjunction are in place, however.

Theorem 5. *(CR^{ho}-restriction) Applications of* CR^{ho} *can always be restricted to the case where the conclusion is atomic.*

Proof: Recall, the *degree* of a formula is defined by the number of logical operators occurring in it. In the first half of the proof, we can now follow *verbatim* Prawitz's original proof (Prawitz, 1965, pp. 39-40). Let Π be a deduction in **NHcp** of A from Γ in which the highest degree of a consequence of an application α of CR^{ho} is d, where $d > 0$. Let F be a consequence of an application α of CR^{ho} in Π such that its degree is d but no consequence of an application of CR^{ho} in Π that stands above F is of degree d. Then Π has the form

$$\begin{array}{c} [F \Rightarrow \bot] \\ \Sigma \\ \underline{\bot} \\ F \\ \Pi_1 \end{array}$$

where $[F \Rightarrow \bot]$ is the set of derivations discharged by α and F has one of the following forms: $A \wedge B$, $A \rightarrow B$, $\neg A$, and $A \vee B$.[31] For negation, we just replace $\neg A$ with $A \Rightarrow \bot$. For $\wedge$ and $\vee$, we can remove the application of CR^{ho} by transforming Π into, respectively

[31] Prawitz's original proof only covers the first three cases (since, in his system, $\vee$ is defined).

$$
\dfrac{\dfrac{\begin{matrix}\left[\dfrac{\dfrac{A \wedge B}{A} \quad [\neg A]^1}{\perp}\right] \\ \Sigma \\ \perp\end{matrix}}{A}\,\mathrm{CR}^{ho},1 \qquad \dfrac{\begin{matrix}\left[\dfrac{\dfrac{A \wedge B}{B} \quad [\neg B]^2}{\perp}\right] \\ \Sigma \\ \perp\end{matrix}}{B}\,\mathrm{CR}^{ho},2}{A \wedge B}\,\wedge\text{-I}
$$
$$\Pi_1$$

and

$$
\dfrac{\dfrac{\begin{matrix}\left[\dfrac{\dfrac{[A]^1 \quad A \to B}{B} \quad [\neg B]^2}{\perp}\right] \\ \Sigma \\ \perp\end{matrix}}{B}\,\mathrm{CR}^{ho},2}{A \to B}\,\to\text{-I},1
$$
$$\Pi_1$$

The case for disjunction can now be dealt with as follows:

$$
\dfrac{\dfrac{\dfrac{\begin{matrix}\left[\dfrac{\dfrac{\dfrac{A \vee B \quad [A \Rightarrow \perp]^1 \quad [B \Rightarrow \perp]^1}{\perp}}{A} \quad [\neg A]^2}{\perp}\right] \\ \Sigma \\ \perp\end{matrix}}{A}\,\mathrm{CR}^{ho},2}{\perp}}{A \vee B}\,\vee\text{-I}^{*}_{p},1
$$
$$\Pi_1$$

The new applications of CR^{ho} in Π_1 have consequences of degrees less than d. Hence, by successive applications of the above procedures, we finally obtain a deduction of A from Γ in which every consequence of every application of CR^{ho} is atomic. ■

Theorem 6. *(Normalization) If $\Gamma \vdash_{\mathbf{NHcp}}$, then there is a normal deduction in **NHcp** of A from Γ.*

Proof (Prawitz, 1965, pp. 40-1): Let Π be a deduction in **NHcp** of A from Γ that is as described in Theorem 8. Let F be a maximum formula in Π such that there is no other maximum formula in Π of higher degree than that of F and such that maximum formulae in Π that stand above a formula occurrence side-connected with F (if any) have lower degrees than F. Let Π' be a reduction of Π at F. The new maximum formulae that may arise from this reduction are all of lower degrees than that of F. Moreover, Π' is still as described in Theorem 5. Hence, by a finite number of reductions, we obtain a normal deduction of A from Γ. ■

Theorem 7. *(Subformula property) Each formula occurring in a normal deduction Π of A from Γ is a subformula of A or of one of the formulae in Γ.*

Prawitz (1965, pp. 42-3) proves this result for his own formalization of **CPL**, which includes the rules for $\wedge$, $\rightarrow$, and CR ($\neg A$ is defined as $A \rightarrow \bot$). In Prawitz's system, the theorem holds for every formula in Π, "except for assumptions discharged by applications of CR^{ho} and for occurrences of $\bot$ that stand immediately below such assumptions". It is easy to show that Prawitz's proof carries over to **NHcp**, this time without exceptions.

Proof: Prawitz's original result carries over if we (i) add to his original system the standard rules for $\neg$ and our rules for $\vee$ (since reduction-steps are available in both cases), and (ii) we substitute CR by CR^{ho}. But now, consider the exceptions to Prawitz's original theorem, viz. that assumptions discharged by applications of CR and occurrences of $\bot$ that stand immediately below such assumptions may not be subformulae of either A or some of the formulae in Γ. Notice that it is a consequence of Prawitz's theorem that, if $B \Rightarrow \bot$ is an assumption discharged by CR^{ho} in a normal deduction of A from Γ, then B is a subformula of A or of some subformula of Γ. But then, also the assumption discharged by CR^{ho}, $B \Rightarrow \bot$ only contains subformulae of either A or some subformula of Γ, given that B is the only formula occurring in that assumption. As for the last exception, the problem disappears as soon as we treat $\bot$ as a logical punctuation sign.[32] ■

Theorem 8. *(Separation property) Any normal deduction only consists of applications of the rules for the connectives occurring in the undischarged assumptions, if any, or in the conclusion, plus possibly* CR^{ho}.

Proof: This follows at once from Theorem 7, by inspection of the inference rules. ■

[32]Notice that Prawitz's Lemma on permutative reductions (see Prawitz, 1965, pp. 49-51) need not be repeated here, since $\mathbf{Ncp}^+$ does not contain general elimination rules such as the standard rule of disjunction elimination.

This completes our quick presentation to **NHcp**'s proof-theoretic virtues. On minimal assumptions, and in a standard natural deduction framework, classical logic can be made consistent with the requirements of harmony and separability.

7.4.4 Categoricity again

Let us now turn to the question how classical inferentialists can deal with the Carnap problem, if they are not willing to adopt either a multiple-conclusions or a bilateralist framework.

First, suppose classical logicians take $\bot$ to have content. Then, just as intuitionists see introduction rules as specifying the canonical grounds for introducing complex statements, classical logicians may see them as as specifying necessary and sufficient conditions for asserting their conclusion. What is more, they may insist that, on their view, basic rules determine truth-conditions, i.e. necessary and sufficient conditions for truth. But then, if they adopt Prawitz's rules for $\bot$, they may claim that the empty rule of $\bot$-introduction shows us that there are no necessary and sufficient conditions for $\bot$'s truth, i.e. that $\bot$ is false on any valuation. Indeed, one can show that, if $\bot$ is false on any valuation, the standard rules for negation determine the standard satisfaction clause for negation,

$$(\neg)\ \ v(\neg A) = 1 \text{ iff } v(A) = 0,$$

on the assumption that, if A is false on any valuation, then the derivation from A to $\bot$ preserves truth on that valuation, i.e. the atomic rule $A/\bot$ holds. This assumption can in fact be *proved*.[33]

> *Proof*: If $v(A) = 0$, then both $A \vdash A$ and $A \vdash \neg A$ preserve truth on v. Hence, so does $\bot$, by negation elimination ${\scriptstyle \neg\text{-E}}\,\dfrac{A \quad \neg A}{\bot}$. ■

Now to the derivation of $(\neg)$. On the one hand, if $\neg A$ is true on a valuation v, and if $\bot$ is false on any valuation, A must itself be false on v. But this means that A and $\neg A$ cannot be both true on *any* valuation, if negation elimination is to preserve truth on every valuation. On the other, if A is false on a valuation, and if the derivation of $\bot$ from A preserves truth on that valuation, negation introduction requires that $\neg A$ be true on that valuation, for any valuation.

The argument goes through even if $\bot$ is treated as a logical punctuation sign. If $v(\neg A) = 1$, then the rule of negation elimination again ensures that $v(A) = 0$. Now suppose $v(A) = 0$. Then, on our assumption that one can use the atomic rule $A/\bot$, one can infer $\neg A$ by $\neg$-I, discharging A. Hence, $v(\neg A) = 1$, as required.

[33] See Garson (2010).

Now recall the Carnap problem for disjunction, that its standard introduction and elimination rules fail to determine the fourth line of its truth-table, that disjunctions all of whose disjuncts are false are themselves false. Suppose first that classical logicians are to keep the standard disjunction rules. Then, we have that, if both $v(A) = 0$ and $v(B) = 0$, both $A/\bot$ and $B/\bot$ hold. Hence, by one application of standard $\vee$-elimination, $\bot$ follows from $A \vee B$, discharging the case assumptions A and B. So $A \vee B$ is false on v, as required.

Suppose now classical logicians are willing to adopt our rules for disjunction. Then, the first three lines of $\vee$'s truth-table are taken care of by the standard rules for disjunction, which we have just shown to be derivable from $\vee$-I* and $\vee$-E*. For the last line, if our elimination rule for disjunction is to preserve truth, we have that $A \vee B$, $\neg A$ and $\neg B$ cannot all be true. But this means that, if $\neg A$ and $\neg B$ both hold, $A \vee B$ must be false, as required.

7.5 Conclusions

In this chapter, we have introduced three harmonious, separable, and categorical formalizations of classical logic: the multiple-conclusions ones, the bilateralist ones, and the more conservative one we have just presented. Each of these formalizations *prima facie* appears to be as proof-theoretically kosher as the standard formalizations of intuitionistic logic. Yet, as we have seen, a closer look reveals that each of these formalizations—in particular the first two—comes with a cost. The cost presents itself in the form of more or less defensible non-standard proof-theoretic assumptions: multiple-conclusions, the use of assertion and denial signs in the object language, and the treatment of absurdity as a logical punctuation sign.

Multiple-conclusions and bilateral formalizations of logic, I have suggested, are ultimately unacceptable. Multiple-conclusions formalizations appear not to be able to make sense of the normative force of valid inferences. Bilateral formalizations require that the speech-act of denial play the role of an external negation. Comparatively, the regimentation of classical logic we have introduced in § 7.4.2-3, **NHcp** is much more appealing, or at least so I have argued. On very minimal assumptions, it enjoys the same proof-theoretic properties as intuitionistic logic. All the same, as we have seen, it requires that we *assume* a structural version of classical *reductio*. The upshot is that we can be in harmony with classical logic, but the classicist needs to make a stronger effort than the intuitionist in order to satisfy the proof-theoretic requirements introduced in Chapter 5 and

Chapter 6. More precisely, she needs to assume some classical principle among her structural assumptions. Classicality need not disrupt the proof-theoretic properties of minimal and intuitionistic logic, but it would seem, it does not flow from the operational rules alone.

Chapter 8

Conclusions

How can we come to know logical truths, and facts about validity more generally? We have focused on two different answers to these questions. The first was that at least some logical beliefs have metaphysical consequences that clash with some of our deepest metaphysical convictions. Thus, for instance, our belief that the Law of Excluded Middle holds unrestrictedly is inconsistent with our belief that (i), for some region of discourse D, we do not believe that every $A \in D$ is decidable, but (ii) we believe every $A \in D$ is knowable, if true. If (i) and (ii) hold, we cannot know that the Law of Excluded Middle holds unrestrictedly. We observed in Chapter 2, however, that this only follows if we hold to *other* logical beliefs, such as our belief that the Law of Non-Contradiction holds unrestrictedly. But on which grounds should we hold on *some* logical beliefs at the expenses of others? Virtually *every* classical inference has been challenged in the last two or three decades. How can we take some challenges seriously, while dismissing some others?

Logical intuition

It may be tempting to respond: "Well, *obviously* the Law of Non-Contradiction is more certain than the Law of Excluded Middle: if you start doubting *that*, we can start doubting just about *any* logical belief". This answer, however, can hardly be satisfactory. In certain cases, 'obvious' logical principles have been challenged. For instance, Hartry Field's suggests that we adopt a conditional that validates *modus ponens*

$$\to\text{-E}\ \frac{A \qquad A \to B}{B}$$

but fails to satisfy principles such as *pseudo modus-ponens*

$$(A \wedge (A \to B)) \to B$$

contraction

$$\frac{A \to (A \to B)}{A \to B},$$

permutation

$$\frac{A \to (B \to C)}{B \to (A \to C)},$$

and *arrow introduction*

$$\begin{array}{c} [A]^n \\ \vdots \\ \end{array}$$
$$\to\text{-I}, n\ \frac{B}{A \to B}.$$

(see Field, 2008, Chapters 15-6). This means, however, that we are not given a set of rules sufficient for reasoning with $\to$. Not only does $\to$ fails to satisfy $\to$-I. It altogether lacks an introduction rule: we are not told how to introduce, in general, conditional statements. It is no mystery that, in Field's logic, the derivation of paradoxes such as the Liar and Curry's are blocked. It much less clear, however, whether the logic he recommends can plausibly be taken to govern the logical uses of expressions like 'if'.

Proof-theoretic criteria as a guide for selecting admissible rules

How to know whether we have gone too far? This leads us to considering a second possible answer to our initial question. The idea—essentially, one of Dummett's main ideas in the *Logical Basis of Metaphysics*—is to provide a framework for assessing what is to count as a legitimate logical principle, and what doesn't. In the second part of the thesis, we considered an *inferentialist* framework, and we offered reasons for thinking that it provides a reliable guide for assessing competing logics. The key thought was that the meaning of a logical expression is fully determined by the rules for its correct use: rules are, in some sense, *complete* with respect to correct uses. Depending on how we cash out this completeness assumption, we are then able to motivate two main proof-theoretic requirements: the local requirement of *harmony*, and the global requirement of *separability*.

Both requirements are widely held to be incompatible with classical logic, and are thus regarded with suspicion by classical logicians. I have argued in Chapter 7 that this suspicion is misplaced: there are formalizations of classical logic, even very natural ones, that satisfy both harmony and separability. By contrast, neither requirement is satisfied by Field's logic—nor, for that matters, is it satisfied by *paraconsistent* logics of paradox, as proposed, for instance, by Priest (2006a) and JC Beall (2009). By inferentialist standards, these logics do not qualify as logic. The reason is simple. These approaches all agree that the key for solving Curry's Paradox is to radically restrict $\rightarrow$-I.

At the same time, however, they all keep the unrestricted rule of $\rightarrow$-E. If harmony is to be a necessary condition for logicality, it follows that none of these approaches succeeds in defining a conditional. If we wish to hold on to the proof-theoretic criteria we introduced and motivated in Chapter 5 and Chapter 6, this means that something *else* has to give, when we are faced with paradoxes such as the Liar and Curry's.

To be sure, the main premise of this argument will be rejected outright by its proponents. Indeed, revisionary logicians like Field, Priest and Beall are very likely to take the incompatibility of the inferentialist approach to logic with their own theories of truth to be a sufficient reason to rule out inferentialism. If inferentialists wish to be able to persuade *these* logicians, they need to resort to a different kind of argument—one to the effect that weakening the logic of the *connectives* is such a way as to make their I- and E-rules disharmonious still leaves us with what these theorists agree we must avoid, viz. *triviality*. But this, I wish to argue, is an almost *trivial* task.

The Validity Curry

The following biconditional captures an essential feature of our notion of validity, viz. that an argument $\langle A, B\rangle$ is valid if and only if $A \vdash B$, where '$\vdash$' is the consequence relation:

$$\text{(VAL) Val}(\ulcorner A \urcorner, \ulcorner B \urcorner) \Leftrightarrow A \vdash B,$$

where 'Val' is a validity predicate. Indeed, let us say, with most logic textbooks, that a two-place relation R is expressed by an open wff $\Phi(x, y)$ with two free variables in a language L if and only if, for any a, b, aRb if and only if $\Phi(\ulcorner a \urcorner, \ulcorner b \urcorner)$ holds. Then, VAL *must* hold if '*Val*' is to *express* validity. And here is the problem: given the standard structural properties of the consequence relation, *irrespective of one's interpretation of the connectives*, VAL is no less paradoxical than the corresponding biconditionals for 'true'. For consider the following rules—a proof-form version of VAL (see Priest, 2010, p. 128):

$$\dfrac{\begin{array}{c}[A]^n\\ \vdots\\ B\end{array}}{\mathsf{Val}(\langle A,B\rangle)}\,\text{Val-I}, n \qquad \dfrac{\mathsf{Val}(\langle A,B\rangle)) \quad A}{B}\,\text{Val-E}$$

And now let Π be $Val(\ulcorner\Pi\urcorner,\ulcorner A\urcorner)$, where A is an arbitrary sentence. We may then reason as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{[\Pi]^1}{\mathsf{Val}(\langle\Pi\rangle,\langle A\rangle)} \quad [\Pi]^1}{A}\,\text{Val-E}}{\mathsf{Val}(\langle\Pi\rangle,\langle A\rangle)}\,\text{Val-I}, 1 \qquad \Pi}{A}\,\text{Val-E}$$

We may call this the *Validity Curry*. Crucially, it does not involve any *operational rule*, though it makes use of the structural rule of *contraction*:

$$\text{If } \Gamma, A, A, \vdash B, \text{ then } \Gamma, A \vdash B,$$

here present in the form of multiple discharge of assumptions. Far from being inconsistent with our best theories of truth, the inferentialist account of logic tells us something that, as we have just seen, we may independently know: the paradoxes are not to be solved by revising the logic of the *connectives*.[1]

Inferentialism and structural assumptions

Let us now move on to considering a potential limitation of the inferentialist framework we have partially sketched, viz that it is—perhaps inherently—incomplete: it can at best justify *operational* rules, and it is silent about *structural* rules. Some authors have argued that this is a serious defect—one that casts doubts on the entire inferentialist approach to logic. There are two problems.

The first is that, if meaning is correct use, structural rules *must* have semantic import. And does not this contradict the inferentialist claim that the meaning of a logical expression is *fully* determined by the rules for its correct use? I do not think so. For one thing, structural rules do not affect the truth- or proof-conditions

[1]In his book, Field (2008, p. 305 ff.) considers a different validity paradox

$$(\Sigma)\ \neg\mathsf{Val}(\langle\top\rangle,\langle\Sigma\rangle).$$

He argues, however, that it is not "particularly compelling". Priest (2010) objects that it *is* compelling, and shows that, in order to solve the problem, Field must give up the assumption that every argument is either valid or invalid—otherwise, one can easily derive $\top \vdash \bot$ (Priest, 2010, p. 128). Both Field and Priest, however, overlook the Validity Curry, which emphatically may *not* be blocked by dropping the assumption that every argument is either valid or invalid.

of the logical operators. For another, they describe features of use that apply to *all* expressions—not just to the logical ones. This suggests, however, that their semantic import can be safely set aside.

The second problem is that, if structural rules cannot be proof-theoretically justified, there is a proof-theoretic lacuna in the inferentialist program. Thus, Graham Priest writes:[2]

> [...] the introduction and elimination rules are superimposed on structural inferential rules; for example, the transitivity of deducibility (deductions may be chained together to make longer deductions). Such structural rules are not inevitable, and the question therefore arises as to how *these* rules are to be justified. This becomes patently obvious if the proof-theory is formulated as a Gentzen system where the structural rules are quite explicit [...]. One needs to justify which structural rules one accepts (and which ones one does not), and there is no evident proof-theoretic way of doing this. If [...] one cannot justify every feature of a proof-theory syntactically, the only other possibility would seem to be some semantic constraint to which the rules must answer. (Priest, 2006a, p. 179)

Priest argues that, if proof-theoretic constraints cannot justify the whole of logic, then one should instead adopt a semantic approach to logic—one on which both structural and operational rules can be justified. But I fail to see why this conclusion should follow from Priest's premises. Structural rules define general features of the relation of deducibility: features that allow us to reason with atomic statements, and that are independent of one's choice of the logical vocabulary. Operational rules define the inferential role of the logical vocabulary, given a *context of deducibility*. Why should these different kinds of rules be justified in the same way? Priest does not say.

My hunch is that structural rules reflect a mix of logical and metaphysical commitments. Logical commitments are represented by principles such as weakening:

$$\text{If } \Gamma \vdash B, \text{ then } \Gamma, A \vdash B$$

contraction:

$$\text{If } \Gamma, A, A, \vdash B, \text{ then } \Gamma, A \vdash B$$

and cut:

[2] See also Hjortland (2008).

$$\text{If } \Gamma \vdash B \text{ for all } B \in \Delta, \text{ and } \Delta \vdash A, \text{ then } \Gamma \vdash A.$$

Metaphysical principles, on the other hand, can be found in the formalizations of classical logic we considered in Chapter 7. For the multiple-conclusions logician, depending on how multiple-conclusions are interpreted, they are present either in the form of disjunctions neither of whose disjuncts is assertible, or in the form of existential statements to the effect that *some* A holds, for which we may not be able to provide a witness. For the bilateral logician, they take the form of co-ordination principles, that allow us to assert A if $-A$ leads to absurdity. Likewise, in our system **NHcp**, they take the form of a structural rule that allows us to infer A if $A \Rightarrow \bot$ cannot be consistently assumed.

Is intuitionistic logic the right logic?

We have criticized at some length multiple-conclusions and bilateral formalizations of logic in Chapter 7. Multiple-conclusions logicians have hard time explaining why logical arguments have to be formalized with multiple conclusions, while the arguments we give in our everyday life always seem to be arguments for at most *one* conclusion. Bilateralists, on the other hand, appear to have potentially good reasons for taking denial as a primitive speech act. However, once we look at the bilateralist formalizations, we soon realize that what they take to be a sign for denial is in fact *external negation*—at least if denied sentences are to be the kind of things we *assume*.

If these were the only proof-theoretically kosher formalizations of classical logic available, and if the inferentialist approach to logic sketched in Chapters 4-6 is broadly along the right tracks, then intuitionists would have a compelling argument from proof-theoretic constraints—essentially, harmony and separability—against classical logic. Moreover, they would be in a position to show why the Basic Revisionary Argument does not after all miss its target. On an inferentialist approach to logic, *minimal logic* truly deserves its name: it is difficult to see how it can be weakened, without thereby calling into question logical inferentialism itself.[3] But then, this would allow anti-realists who subscribe to an inferentialist account of logic to claim, with some plausibility, that the reason why LNC is epistemically in better standing than its dual, LEM, is that the former, but not the latter, is valid in virtue of the meaning of the logical vocabulary. The Basic Revisionary Argument would *be* basic, as its proponents, from Brouwer to Wright, contend.

[3]For a different argument for the same conclusion, see Hale (2002).

These two lines of reasoning rest on the premise that we cannot be in harmony, so to speak, with classical logic. But is this premise available even if classical logic is regimented by means of something that at least includes our system **NHcp**? This is a delicate question.

NHcp satisfies both harmony and separability. Hence, it is proof-theoretically kosher, at least if classical logicians are willing to accept higher-order rules, and to treat $\perp$ as a logical punctuation sign.

It is worth stressing, however, that the logical strength of this system almost entirely relies on our structural version of classical *reductio*,

$$\begin{array}{c} [A \Rightarrow \perp]^n \\ \vdots \\ \mathrm{CR}^{ho}, n\ \dfrac{\perp}{A}\ . \end{array}$$

Without CR^{ho}, our revised rules for disjunction do not even allow us to derive the standard rule of disjunction elimination, i.e. proof by cases. The difficulty is that, *qua* structural assumption, CR^{ho} cannot be justified one the basis of proof-theoretic considerations alone.

This observation may invite to the conclusion that, in **NHcp**, LEM is—unlike LNC—not valid in virtue of the meaning of the logical vocabulary, together with the standard structural assumptions (weakening, contraction, and cut). But then, if this reasoning is correct, it would seem that anti-realism are now in a position to rescue the Basic Revisionary Argument from the main objection we raised in Chapter 2, viz. that it validates a parallel argument for conclusions that are unwelcome to classicists and intuitionists alike. For if the conditions for the applying the Basic Revisionary Argument are satisfied, it will now be difficult to argue, as we did in Chapter 2, that the original argument can be turned either into an argument against LNC, or into an argument for Dialetheism. As the reader may recall, the challenge to the intuitionist we raised at the end of § 2.4.2 was to provide reasons for accepting $\mathcal{K}_n(\mathrm{DEC}^*)$, the claim that we presently know that, for every statement, it is not the case that both it and its negation are unknowable, that are not as strong as to be reasons for accepting $\mathcal{K}_n(\mathrm{DEC})$, the claim that we presently know that, for every statement, either it or its negation is knowable. If classical logic is formalized in the way I have suggested, this challenge may be met. For now, at least from an inferentialist perspective, there *is* an asymmetry between our derivations of $\mathcal{K}_n(\mathrm{DEC})$ and $\mathcal{K}_n(\mathrm{DEC}^*)$. The former requires structural assumptions—in **NHcp**, a use of CR^{ho}—that are strictly stronger than the ones required for deriving the latter. Thus, one may coherently accept $\mathcal{K}_n(\mathrm{DEC}^*)$ *and* reject $\mathcal{K}_n(\mathrm{DEC})$, on the grounds that DEC*, but not DEC, is valid in virtue of the meaning of the logical constants alone.[4]

Stephen Read (2000) takes the availability of harmonious and separable formalizations of classical logic to suggest that the real intuitionistic challenge against classical logic is not that of providing a proof-theoretically acceptable formalization of classical logic. Rather, Read maintains, the intuitionist should challenge the classicist's justification, if there is one, for her unrestricted commitment to the Principle of Bivalence. He writes:

> The constructivist can still mount a challenge to classical logic. But we now see where that challenge should be concentrated—and where it is misguided. The proper challenge is to Bivalence, and to the classical willingness to assert disjunctions, neither of whose disjuncts is separately justified [...]. (Read, 2000, pp. 151-2)[5]

But although the intuitionist's challenge *may* be mounted to the Principle of Bivalence and to the Law of Excluded Middle, as we have seen in Chapter 2 and 3, the foregoing considerations suggest that its ultimate target must be the structural assumptions the classicist is seemingly obliged to rely upon—assumptions that, as we have seen, are strictly stronger than the ones required for intuitionistic logic.

To be sure, *pace* Dummett, Prawitz, and Tennant, the fact that classical logic requires stronger structural assumptions is not a good reason to dismiss it. *If* these assumptions can be independently justified, then there is nothing wrong with classical logic, from a proof-theoretic point of view. Can these assumptions be justified? And can they be coherently maintained? This is where revisionary arguments such as the Basic Revisionary Argument, Wright's argument from vagueness, and Dummett's so-called argument from indefinite extensibility can play a crucial role. But we shall not take matters further.

---

[4]It may be objected that CR^{ho} is not required, in **NHcp**, in order to derive $\mathcal{K}_n$(DEC) from WVER and LEM. But it is. As our natural deduction presentation of the Basic Revisionary Argument makes clear (see *supra*, § 2.3.3, fn. 35), the derivation of $\mathcal{K}_n$(DEC) from WVER and LEM requires an application of the standard rule of disjunction elimination, i.e. proof by cases. However, in **NHcp** proof by cases can be derived from our proposed rule of disjunction only if CR^{ho} is in place.

[5]One may consistently reject Bivalence while being willing to assert disjunctions neither of whose disjuncts are separately assertible—supervaluationist semantics for vagueness and future contingent discourse precisely allow us to do that—see e.g. Keefe (2000) and MacFarlane (2003). Read's claim, then, must be that the proper constructivist challenge has to be directed towards the *logical* Law of Excluded Middle, as opposed to the *semantic* Principle of Bivalence. See e.g. Wright (2001).

Appendices

PAGINATED BLANK PAGES ARE SCANNED AS FOUND IN ORIGINAL THESIS

NO INFORMATION MISSING

Appendix A

Manifestability and decidability

Timothy Williamson has recently argued that Dummett's challenges establish too strong a result, viz. that the central semantic concept of a theory of meaning must be *decidable*. In this Appendix, I briefly consider Williamson's objection, and argue that it rests on a mistaken identification of the notions of truth and of a truth-maker. The general point is that semantic anti-realism is not as foolish a doctrine as Williamson would make it seem: whatever Dummett's arguments may establish, they do not require the decidability of the central semantic concept of a theory of meaning. Williamson's attack on Dummett's challenges is twofold. He first argues that the prospects for assertibility-conditional theories of meaning are slim. He then contends that the result of Dummett's challenges prevents anti-realists from adopting a *truth*-conditional account of meaning instead, even when the notion of truth is anti-realistically construed. I consider each of these two arguments in turn.

A.1 Williamson's first argument

Dummett once wrote, in his early paper *Truth*:

> *We no longer explain the sense of a statement by stipulating its truth-value in terms of the truth-values of its constituents, but by stipulating when it may be asserted in terms of the conditions under which its constituents may be asserted.* (Dummett, 1959, pp. 17-8)

In light of the Acquisition and the Manifestation challenges, Dummett suggests, we ought to abandon a conception of meaning as truth-conditions in favour of a conception of meaning as assertibility-conditions. For instance, instead of focusing on what makes sentences like

(1) $68 + 57 = 125$

or

(2) Spinach is tasty

true, one should rather focus on the circumstances under which they may be correctly asserted—their *assertibility-conditions*. In the case of (1), assertibility-conditions will include possession of a *proof*, or of a means of producing a proof, to the effect that the sum of 68 and 57 is 125. In the case of (2), assertibility-conditions will involve, presumbably, reference to the fact that I, or some culinary expert in the linguistic community, find spinach tasty. And so on.

In Williamson's view, however, Dummett has failed his own methodological assumptions. The problem, he argues, is that anti-realists have not really developed an anti-realist, assertibility-conditional semantics, to be substituted to the realist, truth-conditional semantics. He makes the point in his article *Must do Better*:

> In 1957, Michael Dummett was about to open his campaign to put the debate between realism and anti-realism, as he conceived it, at the centre of philosophy. The campaign has a strong methodological component. Intractable metaphysical disputes (for example, about time) were to be resolved by being reduced to questions in the philosophy of language about the proper form for a semantic theory of the relevant expressions (for example, tense markers). The realist's semantic theory would identify the meaning of an expression with its contribution to the truth-conditions of declarative sentence in which it occurred. The anti-realist's semantic theory would identify the meaning with the expression's contribution to the assertibility conditions of those sentences. Instead of shouting slogans at each other, Dummett's realist and anti-realist would busy themselves in developing systematic compositional semantic theories of the appropriate type, which could then be judged and compared by something like scientific standards. But that is not what happened. (Williamson, 2008, p. 281)

Williamson correctly interprets Dummett has having proposed that metaphysical disputes be reduced to *semantic* ones. However, the way Williamson thinks Dummett and his followers are suggesting to actually adjudicate different theories of meaning, together with the different metaphysical assumptions underwriting them, is more controversial. In Williamson's view, Dummett and his followers originally proposed that realists and anti-realists should busy themselves construing, respectively, realist and anti-realist theories of meaning. Then, the competing theories would be "judged and compared by something like scientific standards". But, Williamson continues, this is not what happened. Dummett and his followers, he argues, have spent too much time developing philosophical objections to truth-conditional theories of meaning, instead of attempting to do serious work on an anti-realist alternative. But this is a problem, Williamson suggests, since we are now in no position to adjudicate between realist and anti-realist semantics: one of the two contending theories is simply yet to be worked out. By contrast, the truth-conditional framework has had, so far, a vast empirical success. Yet, anti-realists do not seem to be very much impressed, or at least so Williamson argues:

> Surprisingly [...] most participants in the Dummett-inspired debates between realism and anti-realism have shown little interest in the success of truth-conditional semantics, judged as a branch of empirical linguistics. (Williamson, 2008, p. 282)

In conclusion, the prospects for an assertibility-conditional semantics are hard to assess, from an empirical point of view: with a very few exceptions, no systematic assertibility-based semantic theory has been developed so far. Williamson writes:

> As for assertibility-conditional semantics, it began with one more or less working paradigm: Heyting's intuitionistic account of the compositional semantics of mathematical language in terms of the condition for something to be a proof of a given sentence. The obvious and crucial challenge was to generalize that account to empirical language: as a first step, to develop a working-conditional semantics for a toy model of some small fragment of empirical language. But that challenge was shirked. Anti-realists preferred to polish their formulations of the grand program rather than getting down to the hard and perhaps disappointing task of trying to carry it out in practice. (*Ibid.*)

Williamson concludes, not without sarcasm, that

> the suggestion that the program's almost total lack of empirical success in the semantics of natural language might constitute some evidence that it is mistaken in principle would be dismissed as crass. (*Ibid.*)

Williamson's argument from empirical success crucially relies on the assumption that Dummett and his followers originally thought that the dispute between the realist and the anti-realist should be adjudicated by testing the empirical success of, respectively, realist and anti-realist theories of meaning. But is this assumption correct?

A.2 Response to Williamson first argument

There are at least two problems with Williamson's argument. The first is that Williamson is misrepresenting Dummett's own program. In Williamson's view, Dummett's strategy to solve the realism/anti-realism debate involves the construction of two competing theories of meaning, the realist and the anti-realist one, which should be adjudicated "by something like scientific standards". But this is not Dummett's methodology. Dummett has always been adamant that the difficulties he was finding in realist theories of meaning were difficulties *in principle*. He writes, in his *What is a Theory of Meaning? II*:

> The difficulties that face the construction of [a realist theory of meaning] are not difficulties of *detail*: they are difficulties of *principle*, that face us at the very outset of the enterprise. (Dummett, 1976, p. 68)

To be sure, Dummett may well be wrong about this, and Williamson may well be right in saying that realist and anti-realist theories of meaning should be assessed by "something like scientific standards". The more serious problem, however, is that Williamson is shooting at the wrong target.

Already in the 70's, Dummett argued that the real point at issue in the realism/anti-realism debate is not whether to adopt a truth-conditional theory of meaning. Rather, it concerns the notion of truth involved in our truth-conditional meaning theory. Here is what Dummett says in the *Introduction* to *Truth and Other Enigmas*, of 1978:

> On the way of putting [things] I adopted, one first proposes explaining meaning, not in terms of truth, but in terms of the condition for correct assertion, and then declares that, for statements whose meaning is so explained, the only admissible notion of truth will be one under which a statement is true when and only when we are able to arrive at a position in which we may correctly assert it. But, in that case, it would have been better first to state the restriction on the application of 'true', and then to have held that the meaning if a statement is given by the condition for it to be true in this, restricted, sense of 'true'.

> This would, indeed, have meant rejecting, rather than embracing, the redundancy theory: the point would now be expressed by saying that acceptance of the principle of bivalence renders impossible the required account of the connection between the condition for a sentence to be true and the use of that sentence. Thus I should now be inclined to say that, under any theory of meaning whatever—at least, any theory of meaning which admits a distinction like that Frege drew between sense and force—we can represent the meaning (sense) of a sentence as given by the condition for it to be true, on some appropriate way of contruing 'true': the problem is not whether meaning is to be explained in terms of truth-conditions, but of what notion of truth is admissible. (Dummett, 1978b, p. xxii)

And here is a more recent quote:

> We ought not [...] repudiate the formula 'To understand a sentence is to know what it is for it to be true'; rather, we must enquire with what conception of truth we must replace that held (but never clearly explained) by the truth-conditional theorist. (Dummett, 2006, pp. 64-5)

If the disagreement between the realist and the anti-realist concerns what notion of truth should be adopted in a theory of meaning, and not whether some notion of truth should be adopted, Williamson's argument from empirical success has no bite: Dummett's anti-realist might perfectly agree with Williamson's distrust of assertibility-conditional approaches to meaning. Granted, Williamson might ask at this point whether the large amount of work to which he refers, doing semantics in the truth-conditional framework, survives if one interprets the notion of truth involved in an anti-realist way. Is such work, for example, free of any assumption of bivalence? Tempting though it may be, we will not attempt to settle this issue here. Instead, we will turn to what Williamson himself has to say about the possible rejoinder we have just sketched.

A.3 Williamson's second argument

Williamson considers the possible reply we have just outlined on the anti-realist's behalf. He writes:

> Some participants in the debate denied any need for anti-realists to develop their own semantic theories of a distinctive form. For, it

> was proposed, anti-realists could take over truth-conditional semantic theories by interpreting "true" to mean assertible or verifiable at the limit of enquiry, or some such epistemic account of truth. (Williamson, 2008, pp. 282-3)

However, he dismisses it on the grounds that "it is quite contrary to Dummett's original arguments" (p. 283). He takes such arguments, Dummett's meaning-theoretic challenges, to

> require the key semantic concept in the anti-realistic semantics, the concept in terms of which the recursive compositional clauses for atomic expressions are stated, to be decidable, in the sense that the speaker is always in a position to know whether it applies in a given case. That is what allows anti-realists to claim that, unlike realists, they can give a non-circular account of what understanding a sentence consists in: a disposition to assert it when and only when its assertibility-condition obtains. (*Ibid.*)

But, he continues, "it is supposed to be common ground between realists and anti-realists that truth is not always decidable" (*Ibid.*). Peter Pagin gives an argument along essentially the same lines:

> The Dummett-Prawitz intuitionistic alternative to truth as the central semantic concept is, however, not provability, since it is not in general *decidable* whether a sentence is provable. Instead, the unary concept of being provable is replaced by the relation *x is a proof of A*. It is normally assumed that when presented with a particular object or construction a and a sentence A, we are able to tell whether or not a is a proof of A. (Pagin, forthcoming, p. 8)

Williamson's and Pagin's argument needs some unpacking. There are two main premises:

(i) The upshot of Dummett's challenges is that the central semantic concept of a theory of meaning must be decidable, i.e. that it must always be possible to decide, of any given statement, whether it falls under the concept or not (say whether it is true or false/assertible or not assertible, or what have you),

(ii) Everybody agrees that truth is *not* decidable.

It follows from (i) and (ii) that

(iii) Truth cannot be the central semantic concept of a theory of meaning.

The crucial premise is, of course, (i).

A.4 Response to Williamson's second argument

Why does Williamson think that Dummett's challenges, if successful, establish that the central concept of a theory of meaning must be decidable? We have seen that, in his view, the decidability of the central concept of an anti-realist theory of meaning

> is what allows antirealists to claim that, unlike realists, they can give a non-circular account of what understanding a sentence consists in: a disposition to assert it when and only when its assertibility-condition obtains. (Williamson, 2008, p. 283)

According to Williamson, anti-realists are thus committed to the following biconditional:

(ARU_{TW}) $\forall\varphi\forall S(S$ understands $\varphi \leftrightarrow (S$ is disposed to assert $\varphi \leftrightarrow \varphi$ is assertible)),

where 'φ' ranges over meaningful statements and 'S' ranges over (competent) speakers. This biconditional is either false or inaccurate, however. If φ's assertibility-conditions obtain just in case *there is* a proof of φ, Williamson's biconditional does not hold: for some φ we do understand, φ's assertibility-conditions might obtain, and yet we might not be disposed to assert it, say because we have not yet discovered a proof of φ. Likewise, if φ's assertibility-conditions obtain just in case *we presently have* a proof of φ, the right-to-left direction of Williamson's biconditional may be true, but it is quite unclear why anti-realists should accept the converse direction. There are at least two reasons. For a start, anti-realists themselves have pointed out some very unintuitive consequences of the view. For instance, from the obtaining of the assertibility conditions of

(3) There are infinitely many twin primes

one could infer, via Williamson's suggested anti-realist account of understanding, that someone knows a great deal about prime numbers! Second, we have already seen in Section 2.2.1 that anti-realists are willing to adopt a different, seemingly more reasonable, account of understanding: that to understand φ is to be disposed to recognize a proof of φ, when presented with one. *This* is, in their view, what allows them to give a non-circular account of the manifestability of understanding. We may represent the account as follows:

(ARU) $\forall\varphi\forall S$(S manifests understanding of $\varphi \leftrightarrow$ (($\varphi \rightarrow$ it is (metaphysically) possible that, at some time t, S recognizes a correct argument for φ if presented with one) $\wedge$ ($\neg\varphi \rightarrow$ it is (metaphysically) possible that, at some time t, S recognizes a correct refutation of φ if presented with one))).[1]

As we saw, the account requires that a statement is true only if there is a proof of, or a correct argument for, it. But does it also require that the central concept of the anti-realist theory of meaning be decidable?

As far as I can see, there are no reasons for answering this question in the affirmative. To be sure, ARU requires that the relation 'x is correct argument for y' be *decidable*: for any speaker S and for any statement φ he or she understands, S must always be in a position, at least in principle, to decide whether a given object Π is a correct argument for φ, when presented with Π. However, it does not follow from this that the central concept of an anti-realist theory of meaning must itself be identified with the relation 'x is a correct argument for y'. The central concept may well be (an epistemically constrained notion of) truth, as Dummett claims, and meanings may well be equated with truth-conditions. Then, knowledge of truth-conditions can be manifested in the following sense: for every φ a competent speaker S understands, *if* S were presented with a correct argument Π for φ, S would be disposed to recognize Π as a correct argument for φ. Thus, the idea is, I can manifest my understanding of Goldbach's Conjecture by recognizing either a proof or a disproof of it (depending on whether the conjecture is true or false), if presented with one. Hence, knowledge of meaning is manifestable in the very weak sense specified by ARU: if φ is true (false), then it is (metaphysically) possible that, at some time, S recognizes a correct argument for φ ($\neg\varphi$). By contrast, on a *realist* account of truth, understanding is not even manifestable in this very weak sense: φ may be true (false), one could *never* be disposed to recognize an argument for φ ($\neg\varphi$), since the argument could simply not be there. In short: if, for the anti-realist, proofs and correct arguments are what make statements true, manifestability requires the decidability of the anti-realist's notion of a truth-maker, but not that of truth itself.

[1]See also Tennant (1997, p. 199).

It might be objected that the foregoing notion of manifestability is too weak for the anti-realist's purposes. Thus, for instance, Peter Pagin writes:

> The underlying reason for [the claim that knowledge of meaning must be publicly manifestable] is Dummett's view that successful communication requires that the communicators know that they understand the linguistic expressions the same way. It is not enough that you and I in fact mean the same by the same expressions, for then we cannot make sure that we understand each other. (Pagin, forthcoming, p. 6)

Pagin observes that, in Dummett's view, successful communication requires that speakers be aware that they mean the same thing by the same expressions, when they do. He offers the following quote of Dummett's in support of his claim:

> If language is to serve as a medium of communication, it is not sufficient that a sentence should in fact be true under the interpretation placed on it by one speaker just in case it is true under that placed on it by another; it is also necessary that both speakers should be aware of the fact. (Dummett, 1978a, p. 132)

This is too strong a requirement, however. In general, we do not need to *know* that we do understand each other, in other to communicate successfully. We might reasonably *assume* that we do so—an assumption, however, we are ready to drop as soon as our interlocutor starts using the expressions of the language in unexpected ways. Pagin nevertheless insists:

> The idea that successful communication cannot simply rest on an act of faith, that it requires knowledge of mutual understanding, not just belief, is what motivates the manifestability requirement. (Pagin, forthcoming, p. 6)

But this is not the ultimate motivation of Dummett's manifestability requirement. Rather, the motivation is that it must be possible in principle to manifest differences of understanding in our linguistic use (see e.g. Dummett, 1973b). Consider again Goldbach's Conjecture. On the foregoing anti-realist account of understanding, two speakers could manifest their common understanding of the conjecture by recognizing a proof, or a disproof, of the conjecture, if presented with one. By contrast, this possibility seems to be foreclosed to the realist: on a realist view, the conjecture may be true, or false, without there being a proof, or a disproof, of it.

In conclusion, what the realist should show is that an epistemic concept of truth satisfying Dummett's requirement of manifestability cannot be adopted for the purposes of a truth-conditional semantics. To the best of my knowledge, however, such an argument has yet to be given. Dummett's challenges, if sound, establish that truth must be epistemically constrained: a statement is true only if there is a correct argument for it. But, *pace* Williamson, the challenges do *not* establish, if sound, that the central semantic concept of a theory of meaning must be *decidable*.

Appendix B

The Whole Discourse Argument

In this Appendix, we briefly introduce Neil Tenant's own argument for intuitionistic logic—the Whole Discourse Argument, as he calls it. I show that the argument essentially reduces to the Basic Revisionary Argument.

B.1 The central inference

In Chapter 7 of *The Taming of the True, Long Live the Manifestation Argument*, Tennant sets out "a completely new argument proceeding from the Manifestation Requirement" (Tennant, 1997, p. 195). He calls it the Whole Discourse argument. The argument "invokes the effective undecidability of the whole discourse" (*Ibid.*), where the effective undecidability of a discourse D is defined as the claim that "there is no (or at least, we possess no) effective method for determining φ's truth-value" (Tennant, 1997, p. 183), for every $\varphi \in D$. Tennant's argument essentially aims at establishing what he calls "the central inference": a four pages proof that the Manifestation Requirement and what Tennant calls 'constructive bivalence', i.e. the claim that every statement in D or its negation has a constructive proof, jointly yield that all statements are effectively decidable—see Tennant (1997, pp. 206-9). In symbols:

(CI_1) $\mathsf{MR}, c\mathsf{BIV}_D \vdash e\mathsf{DEC}_D$,

where BIV_c expresses constructive bivalence and $e\mathsf{DEC}_D$ is the claim that there is an effective method whose application would determine, in a finite amount of steps, whether, for any statement φ in D, φ is true or false. The general idea, then, is that, together with a "proof of undecidability" for D, one between MR and $c\mathsf{BIV}_D$ will have to go.

B.2 Tennant's proof of the central inference

Tennant offers a rather complicated, and, what is more important, controversial, proof of the central inference. First, he defines the decidability of a discourse D as the existence of a procedure μ such that, for all $\varphi \in D$:

(i) μ is total;

(i) μ is effective;

(i) $\mu(\varphi) = T \to (\varphi \text{ is true}) \wedge \mu(\varphi) = F \to \neg(\varphi)$ is true.

He then proceeds to show that constructive bivalence for D and the Manifestation Requirement yield D's decidability. The crucial step is to prove that μ is in fact total. Tennant's proof is as follows:

> By $c\mathsf{BIV}_D$,
>
> φ is true or $\neg(\varphi$ is true)
>
> Assume first that φ is true. Thus there is (constructively) some truth-maker Π for φ. Find it, and present it to the speaker. By MR, the speaker is able to recognize Π as showing that the truth-condition for φ obtains, or at least is able to get himself into a position where he can so recognize. That is, the speaker will be able to return the verdict T on φ. Therefore
>
> φ is true $\to \mu(\varphi) = T$
>
> Now assume that it is not the case that φ is true. Thus there is (constructively) some falsity-maker Σ for φ. Find it, and present it to the speaker. By MR again, the speaker is able to recognize Σ as showing that the truth-condition for φ does not obtain, or at least is able to get himself into a position where he can so recognize. That is, the speaker will be able to return the verdict F on φ.
>
> $\neg(\varphi$ is true) $\to \mu(\varphi) = F$
>
> It now follows by $c\mathsf{BIV}_D$ that μ as defined is total. (Tennant, 1997, pp. 205-6)

This argument is problematic, however. As Jon Cogburn (2003) observes, and as Tennant himself explicitly acknowledges, this argument only works on a *constructivist* understanding of the existential quantifier. On a classical understanding, the existence of some truth (falsity) maker for φ does not guarantee that we be able to find it, and present it to the speaker. But of course this vitiates Tennant's proof: for one cannot already assume an intuitionistic understanding of some logical constants in an argument for the adoption of intuitionistic logic! How to solve the problem? Fortunately for the anti-realist, a snappier and relatively uncontroversial proof of the central inference is available—one that makes use of inference rules that are accepted by intuitionists and classicists alike.

B.3 A snappy proof of the Central Inference

There is something deeply puzzling about Tennant's set up. One of the key premises of his central inference, $c\mathsf{BIV}_D$, really is the conjunction of two claims: the classical Principle of Bivalence, and the claim that every true statement in D is knowable. Tennant himself concedes this point:

> $c\mathsf{BIV}_D$ [must be] understood as involving a constructive notion of truth. In other words, the anti-realist must presuppose, in order to establish the central inference, that *all truths are knowable* (Tennant, 1997, pp. 213-4; Tennant's terminology is adapted to ours)

It follows that Tennant's proof of the central inference should contain one more premise:

(CI_2) $\mathsf{MR}, \mathsf{KP}, \mathsf{BIV}_D \vdash e\mathsf{DEC}_D$,

where, recall, $e\mathsf{DEC}_D$ expresses the claim that discourse D is effectively decidable, i.e. that there is an effective method whose application would enable us to know, of every sentence $\varphi \in D$, whether φ is true or false. But this is also problematic. For one thing, we have seen that anti-realists may not want to *directly* rely on the Manifestability Requirement in their argument for the rejection of classical logic: the interesting and more general issue is whether anti-realism, however motivated, is incompatible with classical logic. Since Tennant himself thinks that Dummett's Manifestation Challenge compels us to accept the knowability of truth—as he puts it, it "constrain[s] truth to be epistemic" and "turn[s] truth into *knowable* truth" (Tennant, 1997, p. 179)—it seems therefore more appropriate to focus on the argument from the Principle of Knowability and bivalence to effective decidability. In symbols:

(CI_3) $\mathsf{KP}, \mathsf{BIV}_D \vdash e\mathsf{DEC}_D$.

But now, recall our argument to the effect that $e\mathsf{DEC}_D$ and

$$(\mathsf{DEC}_D)\ \forall\varphi(\varphi \in D \to (\Diamond\mathcal{K}\varphi \vee \Diamond\mathcal{K}\neg\varphi))$$

are equivalent (see *supra*, § 2.2). That is:

$$(\mathsf{EQ})\ e\mathsf{DEC}_D \leftrightarrow \mathsf{DEC}_D.$$

Our reasoning was as follows:

> Left-to-right: if there is a method for determining φ's truth-value, for every $\varphi \in D$, then either φ or its negation must be knowable. For if one applied the method in question, one would know either φ or its negation, which is to say that either φ or its negation are knowable. Right-to-left: if either φ or its negation are knowable, then there is a method whose application would enable us to know either φ or its negation. Otherwise, it is very difficult to see how φ or its negation could be knowable in the first place.

If this is correct, Tennant's central inference may be represented as follows:

$(\mathsf{CI}_4)\ \ \mathsf{KP}, \mathsf{BIV}_D \vdash \mathsf{DEC}_D.$

If we grant the equivalence between BIV and LEM (via the Equivalence Thesis), and if we ignore the relativisation to a discourse D, the above inference can be easily proved:

> *Proof*: Assume LEM and KP. Now show, by disjunction introduction and arrow introduction that $\Diamond\mathcal{K}P \to (\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P)$. It follows, by transitivity of '$\to$' and KP, that $P \to (\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P)$. By similar reasoning, show that $\neg P \to (\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P)$. But LEM licenses us to infer $\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P$ from $P \to (\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P)$ and $\neg P \to (\Diamond\mathcal{K}P \vee \Diamond\mathcal{K}\neg P)$ by disjunction elimination and arrow elimination. By $\forall$ introduction, DEC follows. ■

Classical logic and semantic anti-realism jointly entail the solvability in principle of every problem. If we are not prepared to assert that every problem is solvable in principle, it follows that we are not prepared to prepared to assert one between KP and LEM. *Anti-realists*, of course, will stick to the former and give up the latter. This is, in essence, the Basic Revisionary Argument.

Appendix C

Empirical negation

In this Appendix, we briefly consider the idea that intuitionists may solve the Church-Fitch problem by enriching their language with an *empirical* negation '$\sim$', such that (i) it can be applied to contingent, empirical statements, and (ii) no contradiction follows from $\mathcal{K}(P \wedge \sim\mathcal{K}P)$ and $\mathcal{K}(Q \wedge \sim\exists x Ix)$ (where Q is, of course, some feasibly unknowable statement). Eminent intuitionists, such as Dummett, are aware that this is a tall order:

> Negation [...] is highly problematic. In mathematics, given the [constructive] meaning of 'If, ...then," it is trivial to explain "Not P" as meaning 'If P, then $0 = 1$'; by contrast, a satisfactory explanation of "not", as applied to empirical statements for which bivalence is not, in general, taken as holding, is very difficult to arrive at. (Dummett, 1993a, p. 473)

In their *Empirical Negation in Intuitionistic Logic*, De Vidi and Solomon bravely take up task of defining an empirical intuitionistic negation.[1] Their suggestion is to introduce an empirical negation '$\sim$' respecting conditions (i) and (ii) above to Saul Kripke's semantics for intuitionistic logic.[2]

C.1 Kripke's semantics for intuitionistic logic

I begin by introducing Saul Kripke's semantics for intuitionistic logic. For the sake of simplicity, let us confine our attention to the propositional case. Our language $\mathcal{L}$ contains the standard logical connectives $\wedge$, $\vee$, $\rightarrow$, $\neg$ and a falsity symbol $\perp$. We

[1] See De Vidi and Solomon (2006). Thanks are due to David De Vidi for sending me a hard copy of the paper.

[2] See Kripke (1965).

let **Atoms** denote the set of the atomic formulas of $\mathcal{L}$. $\mathcal{L}$ is defined by the usual recursive clauses for $\wedge$, $\vee$, $\rightarrow$ and $\neg$. The variables P, Q, R range over sentences of $\mathcal{L}$; the variables B, C range over atoms. A *frame* is a triple $\langle W, \leq, \mathbf{A}\rangle$, where W is a set, $\leq$ is partial order on W, i.e. a reflexive and transitive relation on W and $\mathbf{A} \subseteq W$. Intuitively, W is a set of *states of information* and $\mathbf{A}$ is the set of states of information which are, have been or will be actualized. $\leq$ may be interpreted as the 'possible development of knowledge relation'.[3] From this perspective, the reflexivity and transitivity of $\leq$ are quite natural constraints. If x is a possible development of what is known at w and y is a possible development of what is known at x, then y is a possible development of what is known at w (where w, x, y are states of information). Similarly, a state of information w is (trivally) a possible development of what is known at w. An interpretation I on $\langle W, \leq, \mathbf{A}\rangle$ is a mapping from W to the powerset of **Atoms** that are subject to the *persistence condition* (or *heredity condition*)[4]: if $w \leq x$, then $I(w) \subseteq I(x)$.

Let $w \Vdash P$ abbreviate "P is *true* (or forced) in w". Then, truth for a given interpretation I on a given frame F is thus defined:

(Atom) If B is atomic, $w \Vdash B$ if and only if $B \in I(w)$.

($\vee$) $w \Vdash P \vee Q$ if and only if $w \Vdash P$ or $w \Vdash Q$.

($\wedge$) $w \Vdash P \wedge Q$ if and only if $w \Vdash P$ and $w \Vdash Q$.

($\forall$) $w \Vdash P \rightarrow Q$ if and only if $\forall x \geq w$, if $x \Vdash P$ then $x \Vdash Q$.

($\exists$) $w \Vdash \neg P$ if and only if $\forall x \geq w, x \nVdash P$.

De Vidi and Solomon usefully define three different notions of validity for the semantics we have just sketched:[5]

(V_1) P is *valid under I* if P is forced at every $w \in W$ under I.

(V_2) P is *valid in F* if P is valid under every I on F.

(V_3) P is *valid* if P is valid in every frame F.

We now turn to De Vidi and Solomon's proposed account of empirical negation.

[3]See Beall and van Frassen (2003), p. 96.
[4]See Beall and van Frassen (2003), p. 97.
[5]See De Vidi and Solomon (2006), p. 157.

C.2 Empirical negation

De Vidi and Solomon suggest that we add to $\mathcal{L}$ the unary connective $\sim$, whose intuitive meaning is 'It is not actually the case that'. The recursion clause for $\sim$ is as follows:

($\sim$) $w \Vdash \sim P$ if and only if, for all x s. th. $x \geq w$, and $x \in \mathbf{A}$, $x \nVdash P$.

The intuitive idea is that $\sim P$ holds at a given state w just in case no actual state of information forces P. De Vidi and Solomon (pp. 158-59) note a few interesting facts about '$\sim$':

1. $P \vee \sim P$ is not valid. For if $x \nVdash P$ but $\exists y > x$ s. th. $y \in \mathbf{A}$ and $y \Vdash P$, then $x \nVdash \sim P$ either.
2. $\neg P \to \sim P$ is valid.
3. $\sim P \to \neg P$ is *not* valid. For it is possible that $x \Vdash \sim P$ while, for some $y \geq x$, $y \notin \mathbf{A}$ and $y \Vdash P$. Hence, $x \nVdash \neg P$.
4. If $x \notin \mathbf{A}$, then it is possible that, for some P, $x \Vdash P \wedge \sim P$. For possibly $x \Vdash P$, but there is no $y \geq x$ such that $y \in \mathbf{A}$ (hence, $x \Vdash \neg P$). Thus $\neg(P \wedge \sim P)$ is not valid. However $\sim(P \wedge \sim P)$ *is* valid, because $P \wedge \sim P$ cannot be true at any $x \in \mathbf{A}$.

The last point is crucial. For while $P \wedge \neg\mathcal{K}P$ is absurd if $P \to \neg\neg\mathcal{K}P$ holds, $P \wedge \sim\mathcal{K}P$ is indeed consistent on this assumption. As De Vidi and Solomon put it, "$P \wedge \sim\mathcal{K}P$ can be forced in a state of information x if for all $y \geq x$ such that $y \Vdash \mathcal{K}P$, $y \notin \mathbf{A}$" (De Vidi and Solomon, 2006, p. 159). Since the same result applies to $\mathcal{K}P \wedge \sim\mathcal{K}P$, it would seem that intuitionists who adopt De Vidi and Solomon's empirical negation are in a position to block the Church-Fitch Paradox. In particular, $\mathcal{K}P \wedge \sim\mathcal{K}P$ can be true at any state of information x such that $x \notin \mathbf{A}$. Ditto for $\exists x Ix \wedge \sim\exists x Ix$ This, of course, raises the problem of whether it can be known, in a non-actual state of information, that something is not known at an *actual* state of information.[6] There are, however, even more problems.

C.3 Never say never

In his *Never Say Never*, Williamson briefly considers the possibility that intuitionists express 'Never' by means of an empirical negation.[7] His conclusion is that no

[6] See Williamson (1994, p. 138).
[7] See Williamson (1994).

negation weaker than $\neg$ can be consistently defined in intuitionistic logic. He writes:

> There is a reason to believe that any alternative negation must be at least as strong as $\neg$. For 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. A warrant for $A \wedge \sim A$ should be impossible. That is, we should have $\neg(A \wedge \sim A)$. By the intuitionistically valid schema $\neg(A \wedge B) \rightarrow (B \rightarrow \neg A)$, this yields $\sim A \rightarrow \neg A$ (Williamson, 1994, p. 139).

If $\neg(A \wedge \sim A)$ holds, then $\sim A \rightarrow \neg A$ holds too. Given, however, that the converse of this latter schema is unexceptionable, the argument has it that $\sim$ inevitably collapses on $\neg$.

Williamson's argument rests on the assumption that $\neg(A \wedge \sim A)$ must hold, if $\sim$ is to be considered a negation at all. However, we have seen that this latter schema does not hold within De Vidi and Solomon's semantics. Should we conclude that $\sim$ does not qualify as a negation? We may not be not forced to take a stand on the matter. In fact, Williamson offers an argument to the effect that $\neg(A \wedge \sim A)$ must hold, *tout court*. His idea is that $A \wedge \sim A$ *cannot* be warranted. Hence, there exists no proof for it. By the intuitionistic meaning of $\neg$, though, $\neg(A \wedge \sim A)$ follows.

C.4 De Vidi and Solomon's reply

De Vidi and Solomon note two things about this argument. They write:

> First, when one moves from the constructivist to the Kripke semantic reading of intuitionistic $\neg$, what it means for a $\neg$ sentence changes. In particular, its meaning is no longer directly tied to the possibility of a warrant in the way necessary for Williamson's argument to go through. Which brings us to the second point. If we make a (simplifying?) equation of being forced at x and being warranted at x, then there is a sense in which $A \wedge \sim A$ *is* impossible to warrant according to the present proposal—it could never *actually* be warranted. There are possible states of information in which both P and $\sim P$ are forced, but these necessarily are non-actual ones. The non-actuality of such states of information dissolves the appearence of Paradox, because $\sim P$

> says, in effect, that P is not forced at any *actual* state of information. So, perhaps, in the relevant sense of "impossible to warrant" the conjunction in question is impossible to warrant, but the[n] the intuitionistic negation of that claim doesn't follow. So it doesn't follow that $\sim$ must be at least as strong as $\neg$. Which is a good thing, because our proposal for $\sim$ is *weaker* than $\neg$. (De Vidi and Solomon, 2006, p. 167)

It is worth unpacking De Vidi and Solomon's reasoning. First, they notice that 'being forced' and 'being warranted' are not obviously one and the same notion.[8] Second, they argue that, even if we equate these two notions, Williamson's argument may be blocked nevertheless. For although there is a sense in which $A \wedge \sim A$ cannot be warranted—it cannot be warranted in the actual state of information—there is also a sense in which it *can* be warranted. This can be seen as soon as we reflect on the meaning of $\sim P$, which can be intuitively interpreted as 'Actually, it is not the case that P'. With this in mind, it is easy to see that, in a non-actual state of information, one can have a warrant for $A \wedge \sim A$. So there is a sense in which this conjunction is impossible to warrant, but this sense is not sufficient to establish $\neg(A \wedge \sim A)$. This is a welcome result, De Vidi and Solomon observe, given that, according to their proposal, $\sim$ is weaker than $\neg$.

C.5 Two problems

It is unclear whether De Vidi and Solomon's proposal can ultimately be made to work. There are at least two problems.

To begin with, De Vidi and Solomon's proposal requires that it can be known, in a non-actual state of information, that something is not known at an *actual* state of information. That is, the proposal requires that *transworld knowledge*—knowledge, in w, that, in w', A—be possible. However, as we shall see in § C.6, it is rather doubtful whether this notion is coherent at all.[9]

Secondly, the proposal requires that statements of the form $A \wedge \sim A$ be true at some (non-actual) state of information. But what could establish the truth of such statements? Plainly, no argument for $A \wedge \sim A$, at any state of information, can count as an intuitionistically acceptable proof, or warrant, for statements of his form. As De Vidi and Solomon correctly realize, the only possible answer to the

[8] In particular, 'P is forced' does not seem to imply 'P is warranted' (since $A \wedge \sim A$ can be forced without being warranted), whereas 'P is warranted' seems to imply 'P is forced'. The notion of warrant is thus stronger than the notion of being forced. Thanks to Bob Hale here.

[9] See also Murzi (2008) for more discussion.

problem is to dissociate intuitionistic logic—or, at least, Kripke semantics—from the notion of proof, or warrant. One is left wondering, though, about what would be left of intuitionism as an anti-realist program, given this assumption. The idea that central notions such as truth or inference could be defined in terms of the more immediate notions of proof, warrant, or justification, is central for the intuitionistic enterprise. If the cost of 'solving' the Paradox of Knowability is to divorce intuitionism, and intuitionist semantics, from the epistemic notions that intuitionists of all brands take as primitive, then De Vidi and Solomon's proposal, albeit interesting on its own right, cannot deserve to be called a 'solution' of the problem we started with.

C.6 Possible knowledge of actual truths?

De Vidi and Solomon's suggested solution to the Paradox of Knowability requires that transworld knowledge—knowledge, in w, that, in w', A—be possible. In this section, we briefly consider Dorothy Edgington's attempt to make sense of the notion, and some of the problems it faces.

C.6.1 All actual truths are knowable

As the reader may recall, in order to argue from the Church-Fitch proof to the falsity of the Knowability Principle, one needs to assume, among other things, that WVER correctly captures its logical form. But this is a very contentious assumption According to Dorothy Edgington (1985), anti-realists can solve the Knowability Paradox by carefully distinguishing our concepts from their formalizations. Within quantified modal logics equipped with an actuality operator, she observes, the Knowability Principle allows for the following non-paradoxical reading:

$$\text{(WAVER) } \forall\varphi(@\varphi \to \Diamond\mathcal{K}@\varphi),$$

where '@' is a rigidifier on circumstances of evaluation.[10] For every actual truth φ, the alternative principle says, it can be known that it is actually the case that φ. As Edgington points out, WAVER does not entail the unwelcome SVER.[11] Substitution

[10]Its truth-conditions are standardly given as follows:

$$\text{(Actually) } v(@A, w) = 1 \text{ iff } v(A, w_c) = 1,$$

where w_c is the world of the utterance context.

[11]See Edgington (1985, pp. 562-3).

of $@(P \wedge \neg\mathcal{K}P)$ into WAVER only gives:

$$(1)\ \Diamond\mathcal{K}@(P \wedge \neg\mathcal{K}P).$$

There is no contradiction, however, in claiming that it is possible to know that it is actually the case that P but nobody actually knows that P. As Crispin Wright puts it:

> we have to ask after the range of the quantifiers 'no one' in $[P \wedge \neg\mathcal{K}P]$'s second conjunct. Whatever it is, it is consistent with each of the subjects who fall within it always lacking warrant both for P and for the perennial ignorance of the matter of each of the subjects in the former range. (Wright, 2003a, p. 68)

Via the introduction of a further necessity operator that reads 'fixedly', Edgington shows that WAVER can be generalized to both actual and possible truths.[12]

Contrary to Dummett's restriction, Edgington's formulation of the Knowability Principle is not obviously *ad hoc*: the principle is, as a matter of fact, ambiguous between at least two readings, and, at least *prima facie*, it is open to the anti-realist to argue that Edgington's reading was the one she had always intended. Furthermore, WAVER *can* serve as the main premise of the Basic Revisionary Argument, provided that the thesis of the Missing Guarantee is formulated as the claim that we do not presently know that every statement or its negation can be known to be actually true. However, there is a well-known problem with WAVER: it requires that anti-realists be able to make sense of the idea of *possible knowledge of actual truths*.

C.6.2 Transworld knowledge and the Trivialization Objection

WAVER requires that non-actual subjects have knowledge of the actual world. But how this could be?[13] Certainly we cannot have causal interactions with merely *possible* situations. But then how could we know anything about them? Edgington suggests that transworld knowledge is *counterfactual* knowledge and that non-actual subjects can refer to the actual world by description. In particular, although we cannot actually know that P but it is not known that P, we nevertheless *could have known* this conjunction:

[12] See Edgington (1985, pp. 567-77).

[13] Edgington's suggested treatment of the Church-Fitch proof is not the only one requiring some account of transworld knowledge—see e.g. Rabinowicz and Segerberg (1994), Kvanvig (1995), (2006) and Brogaard and Salerno (2008). The points in the main text carry over, *mutatis mutandis*, to each of these proposals.

> my idea, in outline, [is] there is a sense in which one can know that, as things actually are, P and it is not known that P, but from a counterfactual perspective—as it were, from a modal distance. The 'world of the knower' need not be the same as the 'world of the truth'. (Edgington, 2010)

Edgington's proposal hinges on two main assumptions: that anti-realists can have knowledge of counterfactual claims, and that they can understand talk of *possibilities*, or *possible situations*, where "possibilities differ from possible worlds in leaving many details unspecified" (Edgington, 1985, p. 584). As opposed to possible worlds, situations typically are (although they need not be) *incomplete*. For present purposes, they can be thought of as subsets of possible worlds—whatever possible worlds may be.[14] As Edgington puts it:

> There are *indefinitely many* possible worlds compatible with [...] one possibility—which vary [...] as to whether it is raining in China at the time, or at any other time, and so on *ad infinitum*. Knowledge of counterfactual situations is never of one specific possible world. [...] This suggests that possible worlds are far too idealised to figure in our ordinary modal talk. When I think of the possibility that I will finish the paper today, I am not thinking of one totally specific possible world. It is not the sort of thing I am capable of thinking of. It, itself, seems to violate the principle of knowability. [...] I am thinking of a possibility or a possible situation, which I can refine, or subdivide, into more specific possible situations if I wish, but which will never reach total specificity. (Edgington, 1985, p. 564)

If the anti-realist grants this much, Edgington contends, the Paradox of Knowability may be blocked. For consider the following counterfactual:

(2) If my parents had not met, I would not have been born.

In Edgington's view, (2)'s antecedent specifies a situation, call it s, that would have obtained if my parents had not met, where everything else in s is left as close as possible to the actual world. Then, the idea is that knowledge of (2) may be seen as knowledge that, in s, I would not have been born. More generally, let s_1 and s_2 be possible situations. Then, if ψ is a sufficient condition for s_2 to obtain, knowledge in s_1 of the counterfactual conditional $\psi \mathrel{\Box\!\!\rightarrow} \varphi$ constitutes knowledge, in s_1, that, in s_2, φ. We may formulate Edgington's proposal as follows:

[14] See Humberstone (1981).

(TK) S knows, in s_1, that, in s_2, φ, if

(i) there is a ψ such that ψ is a sufficient condition for s_2 to obtain and

(ii) S knows, in s_1, that $\psi \mathbin{\Box\!\!\rightarrow} \varphi$.

The Trivialization Objection

Edgington's proposal faces a number of objections—see e.g. Percival (1991) and Williamson (2000, Chapter 12). Here I will focus on what is perhaps the most pressing one: Timothy Williamson's so-called Trivialization Objection. Williamson's idea is that transworld knowledge makes possible knowledge all too easy. He writes:[15]

> suppose that, in the world x, the world w would have obtained if P had been true, and that Q is true in w. Then, in x, w would have obtained if the conjunction $P \wedge Q$ had been true; in the terms of a possible worlds semantics for the counterfactual conditional, if Q is true in w and w is the closest world to x in which P is true then w is the closest world to x in which $P \wedge Q$ is true. The proposal therefore implies that knowledge in x of the counterfactual $(P \wedge Q) \mathbin{\Box\!\!\rightarrow} P$, constitutes knowledge in x that P is true in w. But since P is a truth-functional consequence of $P \wedge Q$, the counterfactual is a trivial necessary truth. (Williamson, 2000, p. 295)

The argument does not aim at showing that transworld knowledge is *always* trivial. This would involve that counterfactual knowledge itself is always trivial, which is surely false. Rather, the argument has it that, on Edgington's account of transworld knowledge, if conditionals of the form $A \mathbin{\Box\!\!\rightarrow} B$ can constitute knowledge of $@B$, then also conditionals of the form $(A \wedge B) \mathbin{\Box\!\!\rightarrow} B$ can. But, of course, knowledge of $(A \wedge B) \mathbin{\Box\!\!\rightarrow} B$ is trivial, given that B is an immediate logical consequence of $A \wedge B$. The argument is usually regarded as a fatal objection to WAVER and, more generally, to any conception of knowability requiring transworld knowledge.[16] Edgington herself acknowledges the force of the objection:

[15]See Williamson (1987). See also Williamson (2000, pp. 290-6). See also Kvanvig (2006, pp. 58-62) for a related formulation of the problem. Notice that, although Williamson is here referring to worlds, nothing in his argument hinges on this assumption: the argument would equally go through if we substitute worlds with situations in Williamson's text.

[16]See Williamson (1987), Cozzo (1994), Rabinowicz and Segerberg (1994), Rückert (2004), Kvanvig (2006), Jenkins (2007) and Brogaard and Salerno (2008).

> Knowing merely that the train leaves at the time the train leaves, is not to know when the train leaves. Knowing merely that the sum of this long list of numbers is the sum of this list of numbers, is not to know which number this is. Similarly, knowing merely that, in the possible situation in which P is an unknown truth, P is an unknown truth, is not to know which possible situation that is. For any kind of entity, merely having a true definite description does not suffice for knowing which entity you are talking about. (Edgington, 2010)

Non-trivial counterfactuals

How could anti-realists react? Since Edgington's proposal is trivialized by logically true conditionals of the form $(A \wedge B) \mathrel{\Box\!\!\rightarrow} B$, a quite natural suggestion would be to restrict her counterfactual account of transworld knowledge to non-logically true counterfactuals. Wlodeck Rabinowicz and Krister Segerberg first entertained this possibility in an early paper on the subject. They write:

> It might be objected that the logically true counterfactual $(P \wedge Q) \mathrel{\Box\!\!\rightarrow} P$, which has been used in th[e] trivialization proof, is itself 'too trivial' to yield any knowledge of the counterfactual situation. Perhaps then we should qualify the suggested sufficient condition by a demand that the relevant counterfactual should not be logically true. (Rabinowicz and Segerberg, 1994, p. 125, fn. 3)

Accordingly, TK could be modified as follows:

(TK*) S knows, in s_1, that, in s_2, φ if

(i) there is a ψ such that ψ is a sufficient condition for s_2 to obtain,

(ii) S knows, in s_1, that $\psi \mathrel{\Box\!\!\rightarrow} \varphi$,

(iii) $\psi \mathrel{\Box\!\!\rightarrow} \varphi$ is not a logical truth.

However, as Williamson (2000, p. 294) points out,[17] this would be a bad move. For let R be some very far-fetched proposition, so that R-situations are much farther from s_2 than $P \wedge Q$-situations are. Then, Williamson notices, $(P \wedge Q) \vee R$ is a trivializing sufficient condition. As he puts it:

> let R state something utterly bizarre, logically quite independent of both P and Q, such that it is obvious in x that there are worlds much

[17] And as Rabinowicz and Segerberg acknowledge in their paper.

> closer to x in which $P \wedge Q$ is true than any in which R is true. Then, as before, w is the closest world to x in which the disjunction $(P \wedge Q) \vee R$ is true. Thus the counterfactual $((P \wedge Q) \vee R) \,\Box\!\!\rightarrow P$ is true but not logically true in x, so knowledge of it constitutes knowledge in x that R is true in w even by the modified proposal. But knowledge of the counterfactual $((P \wedge Q) \vee R) \,\Box\!\!\rightarrow P$ is still trivial by contrast with knowledge of $@R$, because its basis is just that R is a far more outlandish supposition than $P \wedge Q$. (Williamson, 2000, p. 295)

Would a restriction to broadly non-trivial counterfactuals fare any better? It does not look like it. As Williamson observes,[18] a small adaptation of the trivialization argument could meet that point by including the negation of a non-trivial mathematical theorem as an extra disjunct of the antecedent of the previously trivial counterfactual, so that one had to prove the theorem in order to derive the counterfactual. Non-triviality would thus be irrelevant for the anti-realist's purposes. It seems to follow that restrictions to either non-logically true or non-trivially true counterfactuals offer no shelter from the trivialization threat. Rabinowicz and Segerberg are themselves agnostic about the possibility of finding an adequate refinement of TK:

> Is it possible to qualify the condition above in some other way, so as to avoid all the trivialization threats? We are not sure. (Rabinowicz and Segerberg, 1994, p. 125, fn. 3)

C.6.3 Edgington's reply

Twenty-three years after the publication of her 1985 paper, Edgington has recently responded to Williamson's objection. Knowledge of possible situations, she now argues, is best thought of in analogy to knowledge of future situations. It is unclear whether Edgington is clarifying her 1985 proposal, or whether she is rather presenting a new one, pressed by Williamson's objection. The aim of the present section is to introduce Edgington's response, and consider some potential objections.

Future and possible situations

Edgington's starting point is that we *do*, as a matter of fact, refer to, and have knowledge of, merely possible situations. This seems true enough. We refer to the

[18] Williamson, p.c.

possible situation that had obtained if the war to Iraq had not been declared, and we refer to the possible situation that had obtained I had waken up earlier this morning. But how to make philosophical sense of this capacity? Edgington now writes that

> the best analogy for referring to particular counterfactual situations is [...] referring to future situations. One cannot perceive the future, or receive testimony from it. We are causally connected to the future, but not in the direction of receiving information from it. Yet I can think and speak of, and know or have reasonable beliefs about, say, the water in the kettle boiling soon, having plugged it in. The same resources allow me to judge that it will boil if I plug the kettle in, or that it would have boiled if I had plugged the kettle in. Our grasp of possible states of affairs is just like our grasp of actual future states of affairs. (Edgington, 2010)

Edgington's suggestion is to assimilate knowledge of possible situations to knowledge of *future* situations. Just like I can know that, given that I have turned the kettle on, the water will boil soon, I can also know that if I *were to* turn the kettle on, the water *would* boil soon. We may grant this much. But how is this suppose to help answering Williamson's argument? Edgington's response is rather brief. She observes that although we are not in causal contact with many possible objects, we may be in actual contact with their possible components. For instance, we can refer to the possible vase that could be made out of a particular piece of clay, we could refer to "the merely possible person that would have resulted from the union of this particular sperm and this particular egg" (*Ibid.*), and so on. Edgington argues that a similar reasoning applies to possible situations. She writes:

> similarly, I suggest, to have enough handle on which possibility you are talking about, one refers to it as the one that would have developed, had there been a course of history which diverged at a certain point from the actual history. One needs to be able to specify the point of departure and the way things would have developed, in a reasonable amount of detail; that is, one has to be able to reconstruct, in outline, a causal route, beginning with history shared with the actual world, of how things would have deviated to produce such-and-such result. And from a counterfactual perspective, the possibility one refers to in this manner may be the way things actually are. (Edgington, 2010)

Edgington's idea seems to be this: in order to refer, in s_1, to a possible situation $s_2 \neq s_1$, we need to specify s_2's "point of departure" from s_1. For instance, let s_1 be a situation in which

> I am fortunate enough to chance upon a discovery which no one else is in a position to make. I am an astronomer, and am the only person to observe a supernova before it disappears for ever [...]. (Edgington, 1985, p. 563)

Let s_2 be a situation in which nobody, including me, was star-gazing last night: the supernova appears in the sky, but nobody will ever know that. Let Q and P be, respectively, "I was star-gazing last night" and "A supernova appears in the sky". Then, P holds at both s_1 and s_2 (the supernova appears in both situations), Q and $\mathcal{K}P$ hold at s_1 (in s_1 someone knows, namely me, that the supernova appeared), and their negations hold at s_2 (in s_2 nobody knows that the supernova appeared in the sky). Our task now is to describe "in a reasonable amount of detail" s_2's point of departure from s_1: the moment at which s_2's history departs from s_1's. Let D be one such description. D will include $\neg Q$, and possibly other sentences, although it need not be complete. Then, the idea seems to be that the reconstruction of a causal route C leading from D to how things are now in s_2 will constitute a description of s_2—a description available in every possibility s such that $s \cap s_2 \neq \emptyset$. Now let s_2 be the actual situation. Then, Edginston suggests, subjects in s_1 can refer to the actual situation in the way we have just described, i.e. by identifying s_2's point of departure and by describing, in a reasonable amount of detail, a causal route from s_2's point of departure to how things are now in s_2. How, if at all, can Edginton's original proposal be modified in light of the foregoing considerations? This seems what Edgington appears to have in mind:

(TK**) S knows, in s_1, that, in s_2, φ, if

(i) there is a ψ such that ψ is a sufficient condition for s_2 to obtain,
(ii) ψ individuates s_2's point of departure from s_1,
(iii) S is able to reconstruct, in outline, a causal route from ψ to how things are now in s_2,
(iv) S knows, in s_1, that $\psi \mathrel{\Box\!\!\rightarrow} \varphi$.

Points of departure

If this is Edgington's new proposal, then it does not seem to work any better than the original 1985 one. Here is a quick argument. If D can be correctly identified as s_2's point of departure from s_1, so can the conjunction $D \wedge (P \wedge \neg\mathcal{K}P)$—recall, both P and $\neg\mathcal{K}P$ hold at s_2. It follows that both knowledge of $D \mathrel{\Box\!\!\rightarrow} (P \wedge \neg\mathcal{K}P)$ *and* knowledge of $D \wedge (P \wedge \neg\mathcal{K}P) \mathrel{\Box\!\!\rightarrow} (P \wedge \neg\mathcal{K}P)$ constitute knowledge, in s_1, that $P \wedge \neg\mathcal{K}P$ holds at s_2. If this is correct, either TK** is not what Edgington really has in mind, or her proposal is in need of revision. In order to solve the problem, Edgington would have to find some property $\mathcal{P}$ which D has, but which $D \wedge (P \wedge \neg\mathcal{K}P)$ lacks, such that in virtue of having $\mathcal{P}$, D, but not $D \wedge (P \wedge \neg\mathcal{K}P)$, correctly describes s_2's point of departure. It is unclear, however, whether, and, in case, how, this can be done.

Manifestability and understanding

But let us concede to Edgington, for the argument's sake, that transworld knowledge *may* be accounted for, some way or other. Then, one might wonder how much has been achieved. The real issue, I take it, has to do with the link between Edgington's WAVER and the anti-realist's original Manifestation Requirement. Williamson first raised the issue:

> A verificationist principle (WVER) was originally motivated by argument about the nature of meaning. In response to Fitch's argument, the principle was modified. But it was not checked that the meaning-theoretic arguments for WVER could plausibly be reconstructed as an argument for WAVER [...]. (Williamson, 2000, p. 299-300)

The upshot of Dummett's meaning-theoretic arguments was that truth cannot be verification-transcendent. For, if it were, there would be sentences whose understanding could not be manifested by being disposed to recognize a proof of them when presented with one—for undecidable sentences, there may not be any such proof. Now, WAVER *prima facie* meets this requirement: it, too, rules out the existence of verification-transcendent truths. Or does it?

Recall, the Manifestation Requirement tells us that, for every sentence A we happen to understand, there must be some possible situation s such that, in s, we are able to recognize a proof of A, if presented with one. It should be clear that, if anything, this principle only supports WVER. In order to support WAVER, it would have to tell us that, for every sentence A we happen to understand, there must be some possible situation s_2 such that, in $_2$, we are able to recognize a proof that A *is true at* s_1, if presented with one, where possibly $s_1 \neq s_2$. But here is the problem. Originally, A did not say anything *about* any situation whatsoever. Yet, we are now required to recognize proofs to the effect that A *holds at s*, for some situation s. This is puzzling. In order to manifest understanding of A, we are asked to display recognitional capacity with respect to *a different sentence*, A^*, of the form $B \boxright A$. In order to manifest understanding of a sentence A, we are asked to be able to recognize proofs of *logically more complex* sentence A^*. This, however, introduces a circularity in the anti-realist's account of understanding.[19] In order to understand A, we need to understand $B \boxright A$, which in turn demands an understanding of A.[20]

What to learn from counterexamples

Even conceding to Edgington that Williamson's trivialization objection can be adequately answered, Edgington's proposed solution of the Paradox of Knowability faces some very serious problems. For one thing, it is difficult to see how WAVER can be supported by the standard anti-realist meaning-theoretic arguments. For another, we found that WAVER renders the anti-realist's account of understanding viciously circular: in order to understand A, we are required to understand a logically more complex sentence A^*. Williamson concludes that, although

> the defender of WAVER [might have] some way of rending [this explanatory circle] harmless, [...] we should not rush to assume that the defence of those principles can be reconciled with the meaning-theoretic ideas which were supposed to motivate the original weak verificationism. Sometimes we should learn from counterexamples that a philosophical idea was wrong in spirit, not just in letter. (Williamson, 2000, p. 300)

This is a strong claim. After all, we cannot exclude that there may be compelling reasons for giving up the assumption that an adequate theory of meaning should be compositional, and that these reasons are available to anti-realists. Still, I agree with Williamson that there seem to be no obvious way to solve the problems besetting WAVER, and that, pending further argument, it is fair to conclude that anti-realists may want to explore alternative solutions to the Paradox of Knowability.

[19] Williamson (2000, p. 300) makes a similar point.

[20] The point carries over to the weaker formulation of the Manifestation Requirement we gave in § 3.2.3. The weaker requirement demands, among other things, that we are disposed to recognize proofs of atomic sentences, if presented with them. But the argument I have just given makes no assumption concerning the logical complexity of A.

Appendix D

Higher-order logics

In this Appendix, I offer reasons for thinking that inferentialists can accommodate higher-order logics, even though this requires that they give up their unqualified commitment to inferentialist orthodoxy, the view that admissible formalizations of logic should satisfy separability and conservativeness. Section D.1 argues that the non-conservativeness result presented in § 6.4.2 sheds some light on the proof-theoretic interpretation of the higher-order quantifiers. More specifically, I will suggest that, because of the non-conservativeness of higher-order logics, inferentialists can address one standard objection to proof-theoretic accounts of higher-order logics, that, from a purely syntactic standpoint, the interpretation of the higher-order quantifiers is left radically underdetermined.[1] Sections D.2-3 sketch an inferentialist account of the meaning, and of our understanding, of the higher-order quantifiers.

D.1 Higher-order logics: a syntactic approach

It is well known that there are at least two different semantics for higher-order logics: the standard semantics, and the so-called Henkin semantics.[2] Narrowing down our focus on second-order logic, and simplifying a little, the crucial difference between the two is that, in the standard semantics, the second-order variables are assigned semantic values on the full power set of the domain, whereas, in the Henkin semantics, they are allowed to range over a proper subset of the powerset of the domain. This apparently minor difference has very major effects.

[1]Thanks to Ian Rumfitt for pressing me on this point, and to Timothy Williamson for hinting at the solution to the problem to be developed below. Marcus Rossberg also develops an argument similar to the one to be presented below in an unpublished manuscript. See also Restall (2008).

[2]For details, see Shapiro (1991, Chapter 4).

On the standard semantics, second-order PA has only one model, up to isomorphism: the set of natural numbers. On the Henkin semantics, it does not: second-order PA with Henkin semantics has *non-standard* models, i.e. models that are not isomorphic to the intended model of PA. It follows from the categoricity of second-order PA that all the standard meta-results for first-order logics, such as completeness, compactness, and the Löwenheim-Skolem theorems, fail for second-order logic with the standard semantics. By contrast, they all hold on the Henkin semantics. It follows that, on the Henkin interpretation, the standard formalizations of second-order logic are not really second-order: they are multi-sorted first-order logics—first-order theories with two different sorts of first-order quantifiers and comprehension axioms (see Shapiro, 1991, p. 74).[3] To be sure, from a model-theoretic perspective, it is clear what the interpretation of the higher-order quantifiers is: the interpretation is unambiguously settled by our choice of the semantics. But how can we decide how to interpret higher-order quantifiers from an *inferentialist* perspective? The rules for the higher-order quantifiers, we are invited to conclude, radically underdetermine their semantic interpretation.

This conclusion, however, would be a mistake. It is well-known that G_2, the Gödel sentence for second-order PA, is true on the standard semantics, but it is nevertheless *false* in some Henkin models—just as G, the Gödel sentence of first-order PA, is false in some first-order models of PA. It is also well-known that, because of categoricity, these Henking models are non-standard ones, i.e. they are not isomorphic to the natural numbers. But the same is true of $\forall f \forall x(\mathsf{PA_2}^* \rightarrow G_2^*)$: this sentence, too, is false on some Henkin models. We now show that these non-standard Henkin models cannot be extended to third-order ones.

Theorem 9. *Non-standard second-order Henkin models cannot be extended to third-order Henkin models.*

> *Proof*: On the standard semantics, G_2 and $\forall f \forall x(\mathsf{PA_2}^* \rightarrow G_2^*)$ are both true on the standard interpretation. Hence, by categoricity, these sentences are both semantic consequences of second-order PA. On the other hand, we know that they are not provable in the standard axiomatizations of second-order logic. But because these axiomatizations are complete with respect to the Henkin semantics, it follows that G_2 and $\forall f \forall x(\mathsf{PA_2}^* \rightarrow G_2^*)$ must be *false* in some Henkin-models—models that, by categoricity, we know to be non-standard. But then, since both G_2 and $\forall f \forall x(\mathsf{PA_2}^* \rightarrow G_2^*)$ are provable in third-order logic, and because

[3] As the reader may recall, $\forall^2$-E conveys the Comprehension Principle; see *supra*, § 6.4.2, fn. 16.

third-order logic is *sound*, these non-standard second-order models cannot be extended to third-order models. ■

The result generalizes: the higher up we move in the hierarchy of higher-order logics, the more non-standard models are eliminated at the lower orders.

Theorem 10. *For all $n > 2 \in \omega$, non-standard n^{th}-order Henkin models cannot be extended to $n + 1^{th}$-order Henkin models.*

Proof: This follows at once from Theorem 5. ■

Crucially, the result relies on the assumption that rules are *open-ended*: it is because of this assumption that the rules for the second-order quantifiers can be instantiated to formulae containing a truth-predicate defined in third-order logic. But now, recall our original problem, that from a purely syntactic perspective, the rules for the higher-order quantifiers radically underdetermine their interpretation. The foregoing considerations suggest that it is not so. Provided we are willing to ascend high-enough in the hierarchy of higher-order logics, the proof-theoretic relation of logical consequence gets closer and closer to the relation of consequence induced by the standard semantics, although, because of the incompleteness of higher-order logics, it never reaches it.[4]

D.2 Inferentialism and higher-order logics

If the foregoing considerations are correct, higher-order logics are in some sense higher-order, even from within an inferentialist perspective. For the orthodox inferentialist, however, higher-order logics come at a high price: the failure of conservativeness and separability. Inferentialists are confronted with a seemingly uncomfortable dilemma: they must reject either higher-order logics, or inferentialist orthodoxy. Or are they?

[4] I should add that this does not yet show that higher-order logics are, under a proof-theoretic interpretation, effectively higher-order—though, of course, much here depends of what we are to mean by 'higher-order'. Consider the logic we get by adding a countable infinity of quantifiers of ever increasing order to first-order logic—call it L. The semantics for L is the set of Henkin models that satisfy all of the instances of comprehension. Supposing, for simplicity, that all of the predicates (at all levels) are monadic, a Henkin model for L consists of (1) a domain of discourse $\mathcal{D}$, (2) interpretations for all of the non-logical terminology, and (3) for each n a subset of the n-fold powerset of $\mathcal{D}$. One should be able to show that L is complete—at least, this is what Stewart Shapiro suggested in correspondence. But, if L is complete, then it cannot be categorical, and, if so, it is not really higher-order, on some understanding of the term. I hope to be able to investigate the issue further in my future research.

D.2.1 Rejecting higher-order logics?

Inferentialists may be tempted to embrace the first horn. They may argue that higher-order logic is not logic, and treat sentences like $\forall f \forall x(\mathsf{PA}_2{}^* \to G_2^*)$ as they treat *arithmetical* Gödel sentences. Thus, Tennant argues that the Gödel sentence of PA is *synthetic*, although it can be known *a priori*. He writes:

> grasp of the meaning of the [Gödel] sentence itself is not sufficient for one to be warranted in asserting it; that is, the sentence is not epistemically analytic, even though its truth has been established *a priori*. (Tennant, 1997, p. 294)

Similarly, Tennant could argue that $\forall f \forall x(\mathsf{PA}_2{}^* \to G_2^*)$ is synthetic, since knowledge of the meaning of its component expressions is not sufficient for one to be warranted in asserting it. But this by no means constitutes a problem, he may maintain, because neither $\forall f \forall x(\mathsf{PA}_2{}^* \to G_2^*)$ nor the third-order resources needed to prove it are logical.

I do not find this line of argument very persuasive, however. To begin with, we have found no compelling argument for requiring that admissible introduction rules obey the complexity condition. Yet, this failure is the only proof-theoretic anomaly exhibited by the rules for the higher-order quantifiers. For any order $n \in \omega$, the rules for the n^{th}-order quantifiers are harmonious, and, for all we know, they satisfy the Fundamental Assumption. The proof of $\forall f \forall x(\mathsf{PA}_2{}^* \to G_2^*)$, for instance, requires third-order resources, but nonetheless ends with an application of $\forall^2$-I, as the Fundamental Assumption requires. Orthodox inferentialists might insist that the breach of conservativeness exhibited by the rules for the higher-order quantifiers is a sufficient reason not to regard them as logical. But this is problematic too. As we saw, the inferentialist's main reason for requiring conservativeness was that the system to which our new vocabulary is to be added is complete with respect the the intuitively correct uses of its logical vocabulary. However, this reason does not seem available when we move to higher-order logics: the non-conservativeness result presented in § 6.4.2 shows that the standard formalizations of higher-order logics are *incomplete* with respect to the intuitively acceptable uses of the higher-order vocabulary. Inferentialists should rather accept higher-order logics, and provide an account of the meaning of the higher-order expressions.

D.2.2 Higher-order logics and indefinite extensibility

Dummett suggests that mathematical or higher-order expressions have *indefinitely extensible* meanings.[5] He writes:

> The use of a mathematical expression could be characterized by means of a single formal system only if the sense of that expression were perfectly definite; when [...] the expression has an inherently vague meaning, it will be essential to the characterization of its use to formulate the general principle according to which any precise characterization can always be extended. (Dummett, 1963a, p. 198)

Dummett takes this view to require that there be a single general principle for generating a precise characterization of the correct uses of an higher-order expression. Presumably, he holds this view for epistemological reasons. The thought seems to be that, if the meaning of an expression $\$$ cannot be fully captured by a single rule, it may still be captured by a single principle for producing indefinitely many rules—a principle whose grasp is necessary and sufficient for understanding $\$$.

On the face of it, though, Dummett's view appears to be a non-starter—at least when applied to higher-order quantification. To begin with, if our understanding of a higher-order quantifier $\forall^n$ is constituted by our grasp of the rules for $\forall^n$ *together with our grasp of the rules of indefinitely many quantifiers of strictly higher order than n*, then there is no single "general principle". Rather, there are indefinitely many ones. Secondly, to concede that the meanings of the higher-order quantifiers are determined by indefinitely many rules is to give up the inferentialist thought that one should be able to understand $\forall^n$ without thereby understanding $\forall^{n+m}$. How to solve the problem?

D.2.3 The holist's response

Logical holists may be relatively unimpressed by the foregoing non-conservativeness result, just as they may not be impressed by the failure of separability of the standard formalizations of classical logic. Marcus Rossberg writes:

[5]For reasons made clear in Shapiro and Wright (2006), it would be better to reserve this expression to concepts that are properly indefinitely extensible, such as *ordinal* and *set*. Shapiro and Wright persuasively argue that a concept P is indefinitely extensible if and only if there is a one-to-one function from the ordinals into P (Shapiro and Wright, 2006, p. 258 and ff.). It is doubtful, however, whether higher-order logics extend beyond the finite ordinals. For the sake of simplicity, I will stick to Dummett's terminology. Shapiro and Wright's notion of indefinite extensibility may then be termed *indefinite indefinite extensibility*.

> the inference rules for the logical constants of second-order logic do not determine all the logical truths and consequences of second-order logic; the inference rules of other logical constants [...] are needed in order to establish that some second-order sentences express logical truths, or that some second-order sentences are logical consequences of other second-order sentences. (Rossberg, 2006, p. 221-2)

The idea is to assume that our understanding of, say, $\forall_n$ depends on our prior grasp of *all* the rules for the higher-order quantifiers, just like, Rossberg argues, "*all* the sentential operators of (classical) propositional logic combine to determine that Peirce's Law can be proven, not just the rules for the conditional" (Rossberg, 2006, p. 221).

But is this view compatible with logical inferentialism? Rossberg motivates it by appealing to the non-conservativeness of classical *propositional* logic: "whoever wants to hold on to classical propositional logic has to reject the molecular view [logical atomism, see *supra* § 6.1.1] in favour of a more holistic approach" (Rossberg, 2006, p. 221). I find this motivation unconvincing, however. As we saw Chapter 7, there *are* separable formalizations of classical logic. Hence, the failure of separability exhibited by the standard formalizations of classical logic is not itself a reason for the classical logician to commit herself to logical holism. The non-conservativeness of higher-order logics and infinite extensibility of higher-order concepts is an isolated and distinctive phenomenon—one that needs to be understood, and that cannot be assimilated to the non-conservativeness of the standard formalizations of classical logic.

D.2.4 Wright on higher-order logics

Perhaps it is a mistake to think that we learn something new when we come to grasp the rules for some n-order quantifiers, $n > 1$. Consider the standard I- and E-rules for the *third-order* quantifiers:

Example 15. *The third-order universal quantifier:*

$$\frac{\Phi[T/\Xi^n]}{\forall\Xi^n\Phi(\Xi^n)}\ \forall^3\text{-}I \qquad \frac{\forall\Xi^n\Phi(\Xi^n)}{\Phi(T^n)}\ \forall^3\text{-}E$$

The usual restrictions apply. These rules are just like the rules for first- and second-order logic, except that the variables range over properties of properties, as opposed to, respectively, objects and properties of objects. But then, why not replace these rules with just *one* schematic rule—one applying to *all* kinds of variables, and hence all possible orders? Thus, Rossberg considers the possibility that we may

> replace the specific rules for (monadic) universal quantifier elimination
>
> $$\frac{\forall x\Phi(x)}{\Phi(t)}\,\forall^1\text{-E} \qquad \frac{\forall X^n\Phi(X^n)}{\Phi(T^n)}\,\forall^2\text{-E} \qquad \frac{\forall \Xi^n\Phi(\Xi^n)}{\Phi(\mathbf{T}^n)}\,\forall^3\text{-E} \qquad \cdots$$
>
> with one general rule
>
> $$\frac{\forall X^i\Phi(X^i)}{\Phi(\Xi^i)}\,\forall^{i+1}\text{-E}$$
>
> and analogously for the introduction rule for the universal quantifier (and the rules for the existential quantifier. (Rossberg, 2006, p. 217-8).

On this view, no non-conservativeness ever arises, since the rules for the quantifiers, both first-order and higher-order, are all grasped, so to speak, in one flash.

Crispin Wright has recently argued along similar lines, although he restricts his claim to the higher-order quantifiers (more on this restriction in a moment):

> epistemologically, it is a mistake to think of higher-order quantifiers as coming in conceptually independent layers, with the second-order quantifiers fixed by the second-order rules, the third-order quantifiers fixed by the third-order rules, and so on. Rather it is the entire series of pairs of higher-and higher-order quantifier rules which collectively fix the meaning of quantification at each order: there are single concepts of higher-order universal and existential generalisation, embracing all the orders, of which it is possible only to give a schematic statement. (Wright, 2007a, p. 24)

According to Wright, there *is* a single general principle behind our understanding of the higher-order quantifiers: we grasp them all in one flash, via a grasp of their common *schematic representation*. On this assumption, higher-order logics are no longer incompatible with inferentialist orthodoxy, and the non-conservativeness of higher-order logics is merely apparent. As soon as we master the rules for the second-order quantifiers, we master the rules for the quantifiers of *any* order. In particular, when we grasp the rules for the second-order quantifiers, we are already able in principle to derive $\forall f \forall x(\mathsf{PA}_2{}^* \to G_2^*)$.

However, it might be objected, although it seems right to say that the meanings of first-order, second, ..., universal quantifiers have a common core (which is, roughly, captured by the common form of the I- and E-rules), there is also a dimension along which they differ in meaning. To understand a quantifier, we need to know what are the values of its bound variable: something that varies between quantifiers of different orders, and is not captured by the schematic rules for $\forall^n$ and $\exists^n$. More is required for deriving $\forall f \forall x(\mathsf{PA}_2{}^* \to G_2^*)$ than our grasp of the schematic rules for the quantifiers alone.[6]

D.3 Understanding higher-order concepts

I suggest that inferentialists should rather embrace Dummett's suggestion that the uses of higher-order expressions are infinitely extensible, but reject his assumption that all the infinitely extensible uses of a higher-order expression must be derivable by means of a single "general principle". The higher we go in the hierarchy of higher-order logics, the more uses of the higher-order quantifiers we validate. As Shapiro puts it, there is a slight, and constant, "change of subject". This need not entail, though, that our understanding of the higher-order quantifiers is radically unstable. Nor, I wish to argue, does it need to entail that, in order to be competent in the use of a higher-order quantifier, one needs to grasp infinitely many rules. But how can inferentialists account for the *meaning*, and for our *understanding*, of the higher-order quantifiers? There are two main options.

On the first option, inferentialists may insist that we come to learn the meanings of the higher-order quantifiers *in stages*: we begin with the first-order quantifiers, we then add the second-order ones, and so on. The process may stop at any point. The question arises, then, as to what is necessary and sufficient in order to acquire each higher-order concept at each stage. For every $n \in \omega$, let $\forall^n$'s *core uses* be whatever use of $\forall^n$ can be derived by means of the $\forall^n$-rules in $\forall^n$'s language. Now say that $\forall^n$'s *core meaning* is the set of all core uses, and that our *minimal understanding* of $\forall^n$ is given by our grasp of the core uses. Then, the thought is that one's minimal understanding of $\forall^n$ is necessary and sufficient for acquiring the concept of n^{th}-order quantification. To be sure, minimal understanding can

[6]Thanks to Bob Hale for helpful comments on this point. In conversation, Wright has counter-objected that, on his view, we do not need to grasp what the values of a quantifier's bound variable are, in order to understand that quantifier. Even conceding this point, though, an even more pressing issue remains: how far do our single concepts of existential and universal generalization extend? (Greg Restall asked this question at a conference in Aberdeen.) Are these the same concept we apply when we quantify, say, over all countable ordinals? To be sure, this is a pressing question for everybody. However, I submit, it is even more pressing if we assume, with Wrigtht, that there are single concepts of existential and universal generalization—after all, one might want to know about the *extension* of these concepts.

be improved—for instance, by adding new rules for quantifiers of strictly higher-orders, thereby validating new uses at the lower levels. But it does not need to. We are not compelled to *climb up* the hierarchy, in order to know what $\forall^n$ means. For any $n \in \omega$, we can minimally understand the rules for the n^{th}-order quantifiers, without thereby grasping the rules for the $n + 1^{th}$-order quantifiers. Since one's minimal understanding of the higher-order quantifiers is trivially unaffected by the extensions of the language, inferentialists can still maintain that our understanding of the higher-order quantifiers is essentially a matter of grasping their introduction and elimination rules. Thus, one may minimally understand the second-order quantifiers without thereby being willing to accept $\forall f \forall x(\mathsf{PA_2}^* \to G_2^*)$, even if this is a sentence in the language of pure second-order logic. Yet, this is precisely what one would expect: knowledge of the rules for the second-order quantifiers alone does not warrant our acceptance of $\forall f \forall x(\mathsf{PA_2}^* \to G_2^*)$.

It may be objected that the notion of minimal understanding is theoretically ill-motivated: it only serves the purpose of stabilizing our understanding of expressions with an unstable meaning. But it is not. On the assumption that one could understand $\forall^n$ without thereby understanding $\forall^{n+1}$, the notion of minimal understanding precisely defines what *that* understanding must consist in.

Alternatively, and perhaps more plausibly, inferentialists may reject the assumption that our understanding of a logical operator \$ is constituted by the stock of \$-theorems and \$-rules we are in principle able to prove at any given moment.[7] Rather, they may say, our understanding of \$ is constituted either by our knowledge of the \$-rules or by our willingness to infer, under certain conditions, according to these rules.[8] Likewise, inferentialists may argue, the sense of a logical operator \$ is not to be identified with the stock of \$-uses we are able to validate at any give time, but it is constituted by the means of proof by means of which these uses can be validated, i.e. by \$'s I- and E-rules. On *this* view, both our understanding and the sense of a higher-order quantifier are stable, irrespective of whether we keep adding quantifiers of strictly higher orders.

On either of the foregoing options, the orthodox inferentialist's requirement that valid inferences are strictly analytic, i.e. derivable by means of the rules for the logical operators occurring in them, is rejected as being too strict, and ultimately unmotivated. As we have seen, once we reject the assumption that, for any logical operator \$, the \$-rules must be complete, the inferentialist's demand for the separability of understanding no longer in general requires that admissible

[7] Thanks to Dag Prawitz for helpful comments on this point.

[8] See *supra*, § 4.3.

systems be separable. On the other hand, inferentialists may still claim that higher-order logics are *loosely analytic*: Gödelian sentences such as $\forall f \forall x(\mathsf{PA}_2{}^* \to G_2^*)$ can nevertheless be proved by means of rules—in this case, the rules for the third-order quantifiers—that are themselves strictly analytic.

Appendix E

Proof-theoretic consequence

In this Appendix, I quickly introduce the proof-theoretic account of validity, as presented in Prawitz (1985). I then discuss a possible objection, and I show that the account only works for *first-order* logics. I conclude by suggesting how one could define a more simple-minded, but, I think, more adequate, proof-theoretic account of validity—one which, contrary to Prawitz's, can be applied to first- and higher-order logics alike.

E.1 Prawitz's account of validity

Let an argument—a step-by-step deduction—be *closed* if it has no undischarged assumptions and no unbound variables, and let us say that it is *open* otherwise. Let an *immediate subargument* of a closed argument Π be an argument for the premises of Π's last inference rule, and let us call an argument *canonical* if it ends with an introduction rule, and it contains valid arguments for its premises. Thus, for instance, the arguments below are both canonical, but the argument on the left is open, since it an undischarged assumption, A, and the argument on the right is closed, since does not contain undischarged assumptions or unbound variables:

$$\frac{A}{A \vee B} \qquad\qquad \frac{\begin{array}{c}[A]\\ \mathcal{D}\\ B\end{array}}{A \rightarrow B}$$

where $\mathcal{D}$ represents a valid argument from A to B. Finally, let us assume that a set of justification procedures $\mathcal{J}$ for transforming non-canonical arguments in canonical arguments is available: one can always reduce arguments ending with an application of an elimination rule to arguments whose last step is taken into accordance with one of the introduction rules of the main logical operator of the argument's conclusion.

With these assumptions in place, the validity of an argument Π with respect to its set of justification procedures $\mathcal{J}$ may be defined as follows (Prawitz, 1985, pp. 164-165). If Π is a *closed* argument, $\langle \Pi, \mathcal{J} \rangle$ is valid if and only if either (i) Π is in canonical form and each immediate subargument Π' of Π is valid with respect to $\mathcal{J}$, or Π is not in canonical form, but it can be transformed into an argument for which (i) holds, by successive applications of the operations in $\mathcal{J}$. If Π is a *open* argument, on the other hand, $\langle \Pi, \mathcal{J} \rangle$ is valid if and only if all closed instances Π' of Π that are obtained by substituting for free parameters closed terms and for free assumptions closed arguments for the assumptions, valid with respect to an extension $\mathcal{J}$ of $\mathcal{J}'$, are valid with respect to $\mathcal{J}'$. We may then say that A is a logical consequence of a finite set of premises Γ if there is a valid argument from Γ to A. In short: the validity of whole of logic is reduced to the primitive validity of a small set of intuitively valid inference rules, on the assumption that the Fundamental Assumption holds.

Prawitz's definition may seem circular, since the notion of a canonical argument is defined in terms of the notion of a valid argument, and *vice versa*. But this appearance is deceiving. In *first-order* logics, in order to check whether an argument for a complex statement is valid, we only need to check whether the arguments for less complex statements, its immediate subarguments, are valid. In turn, in order to check whether an argument is canonically valid, we only need to verify the validity of arguments for less complex conclusions, *at least if the premises of our I-rules are logically less complex than their conclusion.*[1] Since sentences are finite, it follows that the process will terminate at some point.

An example may prove useful. Where $\mathcal{D}_1$ is a valid closed argument, the following non-canonical argument

$$\dfrac{\begin{matrix}\mathcal{D}_1\\ A \wedge B\end{matrix}}{A}$$

is valid if and only if each of its instances can be transformed into a canonical argument. But they can. For $\mathcal{D}_1$ must be either valid or reducible in canonical form, i.e. in an argument ending with an application of $\wedge$-I. This means that $\mathcal{D}_1$ itself contains valid arguments $\mathcal{D}_2$ and $\mathcal{D}_3$ such that the following is also a valid argument:

[1]We shall return to this point in a moment.

$$\frac{\begin{array}{cc}\mathcal{D}_2 & \mathcal{D}_3 \\ A & B\end{array}}{A \wedge B}$$

But then, there *is*, after all, a valid argument for A, viz.

$$\begin{array}{c}\mathcal{D}_2 \\ A\end{array}$$

Since $\mathcal{D}_2$ is valid, it is either canonical or, by the Fundamental Assumption, it is reducible in canonical form, as required.

It may objected that the foregoing definition requires a distinction between canonical and non-canonical ways of establishing *atomic* statements. After all, in our example, the argument from $\mathcal{D}_1$ to A is valid only if $\mathcal{D}_2$ is either canonical or can be reduced in canonical form. Hence, it would seem, if A is atomic, inferentialists *must* be committed to applying the Fundamental Assumption to atomic statements. However, whereas the distinction between canonical and non-canonical arguments can be sharply drawn for compound statements, the prospects for drawing the distinction for atomic statements look rather bleak. My suggestion is that inferentialists can circumvent the problem by simply refusing to draw a distinction where a distinction cannot be drawn. They may stipulate that all acceptable ways of establishing atomic statements count as canonical, thus forcing the Fundamental assumption to hold—trivially—for atomic statements. Thus, if I have heard on the news that Silvio Berlusconi has won the regional elections in Italy, then there is a canonical argument for 'Silvio Berlusconi has won the regional elections in Italy'.[2]

E.2 Syntactic validity and higher-order logics

A more serious objection, I take it, is that *the definition breaks down for higher-order logics*. Why? The reason is that, as we have seen in § 6.4.2, the I-rules of the higher-order quantifiers are not guaranteed to satisfy Dummett's (unstarred) complexity condition: whether they do or not depends on how rich our language is. But this means that, in some of their applications, their premises will be of greater logical complexity than their conclusion. Thus, in order to verify the validity of a canonical argument, one may need to verify the validity of an argument for a more complex conclusion. Hence, the foregoing recursive definitions of valid argument and canonical argument break down for higher-order logics: the process is not anymore guaranteed to stop at some point.

[2]Thanks to Ian Rumfitt and Timothy Williamson for pressing me on this point.

One of the virtues of Prawitz's definition of validity is that it avoids equating logical consequence with derivability *in a single deductive system*. Thus, Prawitz writes:

> If G is a Gödel sentence in a formalization of Peano arithmetic with the axioms Γ for which we can see intuitively that G follows from Γ, then, provided that it can be seen that G follows from Γ with the help of a language that can be analyzed in the way proposed here, there is a logically valid argument for G from Γ, i.e., G is a logical consequence of Γ in the sense proposed here. (Prawitz, 1985, p. 166)

Although G is not provable in Peano Arithmetic, we can informally prove, outside of Peano Arithmetic, that it is a consequence of what we are implicitly committed to when we accept its axioms. But then, one can formalize this informal proof in an extended system, which will in turn have its own Gödel sentence. And so on. As Dummett puts it: " the class of [the] principles [of proof] cannot be specified once and for all, but must be acknowledged to be an indefinitely extensible class" (Dummett, 1963a, p. 199). However, I suggest, if the definition cannot be applied to higher-order logics, this is a Pyrrhic victory. As we have seen in § 6.4.2, the rules for the higher-order quantifiers seem proof-theoretically unexceptionable: they respect the complexity condition*, and they are perfectly harmonious.

A possible way out of the problem would be to define logical consequence as derivability in any deductive system whose I-rules are single-ended and satisfy the complexity condition*, and whose E-rules are in harmony with the corresponding I-rules. Thus, the relation of logical consequence would be indefinitely extensible, as Dummett suggests, and the foregoing proof-theoretic account of validity would not be undermined by Gödel's First Incompleteness Theorem. On the other hand, the account could be applied to higher-order logics: for instance, we may plausibly take the relation of logical consequence to be indefinitely extended by the I- and E-rules of quantifiers of increasingly higher orders.[3]

Let me conclude by briefly considering a possible objection. At the very beginning of his book on logical consequence, John Etchemendy considers the possibility that consequence be defined as "derivability in some deductive system or other". However, he objects that it cannot be so defined, since "any sentence is derivable from any other in *some* such system". He concludes that

> at best we might mean by "consequence" derivability in some *sound*

[3]Rossberg (2006) offers an argument along these lines, but suggests that the criteria for selecting admissible deductive systems must be pragmatic, rather than proof-theoretic.

> deductive system. But the notion of soundness brings us straight back to the intuitive notion of consequence. (Etchemendy, 1990, pp. 2-3)

Because every formula is provable in *some* system, inferentialists need to find criteria for selecting the correct systems by means of which consequence is to be proof-theoretically defined. We may agree that this is one of the lessons inferentialists have learned from Prior's `tonk` and its ilk. Etchemendy contentiously assumes that the natural criterion is *soundness*, and maintains that it in turn presupposes a prior grasp of logical consequence—the very notion we are trying to account for. It should be clear, however, that Etchemendy's objection is off target in the present context. Our proposed criteria for identifying admissible rules, and admissible deductive systems, are the complexity condition*, single-endedness, and harmony—not soundness.

Index

WVER, 51
AG_a, 54
AG_b, 54
AG, 45
AG^+, 53, 54
$AG^+_\neg$, 53
ARU, 274
ARU_{TW}, 273
BIV, 13, 14
CONTR, 56–58
CR, 162, 165, 243, 245
CR^{ho}, 252, 254, 263
CT5, 62
DEC, 36, 42
DEC*, 54, 56
DIST, 63
DN, 162, 163
DN-I, 164
Dilemma, 181
EC, 16
EFQ, 210, 212, 213
ET, 13
FACT, 63
$FAC_\Diamond$, 52, 54
IVER, 76, 84
KP, 16, 33, 34, 65
LEM, 13, 14, 58, 162, 181, 249, 262
LNC, 15, 48, 54, 58, 262
LNC*, 230
ME, 63
MR, 29
MS, 110
MTP, 110
NEC, 63
PA, 200
PA_2, 203
PL, 194
RED*, 230, 239
RED^*_{cl}, 230, 241, 242, 250
RED^*_{int}, 230, 241
ST, 102
SVER, 64, 65, 286
TK**, 293
TK*, 290
TK, 288
Theorem 5, 61
UND*, 33, 34
VAL, 259
WAVER, 286, 294
WM, 68
WVER, 16, 42–44, 48, 50–55, 59, 64–66, 76, 82–84, 87, 88, 90, 286, 294, 295
WVER*, 88
NHcp, 264

NBcp, 230, 232–234, 236
NBcp^{+}, 236, 238
NHcp, 245, 251, 252, 256, 262, 263
Ncp^{DN}, 197
Nip^{-}, 186

Absurdity, 10, 162, 211, 232
Absurdity as a logical punctuation sign, 10, 213, 214, 254, 255
Absurdity, Dummett's account of, 210
Absurdity, Prawitz's account of, 210
Absurdity, Prawitz's interpretation of, 255
Achourioti, T., iv
Acquisition Challenge, 14, 18, 23, 25, 267
Admissible models, 109
Aimar, S., iv
Analyticity of logic, 101, 102, 129, 188
Anti-realism, moderate, 83
Apodictic numbers, 82
Argument from Analyticity, 188, 205
Argument from Circularity, 23
Argument from Harmony, 6, 8, 162, 188
Argument from Separability, 6, 8, 188, 196, 198, 215
Argument from Structural Conservativeness, 190
Aristotle-consistency, 109
Assertibility-conditions, 268
Assertion, 237, 238

Basic Law V, 187
Basic Revisionary Argument, 3, 10, 14, 33, 35, 39–43, 45, 48, 49, 55, 71, 93, 262–264, 277, 280
Basic statements, 67, 71
Beall, JC, iii, 1, 259
Belnap, N., 183, 190
Bendall, K., 170, 190, 193, 197, 239, 304
Berlusconi, S., 226, 309
BHK interpretation of the logical constants, 208
Bilateralism, 10, 228, 230, 231, 256, 262
Boche, 113, 114, 118, 119
Bogard, A., iv
Boghossian, P., 112
Bonardi, P., iv
Brandom, R., 100
Brouwer, J., 1, 35, 37, 262
Brown, G., 226
Bullet, 171, 176, 179–181, 187

Canonical argument, 307
Canonical grounds, 139, 210
Caret, C., iii, iv
Carnap, C., 101
Carnap, R., 9, 128, 170, 214, 222, 233, 239, 240
Carnap. R., 205
Carroll, L., 104, 132
Categoricity, 216, 217, 298
Categoricity Problem, 9, 170, 205, 206, 222, 233, 234, 254
Church, A., 2, 61, 81, 239, 240
Classical *reductio*, 162, 173, 176, 181, 196, 198, 219, 245, 249, 263
Classical Dilemma, 176
Classical propositional logic, 194, 206, 222, 232, 251, 302
Closed Argument, 307
Closure, 42
Cobreiros, P., iv
Coffa, A., 103
Cogburn, J., 29, 31, 43
Compactness, 200
Completeness, 8, 183

Complexity condition, 178, 179, 181, 188, 202, 204, 309
Complexity condition (definition), 179
Complexity condition*, 181
Complexity condition* , 310, 311
Complexity condition*, definition, 181
Compositionality, Principle of, 178
Comprehension Principle, 202, 298
Conservativeness, 8, 102, 182, 187, 188, 190, 216, 217
Conservativeness, strong, 183
Conservativeness, structural, 190
Conservativeness, weak, 183
Consistency, 109
Contraction, 182, 223, 258
Cook, R., 221, 222
Coordination principles, 231
Correctness valuations, 233
Cozzo, C., iii, 35, 51, 66, 73, 85, 116, 117, 119
Curry's Paradox, 2, 258, 259
Cut, 221, 230
Cut Elimination, 220

Davies, R., 144
de Campos Sanz, W., 159
de Jongh, D., iv
de Vidi, D., 66
De Vidid, D., 67
De, M., iv
Decidability, 16, 26, 267
Degree (of a formula), 252
Denial, 10, 228, 230, 241, 256
Denial, and assumptions, 241
Determination Thesis, 96, 112, 170, 181, 199, 205, 218
Dialetheism, 55, 58, 263
Dietz, R., iv
Dilemma, 162
Disjunction, 161, 225, 245
Disjunction, classical, 10, 234, 245, 246
Disjunction, non-constructive uses of, 137
Disjunction, Weir's rules for, 156
Disjunctive Syllogism, 235
Disjunctive syllogism, 162
Distributive law, 149
Divers, J., iv
Double Negation Elimination, 53, 70, 162–164, 167, 173, 196, 197, 219, 233, 249
Dummett's conjecture, 187
Dummett, E., 310
Dummett, M., 1, 4, 6, 14, 15, 18, 20, 24, 26–29, 31–34, 38, 61, 67, 69, 71, 75–77, 79, 82–84, 88, 93, 95, 100, 102, 108, 113, 120, 126, 127, 131, 134, 136, 137, 139, 143, 163, 164, 166, 170, 177–181, 184, 192, 199, 206, 207, 209–211, 214, 229, 230, 237, 238, 251, 258, 264, 267, 268, 270–272, 275, 294, 310
Dutilh Novaes, C., iv
Dyckhoff, R., iv, 145, 146, 162

E-rules, 95
Edgington, D., 66, 137, 286–294
Edgington, E., 66, 286
Empirical statements, 80
Equivalence Thesis, 13, 31, 280
Etchemendy, J., 310, 311
Ex falso quodlibet, 162, 210, 213, 251
Existential quantifiers, Weir's rules for the, 157
Expansions, 144

Falomi, M., iv
Field, H., iii, 1, 10, 13, 110, 115, 193, 201, 257–260
First Incompleteness Theorem, 200, 310
Fitch, F. B., 2, 61, 68
Fixedly operator, 287
Florian, S., 168
Florio, S., iii, 81
Forbes, G. A., iv
Force signs, 230
Francez, N., 145, 146, 162
Frege, G., 100, 122, 187, 229, 230, 270
Full invertibility, 153, 163, 232, 234, 246
Full invertibility, definition, 161
Fundamental Assumption, 6, 97, 99, 135–137, 181, 189, 198, 204, 210, 215, 223, 234, 244, 245, 249, 300, 308, 309

Gödel, K., 28, 199
Gödelian Optimism, 28, 36
Gabbay, M., iv
Garson, J. W., 206, 216–218
GE harmony, 152
GE schema, 147, 163
General elimination harmony, 7, 146, 148, 151, 163, 171, 234
Generalized Inversion Principle, 146, 152, 163
Gentzen, G., 95, 130, 132, 135, 139, 142, 148, 153, 178, 220, 221
Gibbard, P., 241
Girard, JY, 107
Global holism, 172
Gregory, D., iii, 105, 158, 196, 243
Ground, 147, 155, 160

Hacking, I., 101, 184, 221
Halbach, V., iv, 201
Hale, B., iii, 49, 108, 243, 262, 285, 304
Hand, M., 66
Harmony, 7, 131, 136, 138, 155, 157–159, 161, 188, 213, 232, 234, 244, 246, 258, 263, 310, 311
Harmony and consistency, 172
Harmony as full invertibility, 7, 171
Harrington, K., iv
Hart, W., 61, 64
Hauptstaz, 220
Henkin models, 299
Henkin semantics, 298
Hierarchical approaches to the Paradox of Knowability, 81, 85
Higher-order logics, 9, 170, 201, 309, 310
Higher-order logics, incompleteness of, 205
Higher-order quantifiers, schematic rules for the, 303
Higher-order quantifiers, understanding of, 303, 304
Higher-order rules, 247
Hilbert, D., 28, 37
Hjortland, O. T., iii, 240, 242
Humberstone, L., iv

I-rules, 95
Iacona, A., iv
Immediate subargument, 307
Incompleteness, 9, 199
Incurvati, L., iii, 14, 85, 211, 212, 236, 240
Indefinite extensibility, 1, 264, 301, 310
Infefinite extensibility, 68
Inferences, knowledge-transmitting, 104
Inferential packages, 193
Intrinsic harmony, 140, 161, 187
Introduction rule, 174

Introduction rules, Prawitz's definition, 174
Intuitionistic logic, 238, 262
Inversion Principle, 140
Inversion Principle, Weir's, 154, 155
Inversion Principle, Weir's (emended version), 159, 160
Inversionprinzip, 140

Janssen-Lauret, F., iv
Jenkins, C., iii, 66

Künne, W., 76, 78
Keefe, R., iii
Kneale, W., 96, 128
Knowability Principle, 16, 65, 68, 70, 93, 286, 287
Knowledge, implicit, 21
Kripke, S., 1, 115
Kvanvig, J., 66

Löwenheim-Skolem theorems, 298
Law of Double Negation Elimination, 176
Law of Excluded Middle, 13, 14, 40, 42, 54, 121, 137, 162, 176, 223, 234, 249, 257, 264
Law of Excluded Middle, canonical proof of, 248
Law of Non-Contradiction, 15, 48, 54, 57
Leblanc, H., 196
Liar Paradox, 1, 258, 259
Linsky, B., 81
Local completeness, 144, 152, 183, 191, 194
Local completeness (definition), 145
Local peaks, 139, 141, 142
Local soundness, 144, 160
Logical atomism, 172, 175, 178
Logical consequence, 308
Logical holism, 172, 189, 193, 195, 198, 302
Logical inferentialism, 5, 58
Logical molecularism, 172, 178
Logical revision, 120
Loose analyticity, 186, 202, 305
Lorenzen, P., 140

Magidor, O., 83
Manifestability Principle, 20, 57
Manifestation Challenge, 14, 15, 18, 20, 23, 26–29, 267
Manifestation Requirement, 29, 32, 56, 68, 278, 294
Manifestation Requirement, Weak, 68
Martino, E., 72, 74, 75
Material implication, 179
Maximum formula (definition), 141
Maximum formulae, 142, 251
McDowell, J., 107
McGee, V., 112, 114, 121, 122, 125, 126, 193, 226, 227
McGinn, C., 61, 64
Milne, P., iv, 97, 165, 168, 170, 189, 193, 304
Minimal logic, 161
Missing Guarantee, thesis of the, 35, 40, 47, 287
Modal logic, labelled systems of, 136
Models, admissible, 107
Molecularism, 173
Molecularism, strong, 172
Molecularism, weak, 172
Moruzzi, S., iv
Multiple-conclusions, 10, 220, 222, 224, 225, 256, 262

Multiple-conclusions, Restall's interpretation, 226, 227
Murzi, J., 14, 81, 240

Natural semantics, 216, 217
Negation, classical (sequent rules), 221
Negation, external, 10, 236, 241, 243, 256, 262
Negation, intuitionistic, 233
Negation, intuitionistic (sequent rules), 221
Negri, S., 146
Nelson, D., 55
Non-conservativeness, 200, 203, 302
Non-standard models of arithmetic, 299
Normalisation theorem, 142
Normalization, 142, 150, 219, 245, 251
Normalization theorem, 220
Normalization theorem (for **NHcp**), 253

Oms, S., iv
Open argument, 307
Operational rules, 94
Orsi, F., iv
Orthodox inferentialism, 182, 196, 200, 204, 297, 299, 300

Pagin, P., 272, 275
Paradox of Idealisation, 81, 84
Paradox of Idealisation, proof, 84
Paradox of Knowability, 4, 42, 61, 63, 67, 75, 77, 79, 81, 93, 286
Paradox of Knowability, general form of, 84
Paradox of the Knower, 85
Paseau, A., 81, 85
Payne, J., iv
Peano Arithmetic, 199
Peano Arithmetic (second-order), 298
Peirce's Law, 194, 222, 302
Peirce's Rule, 162, 194
Percival, P., 77
Permutation, 258
Permutative reduction procedures, 142, 150
Pfenning, F., 144
Piecha, T., 159
Popper, K., 96, 128
Possibilities, 287, 288
Post-consistency, 109
Prawitz, D., iv, 1, 6, 14, 20, 38, 66, 99, 120, 139, 140, 163, 164, 174, 176, 187, 199, 206, 207, 209, 214, 254, 264, 305, 307, 309
Priest, G., 1, 121, 130, 259–261
Principle of Bivalence, 13, 14, 18, 27, 29, 31, 33, 73, 264
Principle of Contravalence, 4, 56, 57
Prior, A., 112, 129, 131, 171, 311
Proof by cases, 224, 225
Proof-theoretic requirements, global, 131, 169, 258
Proof-theoretic requirements, local, 131, 169, 258
Proof-tokens, 72, 73
Proof-types, 72, 73
Protasi, S., iv
Pseudo *modus-ponens*, 2, 257
Puddifoot, K., iv
Purity, 173
Putnam, H., 32

Quandary, 52–54
Quantum disjunction, 8, 148–152
Quine, V. W. O., 124
Quine, W. V. O., 116, 130

Raatikainen, P., 9, 170, 205, 207, 211
Rabinowicz, W., 290
Radical revisionism, 196
Read, S., iii, 142, 146, 171, 176, 180, 199, 200, 264
Reduction steps, 141
Restall, G., iv, 226, 241, 304
Revisionism, radical, 121
Revisionism, theoretical, 121
Roadevin, C., iv
Roberts, S., iv
Rossberg, M., iii, 49, 203, 297, 301, 302, 310
Rule completeness, 183
Rule soundness, 183
Rule-purism, 174, 185, 186
Rumfitt, I., iii, 9, 104, 118, 214, 223, 228–231, 234, 244, 297
Russell's Paradox, 187

Sagi, G., iv
Salerno, J., iv, 14, 27, 33, 34, 40–42, 47
Salvi, V., iv
Sauchelli, A., iv
Schroeder-Heister, P., 146, 246, 247, 250
Second-order logic, 202
Segerberger, K., 290
Semantic Anti-realism, 16
Semantic externalism, 116, 123, 125
Semantic Realism, 15
Separability, 8, 175, 182, 186, 188, 195, 196, 258, 263
Separability and understanding, 194
Separability, strong, 183, 191, 234, 244
Separability, weak, 183, 234, 244
Sequent style natural deduction, 182
Sequents, 182
Set abstraction, 187
Sgaravatti, D., iv, 118
Shapiro, S., iii, iv, 28, 36, 130, 199, 201, 299, 301, 304
Shared content, 127
Sheerness, 173
Shoesmith, D. J., 225
Single Premise Argument, 4, 18, 29, 31, 32, 48, 55
Single-endedness, 173, 311
Smilean *reductio*, 231, 241
Smiley, T., 9, 225, 228, 231
Smith, M., iii
Smith, P., 211, 212, 236, 240, 241
Soames, S., 245
Solomon, G., 66, 67
Sorites Paradox, 13
Soundness, 183
Stökke, A., iv
Stability, 143
Standard Argument, 77, 79
Steinberger, F., iii, 142, 144, 150, 203, 250
Stipulation Thesis, 101
Stjernberg, F., iii
Strawson, F. P., 29
Strict analyticity, 186, 198, 201, 202
Strict Finitism, 84
Strict finitism, 82
Strohminger, M., iv
Strong intrinsic harmony, 7, 144, 146, 148, 161, 163, 171
Strong molecularism, 178
Strong Verificationism, 63, 64
Structural assumptions, 260
Structural rules, 190, 231
Subformula (definition), 174
Subformula property, 186, 253
Support, 147, 155, 160

Synthetic *a priori*, 300

Tarski, A., 108
Tennant, N., iv, 1, 14, 25, 27–29, 66, 67, 69, 70, 83, 102, 144, 146, 165, 175, 176, 199, 211, 214, 224, 277, 278, 300
Theories of meaning, assertibility-conditional, 268
Theories of meaning, full-blooded, 24, 108
Theories of meaning, modest, 24, 108
Theories of meaning, truth-conditional, 268
Tonk, 7, 129, 132, 148, 158, 165, 184, 213, 311
Total harmony, 184, 187
Transworld knowledge, 287–291
Trivialization Objection, 289–291
Truth, 26, 74, 267, 270
Truth, decidability of, 272
Truth-maker, 274
Truth-makers, 26, 267
Truth-predicate, 199
Truth-preservation, 101
Truth-rules, 199

Ultra orthodox inferentialists, 186
Ultra strict analyticity, 186
Undecidability, 16, 17, 20, 35
Undecidability, absolute, 17, 34
Understanding, 6, 20, 23, 25, 112–115, 121, 125, 126, 193–195, 272, 273, 304
Understanding, manifestability of, 22, 25, 26, 272–275
Unilateralism, 228
Unitype function, 75
Universal quantifier, inductive uses, 138
Usberti, G., iii, 72, 74, 75

Vagueness, 1, 70, 237, 264
Validity (for signed formulae), 233
Validity Curry, 259, 260
Validity of sequents, 222
Validity paradoxes, 259
Validity, proof-theoretic definition, 244, 307
Verification-transcendence, 56
Verification-transcendence, potential, 17, 19
von Plato, J., 146

Wansing, H., 175, 176
Weak Verificationism, 16, 63, 64, 77, 84
Weak Verifictionism, 76
Weakening, 182
Weir, A., 8, 153, 155, 156, 158
Whole Discourse Argument, 277
Williamson, T., iii, 4, 6, 15, 16, 18, 26, 61, 66, 68–71, 74, 76, 77, 85, 93, 112, 113, 115, 117–119, 121, 122, 125, 126, 128, 132, 267, 268, 270–272, 289, 294, 295, 297
Wittgenstein, L., 128
Woodward, R., iv
Wright, C., iii, 1–3, 14, 15, 26, 37, 39, 41, 42, 44–46, 48–55, 62, 65, 70, 77, 83, 89, 105, 106, 134, 203, 262, 264, 286, 301–304

Zardini, E., iv, 118

References

Aristotle: 1961, *Categoriae et Liber de Interpretatione*, Clarendon Press, Oxford.

Beall, J.: 2003, Fitch's paradox, verificationism and the knower paradox, *The Australasian Journal of Philosophy* **78**, 241–47.

Beall, J.: 2009, *Spandrels of Truth*, Oxford University Press, Oxford.

Beall, J. and van Frassen, B.: 2003, *Possibilities and Paradox: An Introduction to Modal and Many-Valued Logics*, Oxford: Oxford University Press.

Belnap, N.: 1962, Tonk, plonk and plink, *Analysis* **22**(6), 130–34.

Bendall, K.: 1978, Natural deduction, separation, and the meaning of the logical operators, *Journal of Philosophical Logic* **7**, 245–76.

Bendall, K.: 1979, Negation as a sign of negative judgement, *Notre Dame Journal of Formal Logic* **20**(1), 68–76.

Bermudez, J.: 2009, Truth, indefinite extensibility, and Fitch's paradox, *in* J. Salerno (ed.), *New Essays on the Knowability Paradox*, Oxford University Press, Oxford, pp. 76–90.

Boghossian, P.: 2003, Blind reasoning, *Proceedings of the Aristotelian Society* **77**(1), 225–48.

Boričić, B. R.: 1985, On sequence-conclusion natural deduction systems, *Journal of Philosophical Logic* **14**, 359–77.

Brandom, R.: 1994, *Making it explicit*, Harvard University Press, Cambridge, (Mass.).

Brandom, R.: 2000, *Articulating reasons*, Harvard University Press, Harvard (Mass.).

Brogaard, B. and Salerno, J.: 2002, Clues to the paradoxes of knowability: Reply to Dummett and Tennant, *Analysis* **62**, 143–50.

Brogaard, B. and Salerno, J.: 2008, Knowability, possibility and paradox, *in* V. Hendrix and D. Pritchard (eds), *New Waves in Epistemology*, Ashgate.

Brouwer, I. L.: 1908, The unreliability of logical principles, *in* A. Heyting (ed.), *Collected Works 1. Philosophy and Foundations of Mathematics*, Amsterdam: North-Holland, pp. 107–11.

Byrne, D.: 2005, Compositionality and the manifestation challenge, *Synthese* **144**, 101–36.

Carnap, R.: 1934, *Logische Syntax der Sprache*, Springer, Berlin. Page references are to translation in Carnap (2002).

Carnap, R.: 1943, *Formalization of Logic*, Harvard University Press, Cambridge, (Mass.).

Carnap, R.: 2002, *The Logical Synthax of Language*, Open Court, Chicago, IL.

Carroll, L.: 1895, What the tortoise said to Achilles, *Mind* **4**, 278–80.

Church, A.: 1944, Review: Formalization of Logic by Rudolf Carnap, *Philosophical Review* **53**(5), 493–8.

Church, A.: 2009, Anonymous referee reports, *in* J. Salerno (ed.), *New Essays on the Knowability Paradox*, Oxford University Press, Oxford, pp. 13–20.

Coffa, A.: 1991, *The semantic tradition from Kant to Carnap*, Cambrdige University Press, Cambridge.

Cogburn, J.: 2002, Logical revision re-revisited: The Wright-Salerno argument for intuitionism, *Philosophical Studies* **110**, 231–48.

Cogburn, J.: 2003, Manifest invalidity: Neil Tennant's new argument for intuitionism, *Synthese* **134**, 353–62.

Cogburn, J.: 2005, The logic of logical revision: Formalizing Dummett's argument, *Australasian Journal of Philosophy* **83**(1), 15–32.

Cook, R.: 2005, Intuitionism reconsidered, *in* S. Shapiro (ed.), *The Oxford handbbook for Philosophy of Mathematics and Logic*, Oxford University Press, Oxford, pp. 387–411.

Cook, R.: 2006, What's wrong with tonk?, *Journal of Philosophical Logic* **(2)**, 217–26.

Cozzo, C.: 1989, Bivalenza e transcendenza, *Teoria* **10**, 63–86.

Cozzo, C.: 1994, What can we learn from the paradox of knowability?, *Topoi* **13**, 71–78.

Cozzo, C.: 1998, Epistemic truth and the excluded middle, *Theoria* **43**, 2–40.

Cozzo, C.: 2008, On the Copernican turn in semantics, *Theoria* **74**, 295–317.

Davies, R. and Pfenning, F.: 2001, A modal analysis of staged computation, *Journal of the ACM* **48**(3), 555–604.

De Campos Sanz, W.: 2009, Inversion by definitional reflection and the admissibility of logical rules, *The Review of Symbolic Logic* **2**, 550–569.

De Vidi, D. and Solomon, G.: 2001, Knowability and intuitionistic logic, *Philosophia* **28**, 319–34.

De Vidi, D. and Solomon, G.: 2006, Empirical negation in intuitionistic logic, *in* D. De Vidi and T. Kenyon (eds), *A Logical Approach to Philosophy*, Springer, Berlin.

Dummett, M.: 1959, Truth, *Proceedings of the Aristotelian Society* **59**, 141–62. Page references are to reprint in Dummett, 1978b, pp. 1-25.

Dummett, M.: 1963a, The philosophical significance of Gödel's theorem, *Ratio* **5**, 140–155. Reprinted in Dummett, 1978b, pp. 186-201.

Dummett, M.: 1963b, Realism. Published for the first time in Dummett, 1978b, pp. 145-65.

Dummett, M.: 1973a, *Frege: Philosophy of Language*, Duckworth, London.

Dummett, M.: 1973b, The philosophical basis of intuitionistic logic, *in* H. Rose and J. Shepherdson (eds), *Logic Colloquium '73*, North-Holland, Amsterdam. Page references are to reprint in Dummett, 1978b, pp. 215-247.

Dummett, M.: 1975, Wang's paradox, *Synthese* **30**, 301–324. Page references are to reprint in Dummett, 1978b, pp. 248-268.

Dummett, M.: 1976, What is a theory of meaning II, *in* G. Evans and J. McDowell (eds), *Truth and Meaning*, Oxford University Press, Oxford, pp. 67–137.

Dummett, M.: 1977, *Elements of Intuitionism*, Oxford University Press, Oxford.

Dummett, M.: 1978a, Frege's distinction between sense and reference, *in* M. Dummett (ed.), *Truth and Other Enigmas*, Duckworth, pp. 116–44.

Dummett, M.: 1978b, *Truth and Other Enigmas*, Duckworth, London.

Dummett, M.: 1978c, What is a theory of meaning I, *in* S. Guttenplan (ed.), *Mind and Language (Wolfson College Lectures)*, Oxford University Press. Page references are to reprint in Dummett, 1993b, pp. 1-33 .

Dummett, M.: 1979, What does the appeal to use do for the theory of meaning?, *in* A. Margalit (ed.), *Meaning and Use*, Reidel, Dordrecht. Page references are to reprint in Dummett, 1993a, pp. 106-116.

Dummett, M.: 1987a, Reply to Dag Prawitz, *Michael Dummett: Contributions to Philosophy*, The Hague: Martinus Nijhoff Publishers, pp. 281–16.

Dummett, M.: 1987b, Reply to McDowell, *in* B. Taylor (ed.), *Michael Dummett: Contributions to Philosophy*, Martinus Nijhoff, pp. 253–268.

Dummett, M.: 1991a, *Frege: Philosophy of Mathematics*, Duckworth, London.

Dummett, M.: 1991b, *The Logical Basis of Metaphysics*, Harvard University Press, Harvard (Mass.).

Dummett, M.: 1993a, Realism and anti-realism, *The Seas of Language*, Harvard (Mass.): Harvard University Press.

Dummett, M.: 1993b, *The Seas of Language*, Oxford University Press, Oxford.

Dummett, M.: 1994, Reply to Prawitz, *in* B. McGuinness and G. Olivieri (eds), *The Philosophy of Michael Dummett*, Kluwer Academic Publishers, Dordrecht, pp. 292–98.

Dummett, M.: 1998, Truth from a constructive standpoint, *Theoria* **14**, 122–38.

Dummett, M.: 2000, *Elements of Intuitionism (2nd ed.)*, Oxford University Press, Oxford.

Dummett, M.: 2001, Victor's error, *Analysis* **61**(269), 1.2.

Dummett, M.: 2002, 'Yes', 'No' and 'Can't Say', *Mind* **111**, 289–295.

Dummett, M.: 2004, *Truth and the Past*, Columbia University Press, New York.

Dummett, M.: 2005, The justificationist's response to a realist, *Mind* **114**, 671–88.

Dummett, M.: 2006, *Truth and Reality*, Oxford University Press, Oxford.

Dummett, M.: 2007a, Reply to John Campbell, *in* R. E. Auxier and L. E. Hahn (eds), *The Philosophy of Michael Dummett*, Open Court, Chicago, pp. 301–13.

Dummett, M.: 2007b, Reply to McDowell, *in* R. E. Auxier and L. E. Hahn (eds), *The Philosophy of Michael Dummett*, Open Court, Chicago, pp. 267–382.

Dummett, M.: 2007c, Reply to Wolfgang Künne, *in* R. E. Auxier and L. E. Hahn (eds), *The Philosophy of Michael Dummett*, Open Court, Chicago, pp. 345–350.

Dummett, M.: 2007d, Reply to Wright, *in* R. E. Auxier and L. E. Hahn (eds), *The Philosophy of Michael Dummett*, Open Court, Chicago.

Dummett, M.: 2009, Fitch's paradox of knowability, *in* J. Salerno (ed.), *New Essays on the Knowability Paradox*, Oxford University Press, Oxford.

Edgington, D.: 1981, Meaning, bivalence and realism, *Proceedings of the Aristotelian Society* **81**, 153–73.

Edgington, D.: 1985, The paradox of knowability, *Mind* **94**, 557–68.

Edgington, D.: 2010, Possible knowledge of unknown truth. Forthcoming in *Synthese*.

Etchemendy, J.: 1990, *The Concept of Logical Consequence*, Harvard University Press, Harvard (Mass.).

Field, H.: 1991, Metalogic and modality, *Philosophical Studies* **62**, 1–22.

Field, H.: 2006, Compositional principles schematic raesoning, *The Monist* **89**, 9–27.

Field, H.: 2008, *Saving Truth from Paradox*, Oxford University Press, Oxford.

Fine, K.: 1975, Vagueness, truth and logic, *Synthese* **30**, 265–300.

Fitch, F.: 1963, A logical analysis of some value concepts, *Journal of Philosophical Logic* **28**, 135–42.

Florio, S. and Murzi, J.: 2009, The paradox of idealisation, *Analysis* **69**, 461–469.

Francez, N. and Dyckhoff, R.: 2009, A note on harmony. Forthcoming in *Journal of Philosophical Logic*.

Frege, G.: 1977, Negation, *in* P. T. Geach (ed.), *Logical Investigations*, Basil Blackwell, Oxford.

Garson, J. W.: 2001, Natural semantics: why natural deduction is intuitionistic, *Theoria* **67**(2), 114–137.

Garson, J. W.: 2010, Expressive power and incompleteness of propositional logics, *Journal of Philosophical Logic* **39**(2), 159–71.

Gentzen, G.: 1934, Untersuchungen über das logischen schliessen, *Math. Zeitschrift* **39**, 405–31.

Gentzen, G.: 1969, *Collected Papers*, North Holland, Amsterdam. Edited by M. E. Szabo.

Gibbard, P.: 2002, Price and Rumfitt on rejective negation and classical logic, *Mind* **111**(442), 298–303.

Girard, J.-Y.: 1999, On the meaning of logical rules I: syntax vs. semantics, *in* U. Berger and H. Schwichtenberg (eds), *Computational Logic*, Nato series F 165, Springer.

Girard, J. Y.: 2003, From foundations to ludics, *Bulletin of Symbolic Logic* **09**(2), 131–68.

Hacking, I.: 1979, What is logic?, *Journal of Philosophy* **76**, 285–319.

Halbach, V.: 2005, Axiomatic theoris of truth, *Stanford Encyclopedia of Philosophy* . **URL:** *http://plato.stanford.edu/entries/truth-axiomatic/*

Hale, B.: 1997, Modality, *in* B. Hale and C. Wright (eds), *A Companion to the Philosophy of Language*, Basil Blackwell.

Hale, B.: 2002, Basic logical knowledge, *in* A. O'Hear (ed.), *Logic, Thought and Language: Royal Institute of Philosophy Supplement 51*, Cambrdige University Press, Cambridge, pp. 279–304.

Hale, B. and Wright, C.: 2001, *The Reason's Proper Study - Essays towards a neo-Fregean Philosophy of Mathematics*, Oxford University Press, Oxford.

Hand, M.: 1999, Antirealism and falsity, *What is Negation?*, Kluwer Academic Publishers, Dordrecht, pp. 185–198.

Hand, M.: 2003, Knowability and epistemic truth, *The Australasian Journal of Philosophy* **81**(2), 216–28.

Hand, M.: 2009, Performance and paradox, *in* J. Salerno (ed.), *New Essays on the Knowability Paradox*, Oxford: Oxford University Press, pp. 283–301.

Harris, J. H.: 1982, What's so logical about the logical axioms?, *Studia Logica* **41**, 159–71.

Hart, W. D.: 1979, The epistemology of abstract objects: access and inference, *Proceedings of the Aristotelian Society* **53**, 153–65.

Hart, W. and McGinn, C.: 1976, Knowledge and necessity, *Journal of Philosophical Logic* **5**, 205–208.

Hjortland, O. T.: 2008, Proof-theoretic harmony and structural assumptions, *MS* .

Hodes, H.: 2004, On the sense and reference of a logical constant, *The Philosophical Quarterly* **54**(214).

Humberstone, L.: 1981, From worlds to possibilities, *Journal of Philosophical Logic* **10**, 310–39.

Humberstone, L.: 2000, The revival of rejective negation, *Journal of Philosophical Logic* **29**, 331–81.

Incurvati, L. and Murzi, J.: 2008, How basic is the basic revisionary argument?, *Analysis* **68**(4), 303–309.

Incurvati, L. and Smith, P.: 2010, Rejection and valuations, *Analysis* **69**(4), 3–10.

Jenkins, C.: 2007, Anti-realism and epistemic accessibility, *Philosophical Studies* **132**, 525–51.

Kallestrup, J.: 2007, If omniscient beings are dialethist, so are anti-realists, *Analysis* **67**(3), 252–4.

Keefe, R.: 2000, *Theories of Vagueness*, Cambridge: Cambridge University Press.

Khlentzos, D.: 2004, *Naturalistic Realism and the Antirealist Challenge*, MIT Press, Cambridge, (Mass.).

Kneale, W.: 1956, The province of logic, *in* H. D. Lewis (ed.), *Contemporary British Philosophy*, George Allen & Unwin Ltd, London, pp. 237–61.

Kripke, S.: 1965, Semantical analysis of intuitionistic logic, *in* M. Dummett and J. N. Crossley (eds), *Formal Systems and Recursive Functions*, Amsterdam: North-Holland Publishing Co.

Kripke, S.: 1975, Outline of a theory of truth, *Journal of Philosophy* **72**, 690–716.

Künne, W.: 2007, Two principles concerning truth, *in* R. E. Auxier and L. E. Hahn (eds), *The Philosophy of Michael Dummett*, Open Court, Chicago, pp. 315–34.

Kvanvig, J.: 1995, The knowability paradox and the prospects for anti-realism, *Nous* **29**, 481–500.

Kvanvig, J.: 2006, *The Knowability Paradox*, Oxford University Press, Oxford.

Leblanc, H.: 1966, Two shortcomings of natural deduction, *The Journal of Philosophy* **63**(2), 29–37.

Leivant, D.: 1994, Higher-order logic, *in* D. M. Gabbay, C. J. Hogger, J. A. Robinson and J. Sieckmann (eds), *Handbook of Logic in Artificial Intelligence and Logic Programming. Volume 2: Deduction Methodologies*, Oxford University Press, Oxford, pp. 230–321.

Linsky, B.: 2009, Logical types in some arguments about knowability and belief, *in* J. Salerno (ed.), *New Essays on the Knowability Paradox*, Oxford: Oxford University Press.

Lorenzen, P.: 1955, *Einführung in die operative Logik and Mathematik*, Springer, Berlin.

MacFarlane, J.: 2003, Future contingents and relative truh, *Philosophical Quarterly* **53**, 321–336.

MacFarlane, J.: 2005, Logical constants, *Stanford Encyclopedia of Philosophy* .

Magidor, O.: 2007, Strict finitism refuted?, *Proceedings of the Aristotelian Society* **107**, 403–11.

Martino, E. and Usberti, G.: 1994, Temporal and atemporal truth in intuitionistic mathematics, *Topoi* **13**, 83–92.

McDowell, J.: 1981, Anti-realism and the epistemology of understanding, *in* H. Parrett and J. Bouveresse (eds), *Meaning and Understanding*, De Gruyter, New York, pp. 225–48.

McDowell, J.: 1987, In defence of modesty, *in* B. Taylor (ed.), *Michael Dummett: Contributions to Philosophy*, Martinus Nijhoff Publishers, Dordrecht, pp. 59–80.

McDowell, J.: 1997, Another plea for modesty, *in* R. Heck (ed.), *Language, Thought and Logic: Essays in Honour of Michael Dummett*, Clarendon Press, Oxford.

McDowell, J.: 2007, Dummett on truth conditions and meaning, *in* R. E. Auxier and L. E. Hahn (eds), *The Philosophy of Michael Dummett*, Open Court, Chicago, pp. 351–366.

McGee, V.: 1985, A counterexample to *modus ponens*, *The Journal of Philosophy* **82**, 462–71.

McGee, V.: 1997, How we learn mathematical language, *Philosophical Review* **106**(1), 35–68.

McGee, V.: 2000, Everything, *in* G. Sher and R. Tieszen (eds), *Between Logic and Intuition. Essays in Honor of Charles Parsons*, Cambrdige University Press, Cambridge, pp. 54–78.

McGee, V.: 2004, Ramsey's dialetheism, *in* B. A.-G. G. Priest, JC Beall (ed.), *The Law of Non-Contradiction. New Philosophical Essays*, Oxford University Press, pp. 276–294.

Miller, A.: 2003, What is the Acquisition Argument?, *in* A. Barber (ed.), *Epistemology of Language*, Oxford University Press, Oxford, pp. 459–95.

Milne, P.: 1994, Classical harmony: Rules of inference and the meaning of the logical constants, *Synthese* **100**, 49–94.

Milne, P.: 2002, Harmony, purity, simplicity, and a "seemingly magical fact", *The Monist* **85**, 498–535.

Moriconi, E. and Tesconi, L.: 2008, On inversion principles, *History and Philosophy of Logic* **29**(2), 103–113.

Murzi, J.: 2008, *The Paradox of Knowability*, PhD thesis, University of Rome "La Sapienza", Rome.

Murzi, J.: 2009, Knowability and bivalence: intuitionistic solutions to the Paradox of Knowability, *Philosophical Studies* **149**(2), 269–81.

Murzi, J.: 2010a, Carnap's categoricity problem and the meanings of the logical constants. Under review.

Murzi, J.: 2010b, Inferentialism and understanding: on Williamson's case against logical inferentialism. Under review.

Murzi, J. and Hjortland, O. T.: 2009, Inferentialism and the categoricity problem: Reply to Raatikainen, *Analysis* **69**(3), 480–488.

Negri, S. and von Plato, J.: 2001, *Structural Proof Theory*, Cambridge University Press, Cambridge.

Nelson, D.: 1949, Constructible falsity, *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*, Amsterdam: North-Holland, pp. 208–225.

Pagin, P.: forthcoming, Compositionality, understanding, and proofs, *Mind* .

Paseau, A.: 2008, Fitch's paradox and typing knowledge, *Notre Dame Journal of Formal Logic* **49**(2), 153–176.

Percival, P.: 1990, Fitch and intuitionistic knowability, *Analysis* **50**, 182–7.

Percival, P.: 1991, Knowability, actuality and the metaphysics of context-dependence, *Australasian Journal of Philosophy* **69**, 82–97.

Pfenning, F. and Davies, R.: 2001, A judgmental reconstruction of modal logic, *Mathematical Structures in Computer Science* **11**, 511–40.

Philie, P.: 2007, Carroll's regress and the epistemology of logical laws, *Philosophical Studies* **134**, 183–210.

Popper, K.: 1947, New foundations for logic, *Mind* **56**(223), 193–235.

Prawitz, D.: 1965, *Natural Deduction*, Almqvist and Wiksell, Stockholm.

Prawitz, D.: 1973, Towards a fundation of a general proof theory, *in* P. Suppes, L. Henkin, A. Joja and G. C. Moisil (eds), *Logic, Methodology and the Philosophy of Science IV: Proceedings of the Fourth International Congress*, North Holland, Amsterdam, pp. 225–50.

Prawitz, D.: 1977, Meaning and proofs: on the conflict between classical and intuitionistic logic, *Theoria* **43**, 1–40.

Prawitz, D.: 1980, Intuitionistic logic: A philosophical challenge, *in* G. H. von Wright (ed.), *Logic and Philosophy*, Martinus Nijhoff Publishers, The Hague, pp. 1–10.

Prawitz, D.: 1985, Remarks on some approaches to the concept of logical consequence, *Synthese* **62**, 153–71.

Prawitz, D.: 1994, Review of 'The Logical Basis of Metaphysics', *Mind* **103**, 373–376.

Prawitz, D.: 1998a, Comments on Lars Bergström's paper: "Prawitz's version of verificationism", *Theoria* **64**, 293–303.

Prawitz, D.: 1998b, Comments on Michael Dummett's paper 'truth from the constructive standpoint', *Theoria* **64**.

Prawitz, D.: 2005, Logical consequence: a constructivist view, *in* S. Shapiro (ed.), *The Oxford handbbook of Philosophy of Mathematics and Logic*, Oxford University Press, Oxford, pp. 671–95.

Prawitz, D.: 2006, Meaning approached via proofs, *Synthese* **148**, 507–24.

Prawitz, D.: 2010, Perfect syllogisms, proofs, and knowledge. Unpublished manuscript.

Price, H.: 1990, Why "not"?, *Mind* **99**, 221–38.

Priest, G.: 2006a, *Doubt Truth to be a Liar*, Oxford University Press, Oxford.

Priest, G.: 2006b, *In Contradiction*, Oxford University Press, Oxford. Expanded edition.

Priest, G.: 2009a, Beyond the limits of knowledge, *in* J. Salerno (ed.), *New Essays on the Knowability Paradox*, Oxford University Press, Oxford, pp. 93–104.

Priest, G.: 2009b, Is the ternary relation depraved? Unpublished manuscript.

Priest, G.: 2010, Hopes fade for saving truth, *Philosophy* **85**, 109–40. Critical notice of Field (2008).

Prior, A.: 1960, The runabout inference-ticket, *Analysis* **21**(2), 38–9.

Prior, A.: 1964, Conjunction and contonktion revisited, *Analysis* **24**, 191–5.

Putnam, H.: 1980, Models and reality, *Journal of Symbolic Logic* **45**, 464–482. Reprinted in Putnam, 1983, pp. 1–25.

Putnam, H.: 1981, *Realism, Truth and History*, Cambridge: Cambridge University Press.

Putnam, H.: 1983, *Realism and Reason. Philosophical Papers*, Vol. 3, Cambridge University Press, Cambridge.

Quine, W. V. O.: 1951, Two dogmas of empiricism, *The Philosophical Review* **60**.

Quine, W. V. O.: 1970, *Philosophy of Logic*, Harvard University Press, Harvard (Mass.).

Raatikainen, P.: 2008, On rules of inference and the meanings of logical constants, *Analysis* **68**(4), 282–87.

Rabinowicz, W. and Segerberg, K.: 1994, Actual truth, possible knowledge, *Topoi* **13**, 101–15.

Read, S.: 2000, Harmony and autonomy in classical logic, *Journal of Philosophical Logic* **29**, 123–54.

Read, S.: 2008, Harmony and necessity, *in* L. K. C. Dégremont and H. Rückert (eds), *On dialogues, logics and other strange things*, King's College Publications, pp. 285–303.

Restall, G.: 2004, Laws of non-contradiction, laws of the excluded middle and logics, *in* G. Priest, J. Beall and B. Armour-Garb (eds), *The Law of Non-Contradiction. New Philosophical Essays*, Oxford University Press, Oxford, pp. 73–85.

Restall, G.: 2005, Multiple conclusions, *in* L. V.-V. Petr Hajek and D. Westerthal (eds), *Logic, Methodology and the Philosophy of Science: Proceedings of the Twelfth International Congress*, King's College Publications, London, pp. 189–205.

Restall, G.: 2008, Proof-theory and meaning: on second-order logic, *in* M. Pelis (ed.), *Logica 2007 Yearbook*, Filosofia, pp. 157–70.

Rossberg, M.: 2006, *Second-order Logic: Ontological and Epistemological Problems*, PhD thesis, University of St Andrews.

Rückert, H.: 2004, A solution to Fitch's paradox of knowability, *in* Gabbay, Rahman, Symons and V. Bendegem (eds), *Logic, Epistemology and the Unity of Science*, Kluwer Academic Publishers, Dordrecht, pp. 351–80.

Rumfitt, I.: 1997, The categoricity problem and truth-value gaps, *Analysis* **57**(4), 223–36.

Rumfitt, I.: 2000, "Yes" and "No", *Mind* **109**, 781–824.

Rumfitt, I.: 2002, Uniteralism disarmed: a reply to Dummett and Gibbard, *Mind* **111**, 305–21.

Rumfitt, I.: 2003, Savoir faire, *The Journal of Philosophy* **100**(3), 158–166.

Rumfitt, I.: 2008, Knowledge by deduction, *Grazer Philosophische Studien* **77**, 61–84.

Salerno, J.: 2000, Revising the logic of logical revision, *Philosophical Studies* **99**, 211–27.

Schroeder-Heister, P.: 1984, A natural extension of natural deduction, *Journal of Symbolic Logic* **49**, 1284–1299.

Schroeder-Heister, P.: 2007, Generalized definitional reflection and the inversion principle, *Logica Universalis* **1**(2), 355–76.

Sgaravatti, D. and Zardini, E.: 2008, Knowing how to establish intellectualism, *Grazer Philosophische Studien* **77**(1), 217–261.

Shapiro, S.: 1991, *Foundations without Foundationalism: A Case for Second-Order Logic*, Oxford University Press, Oxford.

Shapiro, S.: 1993, Anti-realism and modality, *in* J. Czermak (ed.), *Philosophy of Mathematics: Proceedings of the 15th International Wittgenstein Symposium*, Verlag Hölder-Pichler-Tempsky, Vienna, pp. 269–287.

Shapiro, S.: 1998, Induction and indefinite extensibility: The gödel sentence is true, but did someone change the subject?, *Mind* **107**(427), 597–624.

Shapiro, S. and Wright, C.: 2006, All things indefinitely extensible, *in* A. Rayo and G. Uzquiano (eds), *Absolute Generality*, Oxford University Press, Oxford, pp. 255–304.

Shieh, S.: 1998, Undecidability in anti-realism, *Philosophia Mathematica* **6**, 324–333.

Shoesmith, D. J. and Smiley, T.: 1978, *Multiple-conclusion logic*, Cambridge University Press, Cambridge.

Smiley, T.: 1996, Rejection, *Analysis* **56**(1), 1–9.

Soames, S.: 2003, *Philosophical Analysis in the Twentieth Century: The Age of Meaning*, Princeton University Press.

Stanley, J. and Williamson, T.: 2001, Knowing how, *The Journal of Philosophy* **98**(8), 411–444.

Steinberger, F.: 2009a, *Harmony and logical inferentialism*, PhD thesis, University of Cambridge, Cambridge.

Steinberger, F.: 2009b, Not so stable, *Analysis* pp. 665–61.

Tennant, N.: 1981, Is this a proof i see before me?, *Analysis* **41**, 115–0.

Tennant, N.: 1984, Were those disproof i saw before me?, *Analysis* **44**(3), 97–105.

Tennant, N.: 1992, *Autologic*, Einburgh University Press, Edinburgh.

Tennant, N.: 1997, *The Taming of the True*, Oxford University Press, Oxford.

Tennant, N.: 1999, Negation, absurdity and contrariety, *in* D. M. Gabbay and H. Wansing (eds), *What is Negation?*, Kluwer Academic Publishers, Dordrecht, pp. 199–222.

Tennant, N.: 2002, Victor vanquished, *Analysis* **62**, 135–42.

Tennant, N.: 2005a, Deflationism and the Gödel phenomena: Reply to Ketland, *Mind* **114**(453), 89–96.

Tennant, N.: 2005b, Rule-circularity and the justification of deduction, *Philosophical Quarterly* **55**(221), 625–48.

Tennant, N.: 2010, Harmony in a sequent setting. Forthcoming in *Analysis*.

Tennant, N.: forthcoming, Inferentialism, logicism, harmony, and a counterpoint, *in* A. Miller (ed.), *Essays for Crispin Wright: Logic, Language and Mathematics*, Oxford University Press, Oxford.

Thomason, R.: 1970, Indeterminist time and truth-value gaps, *Theoria* **36**, 23–42.

Troelstra, A. S. and Van Dalen, D.: 1988, *Constructivism in Mathematics, Vol. 1*, North-Holland, Amsterdam.

von Plato, J.: 2001, Natural deduction with general elimination rules, *Archives of mathematical logic* **40**, 541–67.

von Plato, J.: 2008, Gentzen's proof of normalization for natural deduction, *Bulletin of Symbolic Logic* **14**, 240–57.

Wagner, S.: 1981, Tonk, *Notre Dame Journal of Formal Logic* **4**, 289–300.

Wansing, H.: 2000, The idea of a proof-theoretic semantics and the meaning of the logical operations, *Studia Logica* **64**, 3–20.

Wansing, H.: 2002, Diamonds are philosopher's best friends, *Journal of Philosophical Logic* **31**(6), 591–612.

Weir, A.: 1986, Classical harmony, *Notre Dame Journal of Formal Logic* **27**(4), 459–82.

Williamson, T.: 1982, Intuitionism disproved?, *Analysis* **42**, 203–7.

Williamson, T.: 1987, On the paradox of knowability, *Mind* **96**, 256–61.

Williamson, T.: 1988, Knowability and constructivism, *The Philosophical Quarterly* **38**, 422–32.

Williamson, T.: 1992, On intuitionistic modal epistemic logic, *Journal of Philosophical Logic* **31**, 591–612.

Williamson, T.: 1994, Never say never, *Topoi* **13**, 135–45.

Williamson, T.: 2000, *Knowledge and its Limits*, Oxford University Press, Oxford.

Williamson, T.: 2003, Blind reasoning: understanding and inference, *Proceedings of the Aristotelian Society* **77**(1), 249–293.

Williamson, T.: 2006, Conceptual truth, *Proceedings of the Aristotelian Society* **80**, 1–41.

Williamson, T.: 2008, *The Philosophy of Philosophy*, Oxford University Press, Oxford.

Williamson, T.: 2009a, Reference, inference and the semantics of pejoratives, *The Philosophy of David Kaplan*, Oxford University Press, Oxford, pp. 137–158.

Williamson, T.: 2009b, Tennant's troubles, *in* J. Salerno (ed.), *New Essays on the Knowability Paradox*, Oxford: Oxford University Press.

Wright, C.: 1982, Strict finitism, *Synthese* **51**(2), 203–82.

Wright, C.: 1987, *Realism, Meaning and Truth*, Blackwell, Oxford.

Wright, C.: 1992, *Truth and Objectivity*, Harvard University Press, Cambridge, (Mass.).

Wright, C.: 1993, *Realism, Meaning and Truth (2^{nd} ed.)*, Harvard University Press, Harvard (Mass.).

Wright, C.: 2000, Truth as sort of epistemic: Putnam's peregrinations, *The Journal of Philosophy* **97**(6), 335–64.

Wright, C.: 2001, On being in a quandary, *Mind* **110**, 45–98.

Wright, C.: 2003a, *Saving the Differences*, Harvard (Mass.): Harvard University Press.

Wright, C.: 2003b, Vagueness: a fifth column approach, *in* J. Beall (ed.), *Liars and Heaps*, Oxford University Press, Oxford, pp. 84–105.

Wright, C.: 2004a, Intuition, entitlment and the epistemology of logical laws, *Dialectica* **58**(1), 155–75.

Wright, C.: 2004b, Warrant for nothing (and foundations for free)?, *Proceedings of the Aristotelian Society Supplementary Volume* **78**(1), 167–212.

Wright, C.: 2007a, On quantifying into predicate position: steps towards a new(tralist) perspective, *in* M. Leng, A. Paseau and M. Potter (eds), *Mathematical Knowledge*, Oxford University Press, pp. 150–71.

Wright, C.: 2007b, Wang's paradox, *in* R. E. Auxier and L. E. Hahn (eds), *The Philosophy of Michael Dummett*, Open Court, Chicago, pp. 415–444.

Wright, C.: 2009, Inferentialism, revision, harmony and (intuitionist) negation. Unpublished manuscript.