

# Rules, Logic, and Meaning

Pedro del Valle-Inclán

Scuola Normale Superiore of Pisa

In partial fulfillment of the requirements for the degree of Doctor  
of Philosophy

# Acknowledgements:

I'm lucky to have a lot of people to thank, so please settle in. I am grateful to my supervisors, Mario Piazza and Bogdan Dicher, for their help and support when I was working on this thesis. Luca Incurvati, Nils Kürbis, Leonardo Ceragioli, Camillo Fiore, Bruno Jacinto, Piet Royle, Kat T. Bough, Julien Murzi, Brett Topey and Ryan Simonelli read early versions of some of this material; thanks to all of them for their comments and their time.

I am also very grateful to Julian J. Schloeder. They were a collaborator (and a guide) in some of the work on bilateralism that made it into Chapter 4, and gave me plenty of feedback, advice, and encouragement during the PhD. Thank you for all of it.

Bas Kortenbach (who must be livid that I haven't mentioned him yet) read this thesis several times, in the form of stand-alone papers and when it all came together. He gave me thorough comments each time, and frankly, at this point all remaining mistakes are his fault. In any case, thank you, Bas; this thesis would have been much worse without your help.

Thanks also to the friends who have spent time with me during these years, in Italy and back home. Some of them have let me know that they'd prefer to be mentioned by name. They are, in alphabetical order, Ricardo de Balbín, Maria Sole Barbieri, Alicia Crespo, Sergio Diez, Lois Kalb, Bas Kortenbach, Carlos Martínez, Guillaume Massas, Ángel Pulido, Miguel Rodríguez, Thomas Sachen and Pablo Urbita. I slightly resent having had to type your names, but I will look back on our time together, and am looking forward to seeing you again.

I'm also grateful to my brother Martín, and to my mum and dad. This is going to sound cliched, but my parents actually *are* wonderfully loving and supportive, and I'm very lucky to be their son. *Gracias papá, gracias mamá.*

Finally, I am grateful to Mónica. She has helped me in every way I can think of and, if I had to guess, in several others that I haven't realised. This is just shorthand for many other things I hope to say over the years, but thank you for everything.

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Overview . . . . .	2
1.2	Thesis Outline . . . . .	8
1.3	List of Included Papers . . . . .	8
<b>2</b>	<b>Inferentialism: a primer</b>	<b>10</b>
2.1	Strict and Moderate Inferentialism . . . . .	10
2.2	More Inferentialisms . . . . .	14
2.3	Why Inferentialism? . . . . .	17
<b>3</b>	<b>The Problem of Harmony</b>	<b>21</b>
3.1	What is the Problem of Harmony? . . . . .	21
3.2	Some Criteria of Harmony . . . . .	24
3.2.1	Harmony and Conservativeness . . . . .	25
3.2.2	Harmony and Inversion . . . . .	29
3.2.3	Harmony and Strength . . . . .	38
3.3	Taking stock: Harmony and Classical Logic . . . . .	47
<b>4</b>	<b>Harmony and Bilateralism</b>	<b>48</b>
4.1	Assertion, Rejection and Negation . . . . .	49
4.2	Bilateralism, Classical Logic, and Harmony . . . . .	51
4.3	Harmony at last . . . . .	59
4.4	Harmony and Normalisation: $\mathcal{HB}_1$ . . . . .	61
4.4.1	Operational Reductions . . . . .	63
4.4.2	Reducing (Re) to Atoms . . . . .	64
4.4.3	Reducing (SR) to Atoms . . . . .	64
4.4.4	Ancillary Reductions . . . . .	66
4.4.5	Normalisation and Corollaries . . . . .	67
4.5	Harmony and Normalisation: $\mathcal{HB}_2$ . . . . .	70
4.5.1	Reducing Conclusions to Atoms in (Ex) . . . . .	71
4.5.2	Reducing Assumptions to Atoms in (Bem) . . . . .	72
4.5.3	Reducing Conclusions to Atoms in (Bem) . . . . .	72
4.5.4	Ancillary reductions . . . . .	73
4.5.5	Normalisation and corollaries . . . . .	74

Appendix I . . . . .	78
<b>5 Logic, Translation, and Minimalism</b>	<b>79</b>
5.1 Logic and Translation . . . . .	80
5.2 Logic and Use . . . . .	90
5.3 Minimalism (and some bureaucracy) . . . . .	93
5.4 Minimalism and Logics . . . . .	95
5.5 Minimalism and Rules . . . . .	101
5.6 Another Look at Structural Rules . . . . .	107
5.7 A final look at Meaning-Variance . . . . .	110
Appendix II . . . . .	116
<b>6 Rules, Models, and Meanings</b>	<b>120</b>
6.1 What is Carnap's Problem? . . . . .	121
6.2 Garson: Local and Global Models . . . . .	122
6.2.1 Local Models . . . . .	123
6.2.2 Global Models . . . . .	128
6.2.3 But what about the Quantifiers? . . . . .	136
6.3 Smiley: Bilateral Constraints . . . . .	141
6.4 McGee: Open-Endedness . . . . .	147
6.5 Bonnay and Westerståhl: Semantic Constraints . . . . .	154
6.5.1 Carnap's Problem and Definability . . . . .	158
6.5.2 Carnap's Problem and Compositionality . . . . .	162
6.6 Another Approach to Carnap's Problem . . . . .	169
6.7 Carnap's Problem and Interpretations . . . . .	175
6.8 Coda: Harmony Again . . . . .	178
<b>7 Concluding Remarks</b>	<b>182</b>
<b>Bibliography</b>	<b>183</b>

# Chapter 1

## Introduction

So you want to write a fugue?  
You've got the urge to write a fugue  
You've got the nerve to write a fugue  
So go ahead and write a fugue

---

Glenn Gould, *So You Want to  
Write a Fugue*

### 1.1 Overview

Let's start with a platitude. Words and sentences are conventional signs: they mean what they do because speakers use them (or have used them) in a certain way. There is nothing wrong with the idea, as far as it goes, but it doesn't go far. 'Use determines meaning', understood this way, says too little to be useful.

Inferentialism is one way to flesh out the slogan. It claims that the meaning of linguistic expressions is determined by the way they are used *in inferences*. If we want to know what a sentence means, an inferentialist would say, we should look at what it follows from, and to what follows from it. Likewise, if we want to know what a word means, we should look at how it contributes to the inferential connections of the sentences where it occurs.

The idea, to some extent, goes back to Frege's early work. In Section 3 of the *Begriffsschrift* he explains the notion of 'conceptual content' as follows:

There are two ways in which the content of two judgements may differ; it may, or it may not, be the case that all inferences that can be drawn from the first judgement when combined with certain other ones can always also be drawn from the second when combined with the same other judgements. The two propositions 'the Greeks defeated the Persians at Plataea' and 'the Persians were defeated by the Greeks at Plataea' differ in the former way; even if a slight

difference of sense is discernible, the agreement in sense is preponderant. Now I call that part of the content that is the same in both the conceptual content. (Frege 2002, p. 12)

Two statements have the same conceptual content if and only if we can substitute one for the other without turning good arguments into bad ones. Put another way, the content of a statement can be pinned down by the role it plays in inferences.

Another source of inspiration comes from Gentzen’s work in proof-theory. In the 1930’s Gentzen set out to provide “a formalism that reflects as accurately as possible the actual logical reasoning involved in mathematical proofs” (Gentzen 1969, p. 74). To achieve this he devised his famous natural deduction calculi, where logical operators are governed by introduction and elimination rules. Introduction rules specify conditions under which formulas containing a certain operator can be deduced. Gentzen’s introduction rule for conjunction, for example, was:

$$\frac{A \quad B}{A \wedge B} (\wedge I)$$

Elimination rules specify what may be deduced from formulas containing a given operator. Gentzen’s elimination rules for conjunction were:

$$\frac{A \wedge B}{A} (\wedge E_1) \quad \frac{A \wedge B}{B} (\wedge E_2)$$

After presenting his calculi, Gentzen remarks that:

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. [...] Note that in saying this we need not go into the ‘informal sense’ of the symbol. (Gentzen 1969, pp. 80-81)

We can put put Gentzen’s emphasis on the importance of introduction rules aside for now. What matters, at this stage, is the claim that inference rules don’t just describe the behaviour of connectives. In laying down certain rules, he says, we are *defining* the symbols concerned; we are specifying what they mean.

These ideas were further developed, under Wittgenstein’s influence, by philosophers like Carnap (1937), Sellars (1953, 1954), and Dummett (1991). On the more technical side, Prawitz’s work (e.g. 1974, 1987) has also spurred an effort to develop inferentialist semantics for logical formalisms, that is, semantics centred around the notion of proof rather than truth.<sup>1</sup>

<sup>1</sup>This is known as proof-theoretic semantics, see (Schroeder-Heister 2023) for an overview.

Perhaps the best way to see what inferentialists are getting at is to look at what they say about good inferences. Consider a toy example:

$$\frac{\text{Frida is a cat}}{\text{Frida is an animal}}$$

A traditional way to explain what makes this a good argument goes roughly as follows: ‘Frida’ denotes a certain cat, ‘cat’ denotes the set of cats, and ‘animal’ the set of animals. A sentence of the form ‘ $x$  is a  $P$ ’ is true if and only if the denotation of ‘ $x$ ’ is an element of the denotation of ‘ $P$ ’, and the set of cats is (necessarily) a subset of the set of animals. Therefore, if it is true that Frida is a cat, then it must be true that Frida is an animal.

The details matter less than the order of explanation. We start from claims about the meaning of ‘animal’ and ‘cat’, and account for the goodness of our inference in terms of them (and truth-preservation).<sup>2</sup> Inferentialists turn this explanation upside down. For them, the fact that speakers take inferences from ‘ $x$  is a cat’ to ‘ $x$  is an animal’ to be correct is part of what makes ‘cat’ and ‘animal’ mean what they do. They start with a primitive notion of good inference, understood in terms of what is accepted by a community or an individual speaker, and use it to explain our talk of meaning (and sometimes talk of truth as well).

There is a subtlety to keep in mind. Strictly speaking, the traditional and inferentialist accounts, as I have presented them, aren’t really inverses of each other. The traditional account starts from claims about meaning and gets to a notion of good inference, where a good inference is not just one that is accepted by a community or an individual. The inferentialist account, on the other hand, starts from a notion of good inference understood in terms of what speakers or communities accept, and explains meaning in terms of it. It may be tempting to say that the traditional account goes from meaning to validity, whereas the inferentialist account goes from validity to meaning, but this would erase an important difference: that between inferences that are ‘good’ in the sense of being accepted, on the one hand, and inferences that are ‘good’ in some speaker-independent sense, on the other. A better way to put the point, then, is to say that inferentialists go from claims about inferential use to claims about the meaning of expressions.<sup>3</sup>

There are many different ways to be an inferentialist. A moderate option is to say that meaning is *determined by* inferential use —or as it is sometimes put, that the meaning of an expression should be ‘read off’ the way it’s used in inferences. Moderate inferentialists may say that the meaning of ‘and’ is a

---

<sup>2</sup>I am identifying the meaning of terms with their reference for the sake of simplicity, but the same order of explanation is at play in more sophisticated accounts where meanings are identified with e.g. functions from ‘possible worlds’ to extensions.

<sup>3</sup>Of course, this raises the question of how inferentialists can carve out a notion of ‘good inference’ that goes beyond being conventionally accepted. This is related to the notion of proof-theoretic harmony, and will loom large in Chapters 3 and 4.

certain truth-function, for example, but they will also insist that this so because speakers endorse certain inferences involving the word ‘and’. According to strict versions of inferentialism, in contrast, the meaning of an expression just *is* the role it plays in inferences. The thought goes roughly as follows: being a bishop or a rook, in the game of chess, consists in being moved according to certain rules. Similarly, being a conjunction or a disjunction consists in being used a certain way in inferences (in the case of conjunction, for example, according to Gentzen’s rules above). In this vein, a strict inferentialist may say that all there is to know about the meaning of ‘and’ is that one can draw inferences from ‘*A*’ and ‘*B*’ to ‘*A* and *B*’ (and vice versa).

These ideas are relatively popular when applied to logical vocabulary. Words like ‘or’, ‘some’ and ‘not’ play a central role in arguments, and conversely, the way we use those words in arguments seems especially telling about what they mean. For this reason, connectives, quantifiers and the like are seen as a promising starting point for inferentialism.

This thesis examines three central debates in inferentialist approaches to logical vocabulary. These debates could be treated independently of each other, but they are interrelated: stances in each of them exclude stances in the others, and answers to some of them can pave the way for answers to the rest. The hope is that by treating them together we can get a fuller picture of the strengths and weaknesses of inferentialism as a whole.

The first debate concerns an intuitive distinction between ‘good’ and ‘bad’ rules. We have already seen that, according to inferentialists, some rules of inference define the expressions that they govern. A question that immediately arises is whether *any* set of inference rules can define a logical operator. In his (1960) Arthur Prior gave a compelling argument against this view. Consider a connective ‘tonk’ governed by the following rules:

$$\frac{A}{A \text{ tonk } B} \text{ (tonk I)} \qquad \frac{A \text{ tonk } B}{B} \text{ (tonk E)}$$

By chaining applications of (tonk I) and (tonk E) we can infer two arbitrary sentences from each other. If these rules define a connective, Prior says, then we have to conclude that any sentence can be inferred from any other. Inferentialists, on their part, have typically replied that Prior only shows that *some* sets of rules can’t be taken to define a connective. They have also proposed different criteria to pin down which sets of rules are acceptable definitions, known as criteria of *proof-theoretic harmony*. Different benchmarks of acceptability draw the line at different places, but they tend to agree that there is something wrong with the standard rules for some classical connectives. Simply put, then: inferentialism seems incompatible with the acceptance of classical logic. There has been a lot of debate about the justification of different harmony criteria, and about whether harmony can really pose a problem for classical logicians. I will refer to this, somewhat loosely, as the Problem of Harmony.

The second debate concerns the relation between different logics. Philosophers often disagree about which arguments are logically valid, and it has sometimes been claimed that the parties involved in this type of disagreement are talking past each other. Quine (1986), for instance, famously held that logics which validate different arguments use different logical vocabulary. Take the conflict between classical and intuitionistic logicians. The classical logician, on Quine's view, holds that all sentences of the form ' $P$  or<sub>c</sub> not<sub>c</sub>  $P$ ' are logically true. The intuitionist replies that some sentences of the form ' $P$  or<sub>i</sub> not<sub>i</sub>  $P$ ' are not true, let alone logically so. It's not that they disagree about some sentences involving the fixed particles 'or' and 'not'; rather, they mean different things by those two words. Quine's point is easy to reformulate in inferentialist terms. According to strict inferentialists the meaning of logical vocabulary is given by the rules governing its use. Since different logics are characterised by different sets of inference rules, the worry is that changing the rules one takes as valid amounts to changing one's language.

*Pace* Quine, partisans of rival logics tend see each other as endorsing different cannons of inference for the same language. Intuitionists (some intuitionists, at least) claim that the Law of Excluded Middle doesn't always hold, not that classical logicians should adopt a new disjunction and negation. Likewise, classical logicians tend to claim that intuitionists are wrong about disjunction and negation, not that they are speaking 'intuitionese'. Quine's approach seems not to make justice to this picture, and it also threatens to make substantial disagreement about logic impossible. I will call this the Meaning-Variance Problem.

The third (and final) debate concerns the proper formulation of moderate inferentialism. So far I have been quite vague about this kind of position; I have only said that, according to moderate inferentialists, the meaning of logical vocabulary should be read off the role it plays in inference. What exactly is a role in inference, though? And how does one read a semantics off of it? As it turns out, answering these questions is a complicated task. A major hurdle is the non-categoricity of logical vocabulary. In his *Formalization of Logic* (1943) Carnap showed that there are non-normal interpretations of classical logic, that is, deviant interpretations of the connectives and quantifiers that are consistent with the classical consequence relation. This spells trouble for moderate inferentialists. It follows from Carnap's results that on many natural ways of understanding 'meaning', 'inferential roles', and 'reading off', the usual meaning of logical vocabulary is not pinned down by the role it plays in inference. I will call this Carnap's Problem.

If we squint right, we can see these three problems as instances of wider philosophical debates. A popular way of explaining the force of certain arguments is to say that they hold good in virtue of the meaning of expressions that occur in them. In the case of logically valid arguments, we are often told that they hold good in virtue of the meaning of logical vocabulary. Inferentialists think of rules as definitions, so the whole point of harmony criteria, that is, the whole point of trying to determine which sets of rules 'really' define logical connectives, is to salvage the link between meaning and validity in the face of

tonk-like cases.

The Meaning-Variance Problem shows up also outside the philosophy of logic: whether differences between theories entail differences in the meaning of the words in which they are couched has been a popular topic in philosophy of science since Duhem (1954), Kuhn (1962) and Feyerabend (1962). And finally, the existence of non-normal interpretations that are consistent with the way we use logical expressions (i.e. Carnap's Problem) can be seen as a special case of the so-called Indeterminacy of Reference (Davidson 1979; Quine 1960). Although I will only consider how these problems relate to logic and the prospects of inferentialist theories about it, I hope that the discussion can be of some use, as a series of case studies, to readers who are interested in the more general debates.

It just remains to sketch the line this thesis takes with respect to all three problems, after which I will give an outline of the chapters that follow.

Regarding the Problem of Harmony, I will argue that there is nothing to prevent inferentialists from endorsing classical logic. Admittedly, the usual formalisations of classical logic are disharmonious. However, classical logic can also be given bilateral formulations —see (Rumfitt 2000; Smiley 1996). I will discuss how to extend criteria of harmony to bilateral settings, and present two harmonious bilateral formalisations of classical logic. Both systems, incidentally, are proof-theoretically well-behaved: they have the usual normalisation, separation, and subformula properties.

A common answer to the Meaning-Variance Problem, going back to Putnam (1957) and Morton (1973), is to claim that sides who agree on enough logical principles use the same logical vocabulary. The obvious difficulty with this idea is spelling out what counts as 'enough' agreement. Recently, some inferentialists have tried to do just that (Dicher 2016; Paoli 2003, 2014; Restall 2002, 2014). They are often called *minimalists*, following Hjortland's (2014) terminology. Minimalists draw a boundary roughly along the line between the operational and structural rules of sequent calculi: operational rules confer meaning, structural rules don't. It follows that disagreements about logic that can be seen as disagreements about structural rules need not involve a change of language.

In this thesis I will argue that minimalism, despite its initial plausibility, is a bad idea; it erases the distinction between connectives that we ordinarily take to be different, and identifies connectives that shouldn't be identified. I will also argue that the importance of the 'change of language' issue has been exaggerated. In many cases, I think, there is no fact of the matter as to whether different logics share their stock of operators, but there can still be substantial disagreement between partisans of each of them.

As for the third problem, I will argue that there are simple, natural ways of construing 'reading off' and 'roles in inference' that allow us to read off the usual interpretation of (classical) first-order languages from standard proof-systems, as well as from the classical consequence relation. In other words, I will argue that in some sense Carnap's Problem can be 'solved'.

## 1.2 Thesis Outline

- **Chapter 2** presents the general background behind this thesis. It goes through the main tenets of inferentialism, some prominent ways of sharpening them, and the motivation behind different forms of inferentialism.
- **Chapter 3** introduces the Problem of Harmony. After explaining the problem in detail, it examines the best-known criteria of proof-theoretic harmony and sets the stage for Chapter 4.
- **Chapter 4** discusses bilateralism and its importance for the harmony debate. It argues that the usual criteria of proof-theoretic harmony need to be adjusted in bilateral settings, proposes one way of doing so, and presents two novel, harmonious calculi for classical logic. It also proves normalisation theorems for both calculi, and shows that normal derivations have the separation and subformula properties.
- **Chapter 5** discusses the Meaning-Variance Problem. It begins with a discussion of Quine’s arguments for the Meaning-Variance claim. After this, it goes on to argue against recent minimalist proposals, and concludes with a reassessment of the importance of meaning-variance claims.
- **Chapter 6** tackles Carnap’s Problem. It discusses the most important proposals to ‘solve’ it, and goes on to put forward a novel way to get around the issue. The chapter concludes with a coda that looks back at the Problem of Harmony from a moderate inferentialist point of view.

## 1.3 List of Included Papers

A number of published or under review papers form the basis of some chapters of this thesis. These are:

- del Valle-Inclán, Pedro, Schlöder, Julian J. (2023). Coordination and Harmony in Bilateral Logic. *Mind* 132 (525):192-207.
- del Valle-Inclán, Pedro, (2023). Harmony and Normalisation in Bilateral Logic. *Bulletin of the Section of Logic* 52 (3):377-409.
- del Valle-Inclán, Pedro (2024). Carnap’s Problem, Definability and Compositionality. *Journal of Philosophical Logic* 53 (5):1321-1346.
- del Valle-Inclán, Pedro (2025). Proof-Theoretic Harmony and the Strength of Rules. *Erkenntnis*.
- del Valle-Inclán, Pedro (under review) Global Validity and the Meaning of Connectives.
- del Valle-Inclán, Pedro (under review) Minimalism, Structural Rules, and Meaning-Variance.

By chapter:

- Chapter 3, especially Section 3.2.3, draws on (del Valle-Inclán 2025)
- Chapter 4 is based on (del Valle-Inclán and Schlöder 2023) and (del Valle-Inclán 2023).
- Chapter 5, especially Sections 5.3-5.7, draws on *Minimalism, Structural Rules, and Meaning-Variance*.
- Chapter 6 draws on (del Valle-Inclán 2024) and *Global Validity and the Meaning of Connectives*.

## Chapter 2

# Inferentialism: a primer

For the only way to write one,  
Is just to plunge right in and write one

---

Glenn Gould, *So You Want to  
Write a Fugue*

In this chapter I will give an overview of inferentialism. Inferentialists, as we have seen, want to account for meaning in terms of use in inferences. That is a large project, and there are different ways to go about it. I have already mentioned two, strict and moderate inferentialism; they are discussed in more detail in the next section. Although the strict/moderate distinction is important, there are other useful ways to sort out types of inferentialism, ways that cut across this divide; I will present them in Section 2. And finally, in Section 3 I will address the elephant in the room: why be an inferentialist at all? This section goes through the usual motivations for different types of inferentialism.

### 2.1 Strict and Moderate Inferentialism

It is customary to draw a distinction between semantics and meta-semantics. Semantics tries to specify the meaning of the expressions of some language. Meta-semantics tries to clarify what makes it the case that linguistic expressions mean what they do. The claim that ‘cavallo’ means horse in Italian, for example, belongs to semantics; the claim that ‘cavallo’ means horse because Italian speakers have certain communicative intentions when they utter it, on the other hand, belongs to meta-semantics.<sup>1</sup>

Moderate inferentialism is a meta-semantic thesis, according to which:

---

<sup>1</sup>As usual the distinction is easier to exemplify than to fully explain, and there is room for disagreement about where semantics ends and meta-semantics begins. Still, a rough and ready understanding of the distinction is enough here. Incidentally, what I am calling semantics and meta-semantics have also been called ‘descriptive’ and ‘foundational’ semantics, respectively (see Stalnaker 1997, García-Carpintero 2012).

**(Inf)** The meaning of linguistic expressions is determined by their use in inference.

Strict inferentialism is a semantic thesis, namely:

**(Inf)<sup>+</sup>** The meaning of linguistic expressions consists in their use in inference.

One important difference between them is that moderate inferentialism is neutral as to how to characterise meanings. The most common semantic theories identify the meaning of sentences with the conditions under which they are true, and explain meaning of subsentential expressions in terms of (broadly) referential relations to non-linguistic objects. Moderate inferentialism is compatible with this approach; it just adds that reference and truth-conditions are fixed by roles in inference. At the same time, moderate inferentialism is also compatible with less orthodox approaches to semantics, where meaning is explained in terms of e.g. assertibility or verifiability conditions (Dummett 1991, Prawitz 1987). Strict inferentialism, in contrast, comes with its own way of talking about meaning. According to strict inferentialists meaning is a roundabout way of talking about inferential use, and talk of meaning boils down to talk of (appropriate) use in inferences.

Let's look at each option a bit closer, beginning with moderate inferentialism. What does it mean to say that the way a logical expression is used in inferences determines reference or truth-conditions? One way to explain this is to look at how inferentialists think semantics relates to language use. Traditionally, semantics associates certain objects ('semantic values') with sentential and subsentential expressions in a systematic way. For inferentialists the only point of doing so is as part of an attempt to explain the linguistic behaviour of speakers; semantic values are posits to explain patterns of use:

It is possible to associate many sorts of things with sentences and other linguistic expressions. What makes the association a *semantic* one is precisely the possibility of appealing to it to explain the proprieties that govern the use of those expressions. Calling what one associates with expressions 'contents', 'propositions', 'sets of possible worlds', 'truth conditions', 'extensions', or 'referents' is at best issuing a promissory note that hints at how what are put forward as their semantic correlates ought to be taken to be relevant to determining how those expressions are correctly used. [...] Semantics answers to pragmatics, attributions of content to explanations of use. (Brandom 1994, pp. 187-8)

This broadly pragmatist approach is popular beyond the ranks of inferentialism. On Davidson's program, for example, sentences are assigned semantic values so as to make sense of the interpretee's verbal behaviour as a rational, intentional activity, and accounting for the use of language in this way is the only criterion of correctness for semantics (Davidson 1973). Quine (1960) also

denied that there could be objective reasons to choose between translation manuals that predict identical patterns of use, and more generally still, mainstream work in natural language semantics often works, implicitly at least, under a similar assumption. Here is Barwise on what he calls ‘thin semantics’:

We have intuitions about the logical behavior of a certain class of sentences. [For example] with attitude reports these are largely intuitions about the phenomenon of “opacity”: reluctance to substitute co-referential terms and the like. We codify these intuitions in a set of logical principles, and then semantics consists of finding a collection of plausible set-theoretic models that makes the logical principles come out correct. I think this is the traditional conception in semantics, and it is the setting for Montague Grammar. (Barwise and Perry 1985, p. 148)

This way of looking at things can help build a bridge between inferential use and truth-conditions. Suppose that we have some account that spells out how the truth of a statement is connected to its use.<sup>2</sup> Then an inferentialist may argue that an assignment of semantic values that ‘does its job’, that is, helps explain proprieties of linguistic use, should make certain basic arguments truth-preserving. For example, it is part of the inferential role of ‘and’, part of how that word is used, that arguments of the form:

$$A, B \vdash A \wedge B$$

are to be accepted as correct. In order for arguments of this form to be truth-preserving, it must be the case that whenever  $A$  is true and  $B$  is true, so is  $A \wedge B$ . Similarly, it is part of the inferential role of ‘and’ that arguments of the forms:

$$A \wedge B \vdash A \quad A \wedge B \vdash B$$

ought to be accepted. For them to be truth-preserving, it must be the case that whenever  $A \wedge B$  is true so are  $A$  and  $B$ . Putting everything together, we get the standard truth-condition for conjunction:  $A \wedge B$  is true if and only if  $A$  is true and  $B$  is true; the semantic value we have read off the role conjunctions play in inference is the usual truth-function. As we will see in Chapter 6, matters are more complicated with the rest of propositional connectives, and much more complicated when we add quantifiers to the mix; we leave further details until then.

What about strict inferentialism? According to strict inferentialists, as we know, meaning is not something distinct from, but somehow associated with, linguistic expressions. Sentences are inferentially articulated—that is, certain transitions from (sets of) sentences to sentences are sanctioned as appropriate,

---

<sup>2</sup>There are many ways to do this. Lewis (1975), for instance, speaks of a convention of truthfulness between speakers, and Weiner (2005) of truth as a standard of assertibility.

and others as inappropriate— and the meaning of a sentence consists in its specific pattern of inferential connections.<sup>3</sup> Sellars puts the point as follows:

To say that ‘rot’ means *red*’ is not to describe ‘rot’ as standing “in the meaning relation” to an entity *red*; it is to use a recognized device (the semantical language game) for bringing home to a *user* of ‘red’ how Germans use ‘rot’. (Sellars 1954, p. 213)

In other words, to state the meaning of an expression is just to state how it functions in its language, and that is all there is to meaning:

It would be a mistake to [claim] that the semantical statement “‘es regnet’ means *it is raining*” gives information about the German use of ‘Es regnet’ which would *supplement* a description of the role it plays in the German language game, making a *complete* description of what could otherwise be a partial account of the properties and relations of ‘Es regnet’ as a meaningful German word. [...] It conveys no information which could not be formulated in terms of the pieces, positions, moves, and transitions [...] of the German language game. (Sellars 1954, p. 213)

Now, how does one go about specifying the inferential role of different expressions? The most popular way to do so is in terms of inference rules, rules that make explicit the allowed moves and transitions between different statements. This is what logicians do (for very restricted classes of expressions) when they set up formal deductive systems. And indeed, strict inferentialists typically hold that the inferential role of logical expressions can be described using Gentzen-style rules familiar from proof-theory. For example, the inferential role of classical disjunction can arguably be spelled out by the following rules:<sup>4</sup>

$$\frac{A}{A \vee B} (\vee I_1) \quad \frac{B}{A \vee B} (\vee I_2) \quad \frac{A \vee B \quad \frac{\Gamma_1, [A]^1 \quad \mathcal{D}_1}{C} \quad \frac{\Gamma_2, [B]^1 \quad \mathcal{D}_2}{C}}{A} (\vee E)^1$$

This approach can be extended to non-logical vocabulary. A popular strategy is to adopt Dummett’s ‘two-aspect’ model of the meaning of sentences (Dummett 1973). The first aspect, Dummett says, consists in the conditions under which a sentence can be appropriately asserted. The second aspect comprises the consequences of the assertion of a sentence, including what can be

<sup>3</sup>Note that it is only *sentences* that are inferentially articulated (we don’t draw inferences between e.g. individual words). Inferentialist semantics starts by explaining the content of sentences, and the content of subsentential expressions is then recovered as the uniform contribution they make to the content of the sentences where they occur.

<sup>4</sup>The usual notational conventions apply: the  $\mathcal{D}_i$  stand for sub-derivations, brackets indicate the discharge of hypotheses, and both vacuous and multiple discharges are allowed.

deduced from it. The thought is that we can characterise the first aspect of meaning by something like introduction rules, and the second aspect by means of something like eliminations. Plausibly, the two aspects of ‘ $x$  is a bachelor’ could be (partially) characterised by rules like:

$$\frac{x \text{ is an unmarried man}}{x \text{ is a bachelor}} \qquad \frac{x \text{ is a bachelor}}{x \text{ is a man}} \qquad \frac{x \text{ is a bachelor}}{x \text{ is unmarried}}$$

In order to flesh out examples like this, and develop a fully fledged theory of meaning, one needs to make theoretical and methodological choices that go beyond the distinction between strict and moderate inferentialism. The most important options are laid out in the next section.

## 2.2 More Inferentialisms

Inferentialists disagree about much more than the strict/moderate distinction. A first, important source of conflict is the intended *target* of the position, that is, the sort of language inferentialism is supposed to be a theory of.

Some inferentialists, like Peacocke (2004) and Boghossian (2000), are interested in the idiolects of individual speakers. Others, like Peregrin (2014) or Dummett (1991), focus on shared, public languages. Each option comes with its own take on what we should zoom in on. If we are interested in idiolects, we should look at the inferential practices of single speakers. If we are interested in public languages, we should look at patterns of use shared across communities of speakers.

Idiolect-inferentialists tend to have a descriptive notion of use in mind when they talk about the inferential use of an expression. Peacocke (2004), for example, takes meaning to be determined by the inferences speakers find ‘primitively obvious’, and Boghossian (2000) spells out use in terms of individual dispositions to infer. Inferentialists who focus on public languages, in contrast, favour normative notions of use. When they talk about the use of an expression they mean its correct use according to some community-wide standard. More importantly, their talk of meaning usually involves normative vocabulary. Brandom, for instance, talks about implicit rules governing commitment and entitlement to statements, as well as relations of incompatibility between them. Speakers are ‘players in the game of giving and asking for reasons’, and keep score of each others commitments and entitlements. The meaning of a sentence is then explained as the overall difference its utterance makes to the commitment-and-entitlement score of a player (see Brandom 1994, Ch. 3).

Inferentialists also disagree about the *scope* of inferentialism, that is, about which expressions inferentialism applies to. Sometimes inferentialism is proposed as an analysis of restricted classes of expressions, like legal or scientific terms; this is known as *local* inferentialism. Other times inferentialism is put

forward as an all-encompassing theory of meaning, applicable to complete languages; this is known as *global* inferentialism. The chapters that follow only deal with inferentialism applied to logical vocabulary, which calls for a quick clarification. When speaking of logical vocabulary, I will not take for granted a sharp, principled boundary between logical and non-logical expressions. There is a long standing debate about what sets logical vocabulary apart.<sup>5</sup> As far as this thesis is concerned, though, we can also see the logical/non-logical distinction as contingent, pragmatic, and somewhat arbitrary (Gómez-Torrente 1989, 2002; Warmbröd 1999).

Finally, a third important distinction between inferentialists concerns the *range* of inferences which are considered relevant to meaning. On one side of the divide we have authors like Peregrin (2014) and Brandom (1994), who claim that every inferential link plays a role in carving out the meaning of each and all expressions. This is often thought to lead to semantic holism, the view that the meaning of any given sentence (and derivatively, of any subsentential expression) can't be disentangled from the meaning of the rest. In Dummett's words:

On such a view, it is illegitimate to ask for the content of any single statement [...]; the significance of each statement [...] is modified by the multiple connections which it has, direct and remote, with other statements in other areas of our language taken as a whole [...]. Rather, even this image is false to the facts: it is not that a statement has, as it were, a primal meaning which then gets modified by the interconnections that are established with other statements [...]. Its meaning simply consists in the place which it occupies in the complicated network which constitutes the totality of our linguistic practices. (Dummett 1975, p. 218)

On the other side of the divide we have inferentialists who think that only part of the inferential use of an expression is relevant for its meaning. Here is Boghossian on conditionals:

The thought is that there is a particular set of inferences involving 'if, then' that are meaning-constituting for a thinker: of all the inferences that 'if, then' can and does participate in, a specific subset is responsible for fixing its meaning. (Boghossian 2000, pp. 248-9)

Inferentialists of this sort typically allow for the possibility of clusters of expressions whose meanings depend on one another. What they reject is that there is an all-encompassing semantic cluster, language as a whole.

The debate is reproduced, at smaller scale, in the case of logical vocabulary. Some inferentialists (e.g. Tennant 1987, 1997) insist that the meanings of logical expressions should be 'separable', that is, that it should be possible

---

<sup>5</sup>See e.g. (Hacking 1979) (Došen 1989) and (Brandom 1994) for different takes from an inferentialist perspective, and (Bonnay 2014) for an overview of how the issue is approached in model-theoretic terms.

to explain the meaning of each logical operator independently of the rest. The most popular way to go about this is to use Gentzen-style introduction and elimination rules which govern single operators, like the rules for conjunction and disjunction we saw earlier. Other inferentialists (e.g. Milne 2002) allow for clusters of logical expressions with interdependent meanings. Correspondingly, they consider rules in which more than one logical constant occurs, such as:

$$\frac{[A]^1 \quad \mathcal{D} \quad B}{\neg A \vee B} 1 \qquad \frac{[B]^1 \quad \mathcal{D} \quad A}{A \vee \neg B} 1$$

A third option is to say that the operators of a logic have to be accounted for *en bloc*. One way to do this is to lay down rules that establish a cyclical dependence between operators. For example, given operators  $o_1, \dots, o_n$ , the rules for  $o_1$  may feature  $o_2$ , whose rules in turn feature  $o_3$  (and so on), until we get to the rules for  $o_n$ , which feature  $o_1$ . Alternatively, we could take meanings to be given by consequence relations on a language, rather than by rules, as in (Bonney and Westerståhl 2016).<sup>6</sup> This thesis looks at positions of each of these kinds.

Choices about the scope and range of inferentialism also affect the notion of inference at play. We have already seen that, in order to extend the rule-based approach to non-logical vocabulary, inferentialists consider material rules of inference (i.e. rules that are not *logically* valid) such as:

$$\frac{x \text{ is a bachelor}}{x \text{ is unmarried}}$$

In fact, global inferentialists think of inferences in an even broader way. An inference, as we normally think of it, is a ‘language-to-language’ move: it takes us from a set of sentences (the premises) to another sentence (the conclusion). This is sometimes seen as a problem for inferentialists who want to account for predicates that are closely tied to observation, like colour terms. The worry is that, by only looking at how linguistic expressions relate to each other, we might end up with a picture of language as something disconnected from the world. Take the sentence ‘This is red’. It certainly stands in inferential relations to other sentences; for instance, it is entailed by ‘This is vermilion’, and it entails ‘This is coloured’. As relevant as these connections may be to its meaning, though, there seems to be something missing from the story: the fact that ‘red’ correctly applies to all and only red things.

<sup>6</sup>We can see this last option as a way to ‘quotient over’ sets of rules that yield the same derivability relation; the idea is that sets of rules which give rise to the same derivability relation determine the same meanings.

Cases like this lead most global inferentialists to allow for what Sellars (1954) called language-entry and language-exit rules. These are inferences (in a loose sense of the word) that take us ‘from world to language’ and ‘from language to world’. The meaning of ‘red’, for example, could be partly given by a language-entry rule to the effect that it is appropriate to assert ‘This is red’ while pointing to a red thing, and a language-exit rule stating that it is appropriate to fetch the red object upon being told ‘Fetch the red thing’.<sup>7</sup> Brandom sums up the idea as follows:

Inferential relations between *noninferential* circumstances of appropriate application and *noninferential* appropriate consequences of application are also taken into account. [...] Thus the visible presence of red things warrants the applicability of the concept red —not as the conclusion of an *inference*, but *observationally*. And the point is that the connection between those circumstances of application and whatever consequences of application the concept may have can be understood to be *inferential* in a broad sense, even when the items connected are not themselves sentential. (Brandom 2007, p. 658)

While this is important in its own right, it will not play a role in what follows. Here I am concerned only with logical vocabulary, and according to inferentialist orthodoxy it can be accounted for in terms of inferences in the usual ‘language-to-language’ sense.

## 2.3 Why Inferentialism?

Now that we have a clearer idea of what inferentialism looks like, it is important to ask: why be an inferentialist at all? Naturally, different types of inferentialism can be argued for on different grounds, so let’s start with the main concern of this thesis, strict and moderate inferentialism applied to logical vocabulary.

Moderate inferentialists hold that the meaning of logical operators is determined by the way they are used in inferences. The idea that use determines meaning is widespread in contemporary philosophy of language, but why inferential use in particular? A good way to motivate this choice is to compare logical operators with other types of expressions. The meaning of proper names, like ‘Bob’, or mass nouns, like ‘water’, can perhaps be learnt by something like ostension. If one points to sufficiently many bodies of water while saying ‘water’, an interlocutor with enough good will and good luck may find out what we mean.<sup>8</sup> Logical constants, however, seem to call for a different type of explanation; after all, it is difficult to imagine what pointing to an instance of ‘and’ may look like.

---

<sup>7</sup>That being said, some inferentialists insist that we can make do with language-to-language inferences alone, see (Simonelli 2023).

<sup>8</sup>I am more than oversimplifying here; see e.g. (Quine 1960) for a classic discussion of the problems that arise from this picture.

Plausibly, we pick up the meaning of logical operators by seeing how they are used in sentences. A very prominent aspect of that use consists in producing and evaluating arguments, so the idea that we can build a semantic theory by concentrating on this feature of use has some initial appeal. Moreover, as inferentialists are fond of pointing out, alternative meta-semantic accounts seem to not get off the ground. Take informational semantics, which tries to root meaning-talk in lawlike connections between mental or linguistic items, on the one hand, and the external objects and properties that cause them on the other (Fodor 1990, Dretske 2000). This program seems straightforward in the case of simple nouns and predicates, like ‘horse’ and ‘red’. But it is difficult to make sense of the idea of a causal relation involving, say, truth-functions —or any other sort of abstract mathematical entity, for that matter (Steinberger and Murzi 2017, Adams and Aizawa 2021).

What about strict inferentialism? One advantage of the position, compared to more standard approaches to semantics, is that it is ‘ontologically neutral’:

Traditionally, semantics has been denotational and representational, consisting in a homomorphic valuation from expressions to some range of objects. This approach risks ontological explosion, first in hypostatizing denotations for empty names, predicates, conjunctions, prepositions and so on [...], then in seeking values for false propositions in the form of non-actual states of affairs. It is also regressive, since a criterion is now needed to determine which non-actual states of affairs are possible, which simply repeats the initial problem. Talk of possible worlds is an attractive metaphor, but does little useful philosophical work and much harm. (Read 2010, p. 558)

Strict inferentialism does without the usual denotational apparatus; an expression is contentful in virtue of being inferentially connected to others, and its meaning is not something separate and distinct from it. This sort of approach seems especially fitting for logical constants. The idea that expressions like ‘or’ and ‘not’ have referents, like proper names, is somewhat counter-intuitive and comparatively recent; traditionally, logical expressions were taken to be syncategorematic.<sup>9</sup>

So much so for inferentialism about logic. What about global forms of inferentialism? Clearly one may grant that the meaning of logical expressions is best approached in terms of their inferential use, but it isn’t obvious how this scales up to whole languages. The first thing to keep in mind is that global inferentialists tend to construe the notion of inference very broadly. We have seen that even observation reports like ‘This is red’ count, on Brandom’s construal, as broadly inferential; restricting attention to inferential use therefore turns out not to be much of a restriction. The point can be pushed further. One way

---

<sup>9</sup>That is not to say that truth-conditional semantics must always assign semantic values to logical expressions. In fact, the standard Tarskian semantics for first-order languages treats connectives and quantifiers syncategorematically. The point is just that strict inferentialism is a different way to cash out the shared intuition that logical operators do not work, semantically, like predicates and names.

to understand (some) global inferentialists is to see them as claiming that all genuinely linguistic use is inferential. According to Brandom, for instance, what distinguishes the noises made by a parrot in response to a certain stimulus, on the one hand, from genuine linguistic behaviour on the other, is that the latter is inferentially articulated:

The parrot does not treat “That’s red” as incompatible with “That’s green”, nor as following from “That’s scarlet” and entailing “That’s colored”. Insofar as the repeatable response is not, for the parrot, caught up in practical proprieties of inference and justification, and so of the making of further judgments, it is not a conceptual or a cognitive matter at all. (Brandom 2000, p. 48)

Finally, and changing tack, inferentialism is also sometimes thought to be a good stepping stone for understanding wider philosophical problems. Dummett, for instance, has claimed that inferentialism yields the best account of what linguistic understanding consists in. On a purely truth-conditional approach, to know the meaning of a sentence is to know the conditions under which it is true. Dummett finds this problematic for two reasons. The first is that knowledge of meaning should not in general be construed as explicit theoretical knowledge. The second reason is that the claim that understanding a statement consists in knowing what it is for it to be true, according to Dummett, is 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). In contrast, grasping the meaning of a sentence, on an inferentialist account, consists in accepting certain inferences involving it, where this is understood as a practical ability —a form of ‘knowing how’ rather than ‘knowing that’.

Other authors have seen inferentialism as a way to get a grip on the epistemology of logic. Details vary depending on the specific form of inferentialism involved, but the overall idea is to argue for the analyticity of logic. One strategy, inspired by Gentzen’s ideas, has it that introduction rules define the operators they involve, in that they display necessary and sufficient conditions for the assertion of certain sentences involving them, and elimination rules can be justified by their corresponding introductions (I will say more about this in the next chapter).

Boghossian (2003) proposes a related idea. He begins by asking how justification is transmitted from the premises of a deductive argument to its conclusion. According to Boghossian traditional answers fall into one of two categories: internalist answers demand that we should be in a position to know, by reflection alone, that the premises of the argument provide us with good grounds for believing its conclusion, and externalist answers demand instead that the patterns of argument we employ be necessarily truth-preserving. The problem, according to Boghossian, is that internalist approaches lead to an infinite regress, and externalist approaches are too lax: some inferences are ‘epistemically blameworthy’ *and* truth-preserving. Boghossian’s way out is to argue that accepting certain inferences is a necessary condition for grasping the meaning of logical

operators, and that inference patterns which are ‘constitutive’ of the concepts they involve can be used blindly and blamelessly (Boghossian 2003, p. 239).

It goes without saying that these accounts have faced a legion of objections, and that the rough motivation for different forms of inferentialism we just went through is far from conclusive. Then again, that was never the point. Broad semantic and meta-semantic claims, like the main inferentialist ideas, are best seen as gambits, working hypotheses to be vindicated through the construction of a satisfactory theory along the lines they propose. It is therefore time to take a look at some of the problems that stand in the way of inferentialists, and consider how one may get around them.

## Chapter 3

# The Problem of Harmony

A canon in inversion  
is a dangerous diversion

---

Glenn Gould, *So You Want to  
Write a Fugue*

In this chapter we will look at the ins and outs of the Problem of Harmony. Section 1 explains what the problem consists in, and what the different strategies to tackle it are. Section 2 goes through some examples of the most popular type of solution, namely, laying down criteria about which inference rules correctly define logical operators. Finally, Section 3 takes stock prepares the ground for Chapter 4.

### 3.1 What is the Problem of Harmony?

In his (1960) Arthur Prior set out to refute any view that combines two claims that were relatively widespread at the time.<sup>1</sup> The first is that some inferences are ‘analytically valid’, that is, valid in virtue of the meaning of the vocabulary in them:

(AV): Some inferences are valid in virtue of the meaning of expressions that occur in them.

At this point it would be natural to ask what validity (let alone validity ‘in virtue of meaning’) consists in. Luckily, we don’t need to get bogged down by the details in this case. For the purposes of Prior’s argument, validity can be understood in any way that entails valid inferences ought to be accepted by all speakers.

The second claim Prior is concerned with has to do with the use of inference rules as definitions:

---

<sup>1</sup>See e.g. (Kneale 1956) or (Popper 1947).

**(AR):** Any set of rules can define the meaning of an expression that it governs.

**(AV)** and **(AR)** suggest an attractive picture of the connection between use, meaning, and logical validity. Gentzen’s ( $\wedge$  I) and ( $\wedge$  E<sub>1/2</sub>) describe some aspects of our use of the word ‘and’. Suppose that, in line with **(AR)**, we take them to define the meaning of conjunction. Then by **(AV)** all inferences carried out according to these rules are valid in virtue of the meaning of conjunction. Validity springs from meaning, and meaning from linguistic use.

This picture is plausible enough if we have sensible rules in mind, but **(AR)** is a very permissive principle; it states that *any* set of inference rules can define an operator. Prior takes advantage of this, and proposes to define a connective ‘tonk’ by means of the following rules:

$$\frac{A}{A \text{ tonk } B} \text{ (tonk I)} \qquad \frac{A \text{ tonk } B}{B} \text{ (tonk E)}$$

By chaining applications of (tonk I) and (tonk E) we can infer two arbitrary sentences from each other, and according to **(AV)** those inferences will be valid in virtue of the meaning of ‘contonkion’. This is the core of the Problem of Harmony: it seems to follow from **(AV)** and **(AR)** that any sentence can be validly inferred from any other.

Prior saw his argument as a *reductio ad absurdum* of the inferentialist program, but unsurprisingly, inferentialists have resisted his conclusion. There are, broadly speaking, three ways to reply to Prior’s challenge. The first is to hold on to both **(AV)** and **(AR)**, and try find some fault with Prior’s argument. The second is to reject **(AV)**, and the third is to reject **(AR)** instead.

The first strategy has recently been defended by Jared Warren (2015, 2020). Warren asks us to imagine a community of speakers whose inferential practice was adequately described, at least in part, by the rules (tonk I) and (tonk E). Suppose their language, Tonkish, was just like English except for the addition of the word ‘tonk’, and suppose they accepted inferences like:

**(1)** Two plus two is four, therefore two plus two is four tonk there are six spoons in my pocket, therefore there are six spoons in my pocket.

According to **(AV)** and **(AR)** inferences like **(1)** would be analytically valid. But the mistake, Warren says, is assuming that the Tonkish sentences occurring in **(1)** should be homophonically translated into English. The meaning of a sentence, on an inferentialist approach, is a matter of the role it plays in inference. The inferential role of the Tonkish sentences ‘two plus two is four’ and ‘there are six spoons in my pocket’, however, would be radically different from the inferential roles of their homophonic English counterparts. Therefore, Warren says, **(AV)** and **(AR)** don’t force us to conclude that the claim that

there are six spoons in my pocket can be (correctly) inferred from the claim that two plus two is four. It only follows that two different claims —expressed in Tonklish by deceptively familiar-looking sentences— analytically follow from each other.<sup>2</sup>

The second strategy, rejecting **(AV)**, has also received some attention. At first sight it may seem that, if one rejects **(AV)**, the second premise **(AR)** has to go with it. As Boghossian has pointed out, **(AV)** and **(AR)** appear to be connected by principles like:

**(P)**: If someone's being disposed to make the inference from  $S_1$  to  $S_2$  is constitutive of what  $S_2$  means in that person's idiolect, then the inference from  $S_1$  to  $S_2$  is analytic in that person's idiolect — i.e., valid by virtue of the meanings of  $S_1$  and  $S_2$  alone. (Boghossian 1994, p. 111)<sup>3</sup>

Inferentialists who reject **(AV)** insist that **(P)**, and other principles along its lines, are false. A popular way to explain why is to say that, although meaning is a matter of inferential role, validity depends on truth-preservation:

To say that the inference from 'red' to 'not blue' is constitutive of the meaning of 'red', is to say that someone must be prepared to make that inference if he is to mean *red* by 'red'. If we wish, we may put this by saying that such a person must be prepared to *regard* the inference from 'red' to 'not blue' as valid. But this doesn't *by itself* imply that the inference is valid, let alone that it is valid by meaning alone. Being constitutive of meaning is one thing, being analytic [analytically valid] another. (Boghossian 1994, p. 120)

Stephen Read takes a similar line in his (2015b):

What is good about the notion of proof-theoretic validity is that it recognizes that what rules one adopts determines the meaning of the logical terms involved and commits one to accepting certain inferences as valid. What is bad is to infer from this that those inferences really are valid. Proof-theoretic validity [...] cannot serve the metaphysical function of actually making those inferences valid. Validity is truth preservation, and proof must respect that fact. (Read 2015b, p. 156)

Robert Brandom goes about this second strategy in a different way. Borrowing Dummett's two-aspect model of meaning (see Section 2.1), Brandom holds that using any concept involves a tacit commitment to the 'inference' from

---

<sup>2</sup>Of course, this raises the question of what Tonklish sentences could possibly mean, and what a Tonklish-to-English translation might look like. For details see (Warren 2015, 2020).

<sup>3</sup>Boghossian formulates **(P)** in terms of idiolects and individual dispositions to infer, but the same point can be made in terms more palatable to 'public-language inferentialists' (see Chapter 2, section 2.2).

its grounds of application to its consequences of application (Brandom 1994, p. 126). But crucially, he says, those tacit commitments can be challenged and rejected:

One can refuse to employ a concept, on the grounds that it embodies an inference one does not endorse. [...] Critical thinkers, or merely fastidious ones, must examine their idioms to be sure that they are prepared to endorse and so defend the appropriateness of the material inferential transitions implicit in the concepts they employ. (Brandom 1994, p. 126)

The idea is that, although some inferences partly constitute the meaning of the expressions they govern —as per **(AR)**— this doesn't entail that they are valid, at least in any sense of 'valid' which compels us to endorse them. This leaves open the question of which inferences ought to be endorsed. Boghossian and Read have a short answer: (necessarily) truth-preserving inferences. This answer is not available to Brandom, though, since he takes talk of truth to be dependent on talk of appropriate inferences. His position

starts with a practical distinction between good and bad inferences, understood as a distinction between [conventionally] appropriate and inappropriate doings, and goes on to understand talk about truth as talk about what is preserved by the good moves. (Brandom 2000, p. 12)

Brandom's answer, to the extent that it is an answer, is that there is no one-size-fits-all way to decide which inferences to stand up for, and so, which concepts to use:

This is not the sort of a question to which one ought to expect or even welcome a general or wholesale answer. Grooming our concepts and material inferential commitments in the light of our assertional commitments [...] and the latter in the light of the former is a messy, retail business. (Brandom 1994, pp. 129-30)

Although the first two ways of responding to Prior's challenge, exemplified by Warren, Boghossian, Read, and Brandom, are viable and interesting in their own right, the most popular line of defence has been the third one, abandoning **(AR)** while holding on to **(AV)**. The idea behind this third approach is that only *some* sets of rules will do as definitions. This comes with a burden: explaining which inference rules successfully define logical operators, and why. There are many, often intricate ways to answer this question, and we will look at some of them in the next section.

## 3.2 Some Criteria of Harmony

The moral inferentialists tend to draw from Prior's paper is straightforward: not all sets of rules can be taken to adequately define an expression, and we

should look for criteria to determine which do. Following (Dummett 1975), these different benchmarks are known as criteria of *proof-theoretic harmony*.

Over the years many (non-equivalent) criteria have been proposed, and I will not attempt to exhaustively review them here. Instead, I will focus on the three best-known proposals, explain their motivation and examine their consequences for classical logic. This will be enough to give a clear idea of the state of the harmony debate, and set the stage for what's to come.

### 3.2.1 Harmony and Conservativeness

One of the earliest replies to Prior's challenge is Nuel Belnap's (1962). According to Belnap:

The key to a solution lies in observing that [...] we are not defining our connectives *ab initio*, but rather in terms of an *antecedently given context of deducibility*, concerning which we have some definite notions. (Belnap 1962, p. 131)

This antecedently given context of deducibility, Belnap says, consists of assumptions about the properties of the relation of logical consequence. And what the rules for tonk show is that:

By a too careless use of definitions, it is possible to create a situation in which we are forced to say things inconsistent with those assumptions. (Belnap 1962, p. 131)

Let's make this more precise. Belnap's context of deducibility can be spelled out in terms of structural rules. Consider, for example, the usual rules of Identity, Weakening and Cut:

$$\frac{}{A \vdash A} \text{ (Id)} \quad \frac{\Gamma \vdash C}{\Gamma, A \vdash C} \text{ (W)} \quad \frac{\Gamma \vdash A \quad \Delta, A \Rightarrow C}{\Gamma, \Delta \vdash C} \text{ (Cut)}$$

Belnap's suggests that:

We may take this little system as expressing *all and only* the universally valid statements and rules expressible in the given notation: it completely determines the context. (Belnap 1962, p. 132, my emphasis)

Now suppose that we extend this system with the tonk rules, which in the new notation read:

$$A \vdash A \text{ tonk } B \quad A \text{ tonk } B \vdash B$$

Using (Cut) in combination with the tonk rules we can derive  $p \vdash q$  for all atomic sentences  $p, q$ . This wasn't possible in the original system, which we had assumed to already encode *all* valid deducibility statements not containing any connectives. The tonk rules, then, clash with our antecedent assumption and must be rejected. The idea can be generalised using the notion of a weak conservative extension:

**(Weak) Conservative Extension:** Let  $\mathcal{L}$  be a language and  $\mathbf{D}$  a calculus for  $\mathcal{L}$  that contains only structural rules. Let  $\mathbf{c}$  be a connective not in  $\mathcal{L}$ , and say  $\mathcal{L}^+$  is the result of extending  $\mathcal{L}$  with  $\mathbf{c}$ , and  $\mathbf{D}^+$  the result of closing  $\mathbf{D}$  under the rules for  $\mathbf{c}$ . Finally, let  $\vdash_x$  denote deducibility in a system  $x$ . We say that  $\mathbf{D}^+$  is a weak conservative extension of  $\mathbf{D}$  if, for any  $\Gamma \cup \{A\} \subseteq \mathcal{L}$ , we have that:

$$\Gamma \vdash_{\mathbf{D}^+} A \text{ iff } \Gamma \vdash_{\mathbf{D}} A.$$

According to Belnap an operator  $\mathbf{c}$  is harmonious (relative to a calculus  $\mathbf{D}$ ), if and only if the rules for  $\mathbf{c}$  are weakly conservative with respect to  $\mathbf{D}$ . I will call this B-harmony for short. Once again:

The justification for unpacking the demand for consistency [with the context of deducibility] in terms of conservativeness is precisely our antecedent assumption that we already had *all* the universally valid deducibility-statements not involving any special connectives. (Belnap 1962, p. 132)<sup>4</sup>

A pleasant feature of Belnap's criterion is that it is (relatively) 'framework-independent'. The notion of a weak conservative extension abstracts away from the specifics of different proof-systems, and works equally well in sequent and natural deduction settings; as we will see below, this doesn't hold for many other harmony criteria. At the same time, Belnap's criterion of harmony is 'context-dependent': an operator is B-harmonious only with respect to some calculus. In the usual terminology, Belnap's is a *global* harmony criterion. Some have seen this as a problem (e.g. Steinberger 2011a), arguing that harmony criteria should be *local*, that is, apply to operators regardless of the context in which they are used. We will see some examples of this type of harmony criterion shortly.

Putting this aside, though, it is easy to check that all standard calculi for classical logic are B-harmonious. The reason conservativeness is often thought a problem for classical logic is not Belnap's proposal, but a natural generalisation of it due to Dummett. In his (1991) Dummett argues that the rules governing a connective should form a conservative extension with respect to the entire calculus they belong to, not just with respect to its structural fragment. We will call this a strong conservative extension:

---

<sup>4</sup>Strictly speaking Belnap's notion of harmony also includes a requirement of uniqueness (roughly, a requirement that connectives governed by the same rules should be inter-substitutable). He attaches relatively little importance to uniqueness, though, and it won't play a part in what follows, so I will omit it here. For more on this see (Belnap 1962) and (Došen and Schroeder-Heister 1985).

**(Strong) Conservative Extension:** Let  $\mathcal{L}$  be a language and  $\mathbf{D}$  a calculus for  $\mathcal{L}$ . Let  $\mathbf{c}$  be a connective not in  $\mathcal{L}$ , and say  $\mathcal{L}^+$  is the result of extending  $\mathcal{L}$  with  $\mathbf{c}$ , and  $\mathbf{D}^+$  the result of closing  $\mathbf{D}$  under the rules for  $\mathbf{c}$ . Finally, let  $\vdash_x$  denote deducibility in a system  $x$ . We say  $\mathbf{D}^+$  is a strong conservative extension of  $\mathbf{D}$  when, for any  $\Gamma \cup \{A\} \subseteq \mathcal{L}$ , we have that:

$$\Gamma \vdash_{\mathbf{D}^+} A \text{ iff } \Gamma \vdash_{\mathbf{D}} A.$$

According to Dummett, a connective  $\mathbf{c}$  is harmonious (relative to a calculus  $\mathbf{D}$ ) if and only if the rules for  $\mathbf{c}$  are strongly conservative with respect to  $\mathbf{D}$ . I will call this D-harmony for short.

Contrary to B-harmony, this new criterion flags many standard calculi for classical logic as defective. Often the problem is that the rules for classical negation are non-conservative. Consider, for example, a calculus with Gentzen's rules for  $\wedge$  and  $\vee$  plus the usual rules for the conditional:

$$\frac{\Gamma, [A]^1 \quad \mathcal{D} \quad B}{A \rightarrow B} (\rightarrow \text{I})^1 \qquad \frac{A \rightarrow B \quad A}{B} (\rightarrow \text{E})$$

In this system Peirce's Law  $((A \rightarrow B) \rightarrow A) \rightarrow A$  cannot be proved. If we add any set of rules for classical negation, say:

$$\frac{[A]^1 \quad \mathcal{D}_1 \quad \neg A}{\neg A} (\neg \text{I})^1 \qquad \frac{A \quad \neg A}{B} (\neg \text{E}_1) \qquad \frac{\neg \neg A}{A} (\neg \text{E}_2)$$

then Peirce's Law becomes derivable. The rules for classical negation, therefore, are not D-harmonious with respect to our background calculus.

Important as this observation might be, we should also keep in mind that strong conservativeness doesn't rule out classical logic *tout court*. It is well known that plenty of multiple-conclusion calculi for classical logic, both in sequent and natural deduction style, contain only strongly conservative rules (Borićić 1985; Read 2000; Gentzen 1969). Dummett, and other inferentialists with him, have argued that multiple-conclusion rules cannot be used to define the meaning of logical operators (Dummett 1991; Steinberger 2011b; Tennant 1997). Even if we grant their point, however, there remain classical, single-conclusion calculi in which all rules are conservative. This is typically achieved by incorporating Peirce's Law in the form of a rule (Gordeev 1987; E. Zimmermann 2002), but that is not the only option. The  $\{\exists, \wedge, \neg\}$  signature is

expressively complete for classical first-order logic, and almost any standard calculus for this signature will come out D-harmonious.<sup>5</sup>

Finally, let's turn to one last important question: why should we require strong conservativeness? The usual motivation goes something as follows. Suppose that, in line with **(AV)**, logically valid arguments are valid in virtue of the meaning of logical vocabulary. Suppose, in addition, that the meaning of each logical operator should be specified by *its* rules, regardless of the meaning of the others. Then a logically valid argument from some premises  $\Gamma$  to a conclusion  $A$  should only appeal to the rules for operators that occur in  $A$  or  $\Gamma$ , since they completely determine the meaning of the relevant vocabulary; the demand for strong conservativeness arises naturally.<sup>6</sup>

This argument gives rise to questions of its own, most notably: why should we assume that the meaning of each operator must be specified independently of the meaning of the rest? This is just another formulation of the requirement of separability from Section 2.2. And the problem, as pointed out by Milne (2002) and Steinberger (2011a), is that arguments in favour of separability are few and far between. Besides some appeals to plausibility or intuition, the principle is usually taken for granted, rather than argued for.

Dummett's (1991) is a clear example. Although Dummett is not a holist, he grants that there are clusters of expressions whose meanings depend on one another:

To understand a sentence of a given language, one must know some fragment of that language. [...] The principle of compositionality allows that a range of coordinate expressions [...] may be capable of being understood only *together*. (Dummett 1991, pp. 222-23)

This leaves open the possibility that the meanings of logical constants need not, in general, be given independently of each other. But the possibility is brushed off when we are told that:

The logical constants form a uniquely simple case: [...] to understand ' $A$  or  $B$ ', one need not understand ' $A$  and  $B$ ' or 'if  $A$  then  $B$ '. (Dummett 1991, p. 223)

To make matters more complicated, when considering whether we should specify the meaning of logical vocabulary using only 'pure' rules, that is, rules containing a single operator, Dummett says:

Reflection shows this demand is exorbitant. An impure **c**-introduction rule will make the understanding of **c** depend on the prior understanding of the other logical constants figuring in the rule. Certainly we don't want this dependence to be cyclic; but there would

---

<sup>5</sup>For a simple example in natural deduction, take the  $\wedge$ ,  $\exists$ , and  $\neg$  rules from (Prawitz 1965, Ch. 1). It is easy to show that, for any  $\mathbf{c} \in \{\exists, \wedge, \neg\}$ ,  $\mathbf{c}$  can be conservatively added to the  $\mathbf{c}$ -less fragment of the calculus.

<sup>6</sup>In fact there are two important alternatives here: one could claim that the meaning of each logical term is given by its operational rules, or by its operational rules *and* the structural rules of their calculus. We will examine these alternatives in detail in Chapter 5.

be nothing in principle objectionable if we could so order the logical constants that the understanding of each depended only on the understanding of those preceding it in the order. (Dummett 1991, p. 257)

But this seems to run counter to the separability requirement, and so against the demand for strong conservativeness.

Let's take stock. There are two harmony criteria built around conservativeness, Belnap's weak and Dummett's strong conservativeness requirements. According to Belnap's criterion the usual formalisations of classical logic are harmonious, while according to Dummett's some (though not all) standard calculi for classical logic are defective. Moreover, Dummett's criterion rests on the assumption that logical vocabulary should be separable, and it is unclear what the grounds for this assumption are.

### 3.2.2 Harmony and Inversion

The next harmony criterion, also due to Dummett, calls for some stage-setting. As we saw in Section 2.1, Dummett sorts the principles governing the use of sentences into two types:

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 1973, p. 396)

Inferentialists explain the meaning of subsentential expressions in terms of their contribution to the meaning of a representative range of sentences. In the case of logical constants, Dummett says, this range consists of sentences where they occur as the main operator. The meaning of a constant  $c$  can then be laid out by its introduction and elimination rules: I-rules represent the constant's contribution to the grounds for the assertion of the sentences where it occurs, and E-rules represent the constant's contribution to the consequences of such assertions.

Dummett's key idea is that there should be a certain balance between the two aspects of the use of sentences, and therefore between introduction and elimination rules. He describes this balance as follows:

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 premise, to derive from the premises of the

introduction rule a consequence that we could not otherwise have drawn. (Dummett 1991, pp. 247-8)

Let's say that a *local c-peak* is a formula occurrence (in a derivation) that is simultaneously the consequence of a **c**-Introduction rule and the major premise of a **c**-Elimination rule. Then according to Dummett's second harmony criterion a connective **c** is harmonious if and only if local **c**-peaks are eliminable. I will call this I-harmony for short.<sup>7</sup>

Typically, one shows that local peaks can be eliminated using the sort of reductions that make up the bulk of a normalisation proof. Suppose, for example, that we have a local  $\vee$ -peak which is a conclusion of  $(\vee I_1)$  and major premise of  $(\vee E)$ :

$$\frac{\frac{\Gamma_0}{\mathcal{D}_0} \quad \frac{A}{A \vee B} (\vee I_1)}{A \vee B} \quad \frac{\Gamma_1, [A]^1 \quad \mathcal{D}_1}{C} \quad \frac{\Gamma_2, [B]^1 \quad \mathcal{D}_2}{C} (\vee E)^1}{C}$$

Clearly we can also derive  $C$  without the  $\vee$ -peak  $A \vee B$ ; we just need to rearrange a few sub-derivations:

$$\frac{\Gamma_0, \Gamma_1}{\mathcal{D}_0} \quad \frac{A}{\mathcal{D}_1} \quad C$$

Local  $\vee$ -peaks that are consequences of  $(\vee I_2)$  and major premises of  $(\vee E)$  are eliminated using a similar strategy:

$$\frac{\frac{\Gamma_0}{\mathcal{D}_0} \quad \frac{B}{A \vee B} (\vee I_2)}{A \vee B} \quad \frac{\Gamma_1, [A]^1 \quad \mathcal{D}_1}{C} \quad \frac{\Gamma_2, [B]^1 \quad \mathcal{D}_2}{C} (\vee E)^1}{C} \quad \rightsquigarrow \quad \frac{\Gamma_0, \Gamma_2}{\mathcal{D}_0} \quad \frac{B}{\mathcal{D}_2} \quad C$$

And  $\wedge$ -peaks can be eliminated too:

---

<sup>7</sup>The 'I' stands for 'Inversion', for reasons we will see shortly.

$$\frac{\frac{\Gamma_0 \quad \Gamma_1}{\mathcal{D}_0 \quad \mathcal{D}_1} \frac{A \quad B}{A \wedge B} (\wedge I)}{\frac{A \wedge B}{A} (\wedge E_1)} \rightsquigarrow \frac{\Gamma_0}{\mathcal{D}_0} \frac{B}{A}$$

$$\frac{\frac{\Gamma_0 \quad \Gamma_1}{\mathcal{D}_0 \quad \mathcal{D}_1} \frac{A \quad B}{A \wedge B} (\wedge I)}{\frac{A \wedge B}{B} (\wedge E_2)} \rightsquigarrow \frac{\Gamma_1}{\mathcal{D}_1} \frac{B}{B}$$

These procedures are available because the elimination rules for  $\wedge$  and  $\vee$  just ‘undo’ or ‘invert’ what is done by their respective introductions. Prawitz puts the point as follows:

An elimination rule is, in a sense, the inverse of the corresponding introduction rule: by an application of an elimination rule one essentially only restores what had already been established if the major premise of the application was inferred by an application of an introduction rule. (Prawitz 1965, p. 33)

Naturally, not all pairs of (sets of) introduction and elimination rules are like this. The rules for tonk make this clear:

$$\frac{A}{A \text{ tonk } B} (\text{tonk I}) \qquad \frac{A \text{ tonk } B}{B} (\text{tonk E})$$

On the face of it, the inverse of (tonk I) is not (tonk E), but the rather harmless rule:

$$\frac{A \text{ tonk } B}{A} (\text{tonk E})^*$$

Peaks formed by an application of (tonk I) followed by an application of (tonk E)\* are easy to eliminate, but peaks formed by (tonk I) and (tonk E) can’t be reduced, which makes the tonk rules disharmonious:

$$\frac{\frac{\Gamma_1}{\mathcal{D}_1} \frac{A}{A \text{ tonk } B} (\text{tonk I})}{\frac{A \text{ tonk } B}{B} (\text{tonk E})} \rightsquigarrow \text{?}$$

Perhaps more surprisingly, the usual sets of rules for classical negation are also disharmonious in this new sense. Given the rules for classical negation we saw earlier, for example, there is no general procedure for eliminating  $\neg$ -peaks of the form:

$$\begin{array}{c} [\neg A]^1 \\ \mathcal{D} \\ \frac{\neg\neg A}{\neg\neg A} (\neg I)^1 \\ \frac{\neg\neg A}{A} (\neg E_2) \end{array}$$

Before we compare this new criterion of harmony to the ones we have seen so far, it will be useful to consider another important aspect of I-harmony. According to Dummett, the possibility of eliminating local peaks is tied to the more ambitious project of finding ‘proof-theoretic justifications’ of some logical laws. We already know that Dummett distinguishes between aspects of the use of sentences that relate the grounds for an assertion, on the one hand, and aspects that have to do with the consequence of an assertion, on the other. Building on this idea he distinguishes between two types of theory of meaning. Pragmatist theories take the consequences of the assertion of a sentence to determine its meaning, while verificationist theories take grounds for assertion as meaning-determining instead.

Now, a connective’s introduction rules spell out some grounds on which we may assert a sentence where it is the main operator.  $(\wedge I)$ , for example, licenses the assertion of  $A \wedge B$  on the grounds of the two sentences  $A$  and  $B$ . Clearly, though,  $(\wedge I)$  doesn’t exhaust all conditions for appropriately asserting  $A \wedge B$ ; for instance, if we have proven  $C \rightarrow (A \wedge B)$  and  $C$  we can safely conclude  $A \wedge B$ . Dummett calls the grounds for assertion specified by introduction rules *direct* or *canonical* grounds, and suggests that verificationists should take canonical grounds to determine the meaning of the sentences they warrant.

This idea has two important consequences. First, if introduction rules are meaning-conferring stipulations, then they don’t stand in need of justification. And secondly, if canonical grounds really determine the meaning of some sentences, then the corresponding introduction rules are in a certain sense ‘complete’. This is what Dummett calls the Fundamental Assumption:

A statement may frequently be established by *indirect* means, but to label certain means ‘canonical’ is to claim that, whenever we are justified in asserting the statement, we *could have* arrived at our entitlement to do so by those restricted means. (Dummett 1991, p. 252)

On any view along these lines the validity of introduction rules is more or less unproblematic. But what about elimination rules? Following Prawitz (1973,

1974), Dummett tries to justify them by appeal to their corresponding introductions. In simple cases the justification Dummett has in mind boils down to a re-interpretation of the procedures for eliminating local peaks. Let's stick to  $\vee$  as an example, and suppose that in the following application of  $(\vee E)$  the sub-arguments  $\mathcal{D}_0, \mathcal{D}_1,$  and  $\mathcal{D}_2$  are valid and have no undischarged premises:

$$\frac{\mathcal{D}_0 \quad [A]^1 \quad \mathcal{D}_1 \quad [B]^1 \quad \mathcal{D}_2}{\frac{A \vee B \quad C \quad C}{C} (\vee E)^1}$$

The argument  $\mathcal{D}_0$  is valid by assumption, so it warrants the assertion of  $A \vee B$ . By the Fundamental Assumption, we could have arrived at such a warrant by canonical means, using  $(\vee I_1)$  or  $(\vee I_2)$ . Say, without loss of generality, that we could have arrived at it by means of  $(\vee I_2)$ :

$$\frac{\mathcal{D}_0 \quad [A]^1 \quad [B]^1}{\frac{\frac{B}{A \vee B} (\vee I_2) \quad \mathcal{D}_1 \quad \mathcal{D}_2}{C \quad C} (\vee E)^1}$$

Then the second procedure for levelling out  $\vee$ -peaks shows that we could have arrived at the same conclusion  $C$  without appealing to  $(\vee E)$ :

$$\begin{array}{c} \mathcal{D}_0 \\ B \\ \mathcal{D}_2 \\ C \end{array}$$

The upshot is that, if the Fundamental Assumption is correct, applications of  $(\vee E)$  to valid sub-arguments with no undischarged premises are 'safe': a valid argument for their conclusion is available before (and independently of) them. Dummett takes this to justify  $(\vee E)$ . More generally, he holds that:

Any elimination rule shown to be in harmony with the introduction rules is justifiable, and hence to be considered valid. (Dummett 1991, p. 253)

The fully worked out version of this idea is what Dummett calls the *upwards justification procedure* (Dummett 1991, Ch. 11). Roughly put, the upwards justification procedure is an inductive definition of validity according to which

applications of introduction rules yield valid arguments by default, and applications of elimination rules yield valid arguments if they can be justified along the lines that we just sketched. The details of the inductive definition aren't important for our purposes, but two difficulties with the resulting account will be relevant below.

The first difficulty is that Dummett's Fundamental Assumption doesn't seem to fit well with our actual use of logical vocabulary (Ripley 2017; Rumfitt 2017). Take  $(\vee I_1)$  and  $(\vee I_2)$ —yes, again. If the Fundamental Assumption is right, and if ' $\vee$ ' is to have roughly the same meaning as 'or', it follows that whenever we are justified in asserting ' $A$  or  $B$ ' we *could have* been in a position to assert  $A$  or assert  $B$ . Now we face a choice. If the 'could' in 'could have' is interpreted in a narrow sense, akin to practical possibility, then the Fundamental Assumption gets our conventions about what warrants the assertion of disjunctions wrong. Here is a colourful example due to Ripley:

You know that a certain spaceship has a self-destruct button, that the spaceship is functioning properly and will only self-destruct if the button is pressed, and that Alice and Zebra are the only potential buttonpressers aboard the ship. You see the spaceship undergo its self-destruct process, destroying (alas!) Alice, Zebra, and all records or evidence about what was on the ship before its destruction. Putting this all together, you conclude that either Alice or Zebra pressed the self-destruct button. This conclusion is surely warranted, but you have, and can have, no evidence about which one of them it was. (Ripley 2017, p. 313)

If, on the other hand, the 'could' in 'could have' is interpreted more broadly (say, as possibility in principle for an ideally placed observer), then it is difficult to see how the Fundamental Assumption relates to our actual linguistic practice. As Ripley puts it “the kinds of warrant that are directly connected to our use of language are not plausibly understood as merely possible warrants; they are the actual ones” (Ripley 2017, p. 315).<sup>8</sup>

The second difficulty with the upwards justification procedure is that it validates inferences that seem plainly incorrect. Goldfarb (2016) has pointed out that given any  $A$  and  $B$  which don't share atomic sub-sentences, and where  $A$  does not contain  $\perp$ , the inference from  $A \rightarrow B$  to  $B$  is valid according to Dummett's definition. It is therefore difficult to see how the upwards procedure carries any justificatory force.<sup>9</sup>

---

<sup>8</sup>Dummett himself grants that the status of the Fundamental Assumption is “very shaky” and that “as applied to the conditional and the universal quantifier, we have to concede that the fundamental assumption is not literally true” (Dummett 1991, pp. 277-8).

<sup>9</sup>I am oversimplifying things here. What Goldfarb shows is that, given any sensible set of rules for atomic sentences, inferences like the above will count as valid. The problem traces back to Dummett's treatment of 'atomic bases', i.e. sets of rules for atomic sentences. Other proof-theoretic definitions of validity are available. Stafford and Nascimento (2023), for example, define a notion of proof-theoretic validity that is sound and complete for intuitionistic propositional logic, although it isn't clear that their assumptions are compatible with

We now have a fuller picture of of ‘harmony as inversion’. It is time to compare it with conservativeness-based criteria.

Earlier we saw that conservativeness requirements are framework-independent, in that they apply to different types of proof-system, and global, in the sense that whether a set of rules is conservative partly depends on the calculus to which it belongs. I-harmony, in contrast, is framework-dependent and local.

The framework-dependence of I-harmony is easy to spot: it appeals to introduction and elimination rules, which are characteristic of natural deduction systems. That being said, it has an analogue in sequent settings. The procedures to eliminate local peaks roughly correspond to procedures to eliminate principal cuts.<sup>10</sup> The first procedure for eliminating  $\vee$ -peaks, for example, corresponds to the following cut-elimination step:

$$\frac{\frac{\mathcal{D}_0}{\Gamma_0 \Rightarrow A} \quad \frac{\frac{\mathcal{D}_1}{\Gamma_1, A \Rightarrow C} \quad \frac{\mathcal{D}_1}{\Gamma_1, B \Rightarrow C}}{\Gamma_1, A \vee B \Rightarrow C} \text{ (Cut)}}{\Gamma_0, \Gamma_1 \Rightarrow C} \rightsquigarrow \frac{\frac{\mathcal{D}_0}{\Gamma_0 \Rightarrow A} \quad \frac{\mathcal{D}_1}{\Gamma_1, A \Rightarrow C}}{\Gamma_0, \Gamma_1 \Rightarrow C} \text{ (Cut)}$$

When sequent rules are such that principal cuts involving them can be eliminated we will say that they are cut-inductive, and we will take cut-inductiveness as the sequent correlate of I-harmony.

The observation that I-harmony is a local criterion calls for some explanation. In some calculi peaks for a connective  $\mathbf{c}$  can sometimes be avoided using the rules for a different connective  $\mathbf{c}'$ . In a calculus with tonk and  $\vee$ , for example this may look as follows:<sup>11</sup>

$$\frac{\frac{\Gamma_1}{\mathcal{D}_1} \quad \frac{A}{A \text{ tonk } B} \text{ (tonk I)}}{B} \text{ (tonk E)} \rightsquigarrow \frac{\frac{\Gamma_1}{\mathcal{D}_1} \quad \frac{A}{A \text{ tonk } B} \text{ (tonk I)}}{(A \text{ tonk } B) \vee B} \text{ (\vee I)} \quad \frac{[A \text{ tonk } B]^1}{B} \text{ (tonk E)} \quad \frac{[B]^1}{B} \text{ (\vee E)}^1$$

I-harmony, however, is meant to capture a certain balance between the two aspects of the use of sentences, a balance that supposedly obtains regardless of the rest of the expressions in a language. Implicitly, the demand seems to

---

Dummett’s. For a different way to raise doubts about the justificatory power of Dummett’s strategy, see (Haack 1982).

<sup>10</sup>Principal cuts are applications Cut where both occurrences of the cut-formula have been obtained in the previous step of a derivation using operational rules.

<sup>11</sup>Note that the tonk-peak on the left derivation disappears, but only at the cost of introducing a  $\vee$ -peak in its place. Moreover, eliminating the new  $\vee$ -peak would only re-introduce the tonk-peak we begun with. Strictly speaking, then, we are not so much eliminating peaks as shuffling them around.

be that, given any connective  $\mathbf{c}$ , it should be possible to eliminate  $\mathbf{c}$ -peaks ‘by appeal to the rules for  $\mathbf{c}$  alone’. The simplest way to make this idea sharper is requiring that  $\mathbf{c}$ -peaks be eliminable in any calculus containing only  $\mathbf{c}$ .

This takes care of the comparison between I-harmony and conservativeness-based criteria. Let’s now turn to another important question: what is the motivation for I-harmony? One possibility is to draw on the upwards justification procedure. Eliminating local peaks, according to Dummett, doubles as a justification of the elimination rules involved, so I-harmony can be seen as a requirement that elimination rules be proof-theoretically justified. At times Dummett seems to reject this idea, however:

The Fundamental Assumption [...] will not suffice to validate all laws of classical logic by proof-theoretic means. That is not a condemnation of classical logic, since there is no a priori reason to assume that the meanings of the logical constants can be specified by any set of self-justifying laws. (Dummett 1991, p. 279)

In his (1991) Dummett first introduces I-harmony as “an analogue, within the restricted domain of logic” of the general requirement for strong conservativeness (p. 247). The extent to which I-harmony and strong conservativeness are analogues is problematic, though. As Steinberger (2013) has pointed out, the two requirements neither imply nor are implied by each other. More precisely (a) a set of rules can be strongly conservative with respect to a calculus containing only I-harmonious rules without being I-harmonious itself, and (b) adding a set of I-harmonious rules to an I-harmonious calculus need not give rise to a strongly conservative extension.

We already have the materials for an example of claim (a). The usual rules for  $\vee$  and  $\wedge$  are I-harmonious, while the rules for classical negation given in the previous section aren’t. If we add them to a calculus containing only  $\vee$  and  $\wedge$ , however, we get a strong conservative extension.

An example of claim (b) appears in (Dummett 1991). Consider an operator  $\sqcup$  governed by the same rules as disjunction, except that ( $\sqcup$  E) doesn’t allow extra assumptions in the sub-derivations of the minor premises:

$$\frac{A}{A \sqcup B} (\sqcup I_1) \quad \frac{B}{A \sqcup B} (\sqcup I_2) \quad \frac{A \sqcup B \quad \frac{\frac{\emptyset, [A]^1}{\mathcal{D}_1} \quad \frac{\emptyset, [B]^1}{\mathcal{D}_2}}{C} \quad C}{A} (\sqcup E)^1$$

These rules are I-harmonious. We can eliminate  $\sqcup$ -peaks in the same way that we eliminate  $\vee$ -peaks. A calculus containing the usual rules for  $\wedge$  and the rules for  $\sqcup$  above, therefore, contains only I-harmonious rules. In such a calculus the distributive law  $A \wedge (B \sqcup C) \vdash (A \wedge B) \sqcup (A \wedge C)$  isn’t provable. Still, when we extend it with the usual rules for  $\vee$ , which are also I-harmonious, distributivity follows:

$$\frac{\frac{A \wedge (B \sqcup C)}{B \sqcup C} \quad \frac{[B]^1}{B \vee C} \quad \frac{[C]^1}{B \vee C}}{B \vee C} 1 \quad \frac{\frac{A \wedge (B \sqcup C)}{A} \quad [B]^2}{(A \wedge B) \sqcup (A \wedge C)} \quad \frac{\frac{A \wedge (B \sqcup C)}{A} \quad [C]^2}{(A \wedge B) \sqcup (A \wedge C)} 2}{(A \wedge B) \sqcup (A \wedge C)}$$

This takes us to the last contested issue surrounding harmony-as-inversion. The idea behind I-harmony is to enforce a certain balance between introduction and elimination rules, but some have argued that it falls short of this goal (Steinberger 2011a, Tranchini 2018). Once we fix a set of introduction rules, I-harmony excludes eliminations that seem ‘too strong’ with respect to them. For instance, given what (tonk I) requires in order to prove the sentence  $A$  tonk  $B$ , the rule (tonk E) allows one to deduce ‘too much’. Still, the objection continues, I-harmony does not exclude eliminations that are ‘too weak’ with respect to their corresponding introductions. ( $\sqcup$  E) is often used as an example, but the point is clearer when we look at the following connective  $\sqcap$ :

$$\frac{A \quad B}{A \sqcap B} (\sqcap I) \quad \frac{A \sqcap B}{A} (\sqcap E)$$

Intuitively, ( $\sqcap$  E) doesn’t let us deduce ‘as much as it should’, given how  $A \sqcap B$  is proved using ( $\sqcap$  I). However, it is clear that  $\sqcap$  is I-harmonious.

Partly to address this problem Dummett sketches a converse of the upwards justification procedure, the *downwards justification procedure* (Dummett 1991, Ch. 13). This is an attempt to justify introduction rules on the basis of their corresponding eliminations, much like the upwards procedure was meant to justify eliminations on the basis of their corresponding introductions. Dummett then proposes a further requirement of stability, based on the interaction between upwards and downwards justifications:

Suppose we adopt the downwards procedure, and start with a set  $\mathcal{E}$  of elimination rules. By our procedure, we can determine which introduction rules are valid: say these form a set  $\mathcal{I}$ . Now, with respect to this set  $\mathcal{I}$  of introduction rules the upwards justification procedure is well-defined, so we can use it to determine which elimination rules are valid, according to the criteria of the upwards procedure. If we get back by this means to the set  $\mathcal{E}$ , or to some set interderivable with  $\mathcal{E}$ , in the presence of  $\mathcal{I}$ , stability prevails; otherwise not. (Dummett 1991, pp. 287-8)

The requirement of stability rules out  $\sqcup$  and  $\sqcap$ , but it is somewhat undermined by the problems that surround Dummett’s proof theoretic justifications. Besides the dubious status of the Fundamental Assumption, the upwards justification procedure, as we saw, validates some seemingly incorrect inferences.

Conversely, the downwards justification procedure invalidates some instances of (presumably) valid inference patterns, like modus ponens.<sup>12</sup>

Ultimately, Dummett fell back on conservativeness as his preferred harmony criterion (Dummett 1991, p. 290). As we will see in the next section, however, there have been other attempts to cash out the idea that introduction and elimination rules should offset one another.

### 3.2.3 Harmony and Strength

Our next harmony criterion is due to Neil Tennant. Originally presented in (Tennant 1978), it has seen successive reformulations in (Tennant 1987), (Tennant 1997), and more recently in (Tennant 2020).

Tennant’s goal is to ensure that the introduction and elimination rules for logical operators are ‘balanced’ with each other. In some sense, as we saw earlier, (tonk E) is too strong for (tonk I), and ( $\sqcap$  E) is too weak for ( $\sqcap$  I). In order to make the relevant idea of balance more precise Tennant puts forward two requirements, one about the interplay between introductions and eliminations, and another about the relative strength of rules. I will call them the requirements of *equilibrium* and *maximality*, respectively.

The requirement of equilibrium appeals to the notion of the strongest proposition with a property  $\varphi$ . By a proposition Tennant understands a class of logically equivalent formulas. If  $A$  is a formula, ‘the proposition  $A$ ’ is the class of formulas logically equivalent to  $A$ . Strength, as applied to propositions, is defined as follows (see Tennant 2020, p. 237):

**Strongest proposition with property  $\varphi$ :** The strongest proposition with property  $\varphi$  is that proposition  $A$  with property  $\varphi$  such that any proposition  $B$  with property  $\varphi$  is deducible from  $A$ .

**Weakest proposition with property  $\varphi$ :** the weakest proposition with property  $\varphi$  is that proposition  $A$  with property  $\varphi$  that can be deduced from any proposition  $B$  with property  $\varphi$ .

Clearly, these definitions take some background deducibility relation  $\vdash$  for granted. When I want to emphasise this, or when the relation in question is not clear from context, I will talk of ( $\vdash$ )propositions being ( $\vdash$ )stronger than others.

With this in place, we can now present the first half of Tennant’s harmony criterion, the requirement of equilibrium.<sup>13</sup> Let  $\mathbf{c}$  be a connective governed by ( $\mathbf{c}$  I) and ( $\mathbf{c}$  E). Then in order for  $\mathbf{c}$  to be in equilibrium, ( $\mathbf{c}$  I) and ( $\mathbf{c}$  E) must satisfy the following conditions (Tennant 2020, pp. 241-2):

<sup>12</sup>See (Prawitz 2017) for details. In the same paper Prawitz proposes a modification with which Dummett’s procedure justifies the usual intuitionistic rules, so this last problem is not unsurmountable. More recently Oliveira (2019) has put forward a variant of the downwards justification procedure such that the resulting notion of validity is complete for propositional intuitionistic logic.

<sup>13</sup>I will follow Tennant in formulating the requirement for binary connectives, but the changes needed for  $n$ -ary connectives are obvious. Here and in what follows ( $\mathbf{c}$  I) and ( $\mathbf{c}$  E) can be taken as single rules or as sets of rules.

**(S)**:  $A\mathbf{c}B$  is the strongest conclusion possible under the conditions described by **(c I)**. Moreover, in order to show this:

- (i) One needs to exploit all the conditions described by **(c I)**.
- (ii) One needs to make full use of **(c E)**.
- (iii) One may not make any use of **(c I)**.

**(W)**:  $A\mathbf{c}B$  is the weakest major premiss possible under the conditions described by **(c E)**. Moreover, in order to show this:

- (iv) One needs to exploit all the conditions described by **(c E)**.
- (v) One needs to make full use of **(c I)**.
- (vi) One may not make any use of **(c E)**.

Tennant’s formulation of **(S)** and **(W)** is not entirely clear, so it will be useful to see them in action. Let’s take  $\wedge$  as an example. In order to prove that its usual rules satisfy **(S)**, we need to show that  $A \wedge B$  is the strongest proposition that fulfils the conditions described by **( $\wedge$  I)**. Let  $C$  be an arbitrary proposition that can be inferred from premisses  $A, B$ . We can show that  $A \wedge B$  is stronger as follows:

$$\frac{\frac{A \wedge B}{A} (\wedge E_1) \quad \frac{A \wedge B}{B} (\wedge E_2)}{C}$$

Note that this derivation (i) uses the assumption that  $C$  can be inferred from  $A, B$ , (ii) makes full use of **( $\wedge$  E)** —that is, it uses ‘both halves’ of the rule— and (iii) makes no use of **( $\wedge$  I)**.

In order to prove **(W)** we need to show that  $A \wedge B$  is the weakest proposition that fulfils the conditions described by **( $\wedge$  E)**. Let  $C$  be an arbitrary proposition from which we can infer  $A, B$ . We can show that  $A \wedge B$  is weaker as follows:

$$\frac{\frac{C}{A} \quad \frac{C}{B}}{A \wedge B} (\wedge I)$$

It is also easy to check that this derivation satisfies conditions (iv)-(vi) in much the same way that the previous one satisfied conditions (i)-(iii).

There are still a couple of loose ends to tie. First, **(S)** and **(W)** appeal to the notion of the strength of a proposition, but strength with respect to which deducibility relation? Tennant doesn’t explicitly address this, but he does insist that harmony is a relation in which an operator’s rules stand to each other regardless of the rules governing the rest of the language (Tennant 1997, Ch. 10). This suggests that the intended background deducibility relation is  $\vdash_c$ , the relation induced by **(c I)** and **(c E)**. In what follows I will take this for granted, but nothing substantial hinges on it.

Secondly, the requirement of equilibrium doesn't uniquely determine the appropriate counterpart of a given set of rules. Take  $(\vee I_1)$  and  $(\vee I_2)$ . They are in equilibrium with  $(\vee E)$ , but also with its restricted version  $(\vee E_R)$  from the previous section:

$$\frac{\Gamma_1, [A]^n \quad \Gamma_2, [B]^n \quad \mathcal{D}_1 \quad \mathcal{D}_2}{\frac{A \vee B \quad C}{C} (\vee E)^n} \quad \frac{\emptyset, [A]^n \quad \emptyset, [B]^n \quad \mathcal{D}_1 \quad \mathcal{D}_2}{\frac{A \vee B \quad C}{C} (\vee E_R)^n}$$

Similarly, there can be more than one introduction rule in equilibrium with a given elimination. For example,  $(\rightarrow E)$  is in equilibrium with  $(\rightarrow I)$  and also with the restricted rule  $(\rightarrow I_R)$ , in which no vacuous discharges are allowed:

$$\frac{\Gamma, [A]^n \quad \mathcal{D}}{\frac{B}{A \rightarrow B} (\rightarrow I)^n} \quad \frac{\Gamma, [A]_{\square}^n \quad \mathcal{D}}{\frac{B}{A \rightarrow B} (\rightarrow I_R)^n}$$

Moreover, the different rules in equilibrium with a given introduction or elimination need not be variations of the same shape. The usual rule of conjunction introduction, for example, is in equilibrium with  $(\wedge E_{1/2})$ , but also with the following elimination:

$$\frac{\Gamma, [A]^n [B]^n \quad \mathcal{D}}{\frac{A \wedge B \quad C}{C} (\wedge GE)^n}$$

According to Tennant a criterion of harmony should uniquely determine the counterpart of any given introduction or elimination rules. In order to ensure this, he supplements the requirement of equilibrium with the requirement of maximality. This requirement consists of two conditions (Tennant 2020, p. 239):

**(M1):** Given  $(c E)$ , we determine  $(c I)$  as the strongest introduction rule such that  $(c I)$  and  $(c E)$  are in equilibrium.

**(M2):** Given  $(c I)$ , we determine  $(c E)$  as the strongest elimination rule such that  $(c I)$  and  $(c E)$  are in equilibrium.

Putting all the pieces together, we have Tennant's harmony criterion: a connective  $c$  is harmonious if and only if its rules satisfy the requirements of equilibrium and maximality. We will call this T-harmony for short.

The reader can check that the usual sets of rules for classical negation are not in equilibrium, and therefore aren't T-harmonious. Comparing Tennant's notion of harmony with the preceding ones, however, is more complicated, due to some problems that arise from the requirement of maximality.

Whenever we have a rule and several counterparts in equilibrium with it, the requirement of maximality asks us to pick the strongest one. The obvious question is: strongest in what sense? Unlike in the case of propositions, Tennant offers no account of the strength of rules. This is a problem, since there are many reasonable things one can mean by '(c E<sub>1</sub>) is stronger than (c E<sub>2</sub>) given (c I)'.<sup>14</sup>

Intuitively, an elimination rule is stronger than another if it allows us to 'deduce more'. A natural way to make this more precise is by comparing the resulting deducibility relations. Let  $\vdash_{c_1}$  be the relation induced by (c I) and (c E<sub>1</sub>), and  $\vdash_{c_2}$  the relation induced by (c I) and (c E<sub>2</sub>). Then we can say that:

**Strength<sub>1</sub>:** Given (c I), rule (c E<sub>1</sub>) is stronger<sub>1</sub> than rule (c E<sub>2</sub>) if  $\vdash_{c_2} \subseteq \vdash_{c_1}$ .

This is not the only option, though. If we look back at the rules in the beginning of the section, we can see that ( $\vee E_R$ ) is a special case of ( $\vee E$ ), and ( $\rightarrow I_R$ ) is a special case of ( $\rightarrow I$ ). The standard rules allow for more inferential transitions, more passages from premisses to conclusions, than their restricted versions. This suggests that we can also cash out 'deducing more' in terms of derivations, rather than derivability. Here is the corresponding notion of strength:

**Strength<sub>2</sub>:** Rule (c E<sub>1</sub>) is stronger<sub>2</sub> than rule (c E<sub>2</sub>) if all correct applications of (c E<sub>2</sub>) are correct applications of (c E<sub>1</sub>).

Alternatively, we may want to think of 'deducing more' in terms of derivability between rules, rather than derivability between formulas:

**Strength<sub>3</sub>:** Rule (c E<sub>1</sub>) is stronger<sub>3</sub> than rule (c E<sub>2</sub>), given (c I), if (c E<sub>2</sub>) is derivable from (c E<sub>1</sub>) and (c I).

These notions of strength are not equivalent, and depending on which of them we settle on —if any— different pairs of rules will come out harmonious.<sup>14</sup> The requirement of maximality, then, has simply not been stated in full: we still need to know what notion of strength it assumes.<sup>15</sup>

<sup>14</sup>To see that strength<sub>1</sub> is different from strength<sub>2</sub>, compare the usual ( $\wedge E_{1/2}$ ) rule with the general-elimination rule ( $\wedge GE$ ): they are equally strong<sub>1</sub> but neither is stronger<sub>2</sub> than the other. To see that strength<sub>1</sub> is different from strength<sub>3</sub>, note that ( $\vee E$ ) and ( $\vee E_R$ ) are equally strong<sub>1</sub> but the first is strictly stronger<sub>3</sub>. And finally, to see that strength<sub>2</sub> is different from strength<sub>3</sub>, just compare ( $\wedge E_{1/2}$ ) and ( $\wedge GE$ ) again: neither is stronger<sub>2</sub> than the other, but they are equally strong<sub>3</sub>.

<sup>15</sup>A similar point is made in (Ceragioli 2022), but Ceragioli's take differs substantially from mine. Relevantly for this section, Ceragioli holds that something like strength<sub>1</sub> is the 'right' measure of strength.

The problem runs deeper than this, though. A criterion of harmony is supposed to rule out ‘defective’ or ‘illegitimate’ connectives. This means that we can’t simply pick a notion of strength for the requirement of maximality; we also need to explain why that particular notion of strength is the right one. As far as I can tell, however, there is no a unique right measure of the strength of rules to start with. Rather, each of the notions of strength we have considered picks out a reasonable sense in which we can say that a rule is stronger than another.

Suppose we are focused on the fine-grain of proofs, as when proving normalisation results or designing backwards proof-search algorithms. Then something like  $\text{strength}_2$  seems appropriate: it is a measure of strength that takes the shape and number of proof-trees into account. If we are only interested in the yield of rules (as when assessing soundness and completeness), then  $\text{strength}_2$  will not do; it is sensitive to aspects that don’t affect derivability between formulas, and something like  $\text{strength}_1$  seems more adequate. And in other contexts still (say, when studying the admissible rules of some proof-system),  $\text{strength}_3$  would be more fitting. Which measure of strength is right depends on what we want the notion to be sensitive to, but there is no *independent* reason to prefer one over the other.<sup>16</sup>

The rest of Tennant’s proposal offers little guidance as to which notion of strength is best suited for the job. Take the requirement of equilibrium. Its goal is to ensure some facts about the strength of propositions, which is defined in terms of derivability between formulas. This suggests that the relevant notion of strength should be something like  $\text{strength}_1$ . At the same time, Tennant imposes a long list of conditions regarding what the derivations used to show equilibrium must look like, something that only  $\text{strength}_2$  and  $\text{strength}_3$  are sensitive to, although in different ways. We get (at best) mixed signals about which measure of strength to use.

The requirement of maximality also leads to a second, related problem. This requirement asks us to pick *the* strongest rule among our various candidates. In doing so, it assumes that there is always (exactly) one such rule. But this assumption seems unjustified, and examples suggesting that it is false are easy to come by. Consider  $(\wedge I)$  again. As we saw, it is in equilibrium with each of the following eliminations:

$$\frac{A \wedge B}{A/B} (\wedge E_{1/2}) \qquad \frac{A \wedge B \quad \frac{D}{C}}{C} (\wedge GE)^n$$

---

<sup>16</sup>For a useful analogy, take the question of who is the strongest weightlifter. This may be whoever lifts the most weight (in absolute terms). Or it may be whoever lifts the most weight relative to their body mass (pound for pound strength). And in either case, we have to specify both how much weight is lifted and how (the two Olympic weightlifting events, ‘snatch’ and ‘clean and jerk’, have different records). The point is that these measures of the strength of weightlifters come apart, just like our measures of the strength of rules. And just like in the case of rules, there is no single right way to call the shots.

Now, which one is the strongest of the two? This will depend on the notion of strength we want to use. Notably, if strength is defined in any of the ways we have considered, neither rule is the strongest:  $(\wedge E_{1/2})$  and  $(\wedge GE)$  are equally strong<sub>1</sub> and equally strong<sub>3</sub>, and neither is stronger<sub>2</sub> than the other. What's more, this is the most natural thing to say. There is no compelling reason to see either rule as the strongest of the pair; arguably, a measure of strength according to which one of  $(\wedge E_{1/2})$  and  $(\wedge GE)$  was strictly stronger than the other would be a *bad* measure of strength.

We could try to avoid the problem by claiming  $(\wedge E_{1/2})$  and  $(\wedge GE)$  are in fact 'the same rule', but there seem to be good reasons not to do so. According to the few available accounts of what inference rules are,  $(\wedge E_{1/2})$  and  $(\wedge GE)$  are not identical. I am only aware two accounts that deliver precise identity criteria. One is part of the folklore of proof-theory and the other is due to Tennant (2021). The first account defines sequents as a pairs  $(\Gamma, A)$  where  $A$  is a sentence and  $\Gamma$  is a collection of sentences. It then defines sequent-to-sequent transitions as tuples of sequents, and inference rules as sets of sequent-to-sequent transitions. It is easy to check that  $(\wedge E_{1/2})$  and  $(\wedge GE)$ , seen this way, are different rules. The second account, Tennant's, takes rules as inductive clauses on a definition of derivations;  $(\wedge E_{1/2})$  and  $(\wedge GE)$  are distinct rules according to it too.

More importantly  $(\wedge E_{1/2})$  and  $(\wedge GE)$  behave quite differently in proofs, and extending the same calculus with them can produce non-equivalent results. This is best seen in a substructural setting. In structural logics, like classical or intuitionistic logic, open assumptions are collected into sets. In substructural logics they are collected into multisets or sequences.<sup>17</sup> Suppose we are collecting assumptions into multisets, as in linear logic and its variants, and let our background calculus have only a conditional with the usual rules:

$$\frac{\Gamma, [A]^n \quad \mathcal{D} \quad B}{A \rightarrow B} (\rightarrow I)^n \qquad \frac{A \rightarrow B \quad A}{B} (\rightarrow E)$$

If we add a conjunction governed by  $(\wedge E_{1/2})$  and  $(\wedge GE)$ , we have that  $p \wedge q, p \rightarrow r, q \rightarrow s \vdash r \wedge s$ . But if we add a conjunction governed by  $(\wedge I)$  and  $(\wedge E_{1/2})$ , we don't: the closest we get is  $p \wedge q, p \wedge q, p \rightarrow r, q \rightarrow s \vdash r \wedge s$ . In fact, in intuitionistic linear logic these rules can define distinct connectives;  $(\wedge E_{1/2})$  is an elimination rule for the additive conjunction 'with', whereas  $(\wedge GE)$  is an elimination rule for (something close to) the multiplicative conjunction 'tensor'.<sup>18</sup> The usual accounts of inference rules seem to tell the right story.

<sup>17</sup>Sequences are ordered lists, and multisets are lists without order (equivalently, sets with repetitions).

<sup>18</sup>More precisely, the variant of  $(\wedge GE)$  where no vacuous or multiple discharges are allowed is an elimination rule for tensor. For more details on linear logic in a natural deduction setting see (Negri 2002).

Not only do  $(\wedge E_{1/2})$  and  $(\wedge GE)$  produce different sets of proof-trees, but in some standard contexts they also give rise to different derivability relations, and define distinct connectives. The choice between  $(\wedge E_{1/2})$  and  $(\wedge GE)$  makes an important difference, and this looks like a good reason not to identify them.

Let's summarise the discussion. Tennant's harmony criterion consists of the requirements of equilibrium and maximality. There are two problems with the requirement of maximality. The first is that it appeals to the notion of the strength of a rule, but (a) there are multiple, non-equivalent measures of the strength of a rule, and Tennant doesn't clarify which one is at play, and (b) it seems that no principled choice among various measures of strength can be made. The second problem with the requirement of maximality is that it assumes there is a unique maximally strong rule in equilibrium with any given set of rules. This assumption is questionable and, on closer inspection, seems mistaken.

A different family of problems is pointed out in Florian Steinberger's (2009). Consider these standard rules for the existential quantifier:

$$\frac{A(t/x)}{\exists xA(x)} (\exists I) \qquad \frac{\Gamma, [A(t/x)]^n \quad \mathcal{D} \quad B}{\exists xA(x)} (\exists E)^n$$

As usual,  $A(t/x)$  is the result of substituting the term  $t$  for all free occurrences of  $x$  in  $A$ ,  $x$  is freely substitutable for  $t$  in  $A$ , and in applications of  $(\exists E)$  the term  $t$  cannot occur in  $\Gamma, \exists xA(x)$ , or  $B$ .

It is relatively straightforward to show that  $(\exists I)$  and  $(\exists E)$  satisfy the requirement of equilibrium.<sup>19</sup> But as it turns out, a deviant elimination rule  $(\exists E)^*$ , which is just like  $(\exists E)$  but without any restrictions on variables, is also in equilibrium with  $(\exists I)$  —see (Steinberger 2009, p. 659). Moreover, the combination of  $(\exists I)$  and  $(\exists E)^*$  is trivialising, much like the combination of (tonk I) and (tonk E). For example, it allows us to prove that  $P(b)$  follows from  $P(a)$  for any atomic  $P$  and arbitrary  $a$  and  $b$ :

$$\frac{\frac{P(a)}{\exists xP(x)} (\exists I) \quad [P(b)]^1}{P(b)} (\exists E)^*1$$

Steinberger draws the following conclusion:

<sup>19</sup>In fact, as Steinberger notes there are problems already at this stage. Whether an application of  $(\exists E)$  is correct depends, among other things, on which variables occur free in its premises and conclusions. At the same time, the requirement of equilibrium is meant to apply to *propositions*, i.e. sets of logically equivalent formulas. Logically equivalent formulas need not contain the same free variables, though, so ensuring that Tennant's set-up doesn't break down requires some footwork. For more details see (Steinberger 2009, p. 659)

Surely  $(\exists E)^*$  is the strongest elimination rule [in equilibrium with  $(\exists I)$ ] as it enables us to ‘extract’ the most out of sentences containing  $\exists$  in a dominant position. In particular, of course, it is stronger than the ordinary  $(\exists E)$  elimination rule. It follows that it is the wholly unrestricted elimination rule that is in harmony with the introduction rule given above. (Steinberger 2009, p. 659, with notational changes)

Let’s unpack this passage. Steinberger argues that  $(\exists E)^*$  is stronger than  $(\exists E)$ , since the former “enables us to extract the most out of sentences containing  $\exists$  in a dominant position”. It may not be entirely clear which notion of strength Steinberger has in mind, but there is no need to quibble:  $(\exists E)^*$  is stronger than  $(\exists E)$  in all the senses we have looked at. It follows that the standard rules for  $\exists$  do not satisfy the requirement of maximality, and so are disharmonious. This is an unintended consequence of Tennant’s criterion, and arguably a sign that there is something wrong with his approach:  $(\exists I)$  and  $(\exists E)$  are typically seen as a paradigmatically ‘balanced’ pair of rules.

Steinberger also states that  $(\exists E)^*$  is the *strongest* rule in equilibrium with  $(\exists I)$ , and so “it follows that it is the wholly unrestricted elimination rule that is in harmony with the introduction rule”. This is too quick; for all we know some rule stronger than  $(\exists E)^*$  could be in equilibrium with  $(\exists I)$ . All the same, Steinberger’s comment highlights another difficulty with the maximality requirement: in speaking of the strongest introduction or elimination rule with a certain property, the requirement presupposes a full account of the general shape of introductions and eliminations, but it isn’t obvious what such an account must look like.<sup>20</sup>

In response to Steinberger’s criticism Tennant adds a further conservativeness requirement, which he formulates for sequent settings:

We are arguing, in effect, for a certain *conservativeness requirement*: The rules for a logical operator should not allow one to prove (on the assumption that Cut is admissible) any atomic sequent. (Tennant 2010, p. 467)

Translated back to natural deduction jargon, this is the requirement that rules for a logical operator shouldn’t make it possible to prove new inferences between atomic formulas by means of local peaks. Clearly, whether a set of rules is conservative in this sense partly depends on the calculus to which they belong. This doesn’t sit well with Tennant’s insistence that harmony is a relation

---

<sup>20</sup>Tennant (1997, Ch. 9) argues that meaning-constitutive rules should feature exactly one occurrence of exactly one operator; for introduction rules this occurrence should be in the conclusion, and for elimination rules in one of the premises. If this is taken as an account of the shape of introductions and eliminations, it is arguably too restrictive. Many unproblematic connectives (for instance, connectives that can be defined in terms of others that Tennant finds harmonious) can’t be given rules of this type. Elaborating on this point would take us too far afield, but see (Olkhovikov and Schroeder-Heister 2014), (Read 2015a) and (Humberstone and Makinson 2011) for the sort of result I am alluding to.

between rules, rather than between rules and background calculi, but at least it flags Steinberger’s deviant rules as disharmonious in all interesting contexts: Steinberger’s rules, as we saw earlier, allow us to derive  $P(b)$  from  $P(a)$  for all atomic  $P$  and arbitrary  $a$  and  $b$ , and so are non-conservative with respect to any calculus in which this doesn’t hold without them.

Finally, and for the sake of uniformity with the previous sections, we should ask about the justification of Tennant’s harmony criterion. As far as I can tell, not much motivation has been offered. (Tennant 1978) simply puts forward the harmony criterion as a way to cash out the idea that the rules governing connectives should be balanced (p. 74).<sup>21</sup> In (Tennant 1987), on the other hand, we are told that:

The requirement for harmony emerges clearly if one follows a philosophical method that has the *appearance* of empirical speculation about the origins of language, but is actually designed to focus on constitutive features of meaning. This is the method of enquiring after the *aetiology of entrenchment* of expressions in a language and of conventions governing their use. (Tennant 1987, p. 77)

There follows a long discussion about how some connectives could be introduced into a language that lacked logical operators, after which the requirements of equilibrium and maximality are stated without much indication as to how they relate to the previous remarks. And lastly, (Tennant 2020) has it that:

Harmony is a transcendental precondition for the very possibility of logically structured communication. A would-be logical operator that does not display harmony [...] could not possibly be retained within an evolving language after making a first debut. Because the ‘deductive reasoning’ that it would afford would go so haywire, it would have been rapidly selected against. (Tennant 2020, p. 245)

Let’s summarize the discussion once again. Tennant’s original harmony criterion consists of two requirements, the requirement of equilibrium and the requirement of maximality. The requirement of maximality gives rise to problems related to the notion of the strength of a rule and the existence of a maximally strong rule satisfying certain conditions. Moreover, as Steinberger points out, Tennant’s harmony criterion is vulnerable to certain counterexamples. These can be ruled out (relative to any sensible background calculus) by means of an extra conservativeness requirement, making the resulting harmony criterion ‘global’ or ‘context-dependent’. This, however, doesn’t sit well with Tennant’s overall approach. In addition, it is not clear what the motivation for T-harmony is.

---

<sup>21</sup>Strictly speaking (Tennant 1978) only presents the requirement of equilibrium; the requirement of maximality is a later addition.

### 3.3 Taking stock: Harmony and Classical Logic

It is time to take a step back and put this chapter in perspective. We began with Prior's observation that, if one accepts **(AV)** and **(AR)**, it seems to follow that any sentence can be validly inferred from any other. We then went through two ways of dismissing Prior's challenge. The first was to hold on to **(AV)** and **(AR)** and reject Prior's conclusion. The second was to insist that inference rules can be meaning-constituting without being analytically valid—that is, holding on to **(AR)** and rejecting **(AV)**. Insofar as these replies get rid of the Problem of Harmony, they also block any harmony-based objection to classical logic.

After this we focused on the most popular type of response to Prior's argument: restricting **(AR)** by means of harmony criteria. It follows from our discussion that harmony is less of a problem for classical logicians than it is usually made out to be. According to conservativeness-based proposals, like B-harmony and D-harmony, some formalisations of classical logic are harmonious. In contrast, criteria built around an intuitive notion of balance, like I-harmony and T-harmony, have it that classical logic is disharmonious. Whether this should be concerning is unclear, though; I-harmony and T-harmony face serious problems of their own, and there are plausible grounds to reject them.

As we will see shortly, an even stronger response is available for classicists. There are somewhat unusual formalisations of classical logic that satisfy all the harmony criteria in the literature, including I and T-harmony. Their operational rules straightforwardly 'invert' or 'undo' each other in the sense discussed by Prawitz, and they appear to be balanced according to any plausible measure of balance. Arguably, then, these calculi give a general response to harmony-based objections against classical logic. The details of this approach are laid out in the next chapter.

## Chapter 4

# Harmony and Bilateralism

And a bit of augmentation  
is a serious temptation

---

Glenn Gould, *So You Want to  
Write a Fugue*

Ever since Frege's (1997) it has been common to assume that, although assertion is a fundamental speech act, there is no need to consider an opposing speech act of rejection. More recently, some philosophers have challenged this idea (Price 1990, Smiley 1996, Rumfitt 2000). I will say more about these two positions, known as *unilateralism* and *bilateralism*, in Section 1.

There are many ways to incorporate bilateralist ideas into logical machinery. One approach, due to (Smiley 1996), is to augment formal languages with signs for both assertion and rejection, and use rules governing asserted and rejected formulas. As Rumfitt (2000) has shown, some of the resulting calculi for classical logic satisfy the usual harmony requirements. This has prompted many to claim that bilateralism gives classical logicians a way out of the Problem of Harmony, but on closer inspection the issue is more complicated. As we will see, in bilateral settings the usual (or 'unilateral') harmony criteria aren't enough to rule out a host of trivialising, tonk-like connectives. Therefore, the fact that there are 'unilaterally harmonious' rules for the classical connectives can hardly be an indication that there is nothing wrong with them. What we need, it seems, is a way to adapt the standard harmony criteria to bilateral contexts. I will propose one way of doing so in Section 2, and show that Rumfitt's calculi are not harmonious in this new sense.

In Section 3 I will present two bilateral calculi for classical logic which do satisfy the criterion of harmony from Section 2 (as well as all unilateral harmony criteria), and in that sense vindicate the claim that bilateralism can make classical logic harmonious. Finally, in Sections 4 and 5 I will prove normalisation theorems (and some corollaries) for the calculi of Section 3, showing that they are proof-theoretically 'well-behaved'.

## 4.1 Assertion, Rejection and Negation

What is the relation between the speech acts of assertion and rejection? And how do these speech acts relate to negation? Contemporary debates centre around two opposing views, *unilateralism* and *bilateralism*. According to unilateralism the rejection of a sentence  $A$  is best analysed as the assertion of the negated sentence  $\neg A$ . According to bilateralism, in contrast, assertion and rejection are best seen as distinct speech acts.<sup>1</sup>

The point can be put in terms of answers to polar (or ‘yes/no’) questions. Unilateralists and bilateralists agree that affirmative answers to the question ‘Is it the case that  $A$ ?’ should be analysed as assertions of the sentence  $A$ . For example, given the following exchange between speakers Abelard and Eloise:

- (1) Abelard: Is Astrolabe in the kitchen?  
Eloise: Yes.

both sides of the debate agree that Eloise’s ‘Yes’ amounts to an assertion of the sentence ‘Astrolabe is in the kitchen’. The disagreement concerns the analysis negative answers, such as:

- (2) Abelard: Is Astrolabe in the kitchen?  
Eloise: No.

According to unilateralists, Eloise performs the same speech act in (1) and (2): she is asserting something. The difference is that in (1) Eloise asserts the sentence ‘Astrolabe is in the kitchen’, while in (2) she asserts the sentence ‘Astrolabe isn’t in the kitchen’. According to bilateralists, on the other hand, Eloise is performing different speech acts in (1) and (2): in (1) she is asserting, and in (2) she is rejecting. However, both speech acts concern the same sentence, ‘Astrolabe is in the kitchen’.<sup>2</sup>

Here I will focus on what bilateralism looks like in the context of an inferentialist approach to meaning. The general idea goes as follows. According to bilateralists our inferential practice deals in both assertions and rejections: in the course of an argument we will typically assert some sentences and reject others. According to inferentialists, moreover, the meaning of logical operators is given by the rules that govern our inferential practice. But if this practice features both assertions and rejections, so will the rules in question. In the specific case of negation, the approach more or less reverses unilateralism; instead of analysing rejection in terms of assertion and negation, the strategy is to analyse negation in terms of assertion and rejection.

---

<sup>1</sup>Following (Incurvati and Schlöder 2017) I use ‘assertion’ and ‘rejection’ to refer to speech acts, which express attitudes of assent and dissent, respectively, towards sentences. This choice of terminology is common, but some prefer ‘denial’ for what I am calling rejection and ‘rejection’ for what I am calling dissent (e.g. Restall 2005, Ripley 2017).

<sup>2</sup>The classic argument against bilateralism is Frege’s (1997), while (Price 1990) and (Smiley 1996) are standard bilateralist responses. For general introductions to the unilateralism/bilateralism debate see (Ripley 2011, 2020) and (Incurvati and Schlöder 2023).

To flesh out this idea we have to introduce some distinctions. So far I have been talking of rejections rather generally, but there is reason to be a bit more discriminating. Compare the following exchanges:<sup>3</sup>

- (2) Abelard: Is Astrolabe in the kitchen?  
Eloise: No.
- (3) Abelard: Astrolabe is in the kitchen.  
Eloise: No, he is either in the kitchen or the living room.

The bilateralist wants to say that in both cases Eloise’s ‘No’ expresses a rejection. Still, these rejections are different in some important ways. In (2) Eloise’s ‘No’ is equivalent to (though not necessarily identical with) an assertion that Astrolabe isn’t in the kitchen. Indeed, the dialogue could go on as follows:

- (2) (cont.) Abelard: Well, he can only be in the kitchen or the living room, because all the other rooms are empty. Since he isn’t in the kitchen, I guess he must be in the living room.

In (3), on the other hand, Eloise’s ‘No’ isn’t equivalent to the assertion that Astrolabe isn’t the kitchen; her perfectly coherent continuation makes this clear. Using Incurvati and Schlöder’s (2017) terminology, we will say that (2) features a *strong* rejection, and (3) a *weak* one.

We can make better sense of the difference between strong and weak rejections in terms of the notion of commitment. Speakers who strongly reject a sentence *A*, like Eloise in (2), bring upon themselves certain obligations. They must, for example, allow their interlocutors to assert  $\neg A$  in the course of an argument, and assert  $\neg A$  themselves if asked to take a stance on whether *A*; failure to do so would be considered incoherent behaviour by the other parties in a conversation. We will say that a speaker in this situation is *committed to the strong rejection of A*. In contrast, speakers who weakly reject a sentence *A*, as Eloise in (3), bring upon themselves different obligations. Relevantly, they may object to their interlocutors asserting  $\neg A$  in the course of an argument, and need not assert  $\neg A$  if asked to take a stance on the issue.

A similar point can be made about assertions, as the following examples show:

- (4) Abelard (to Eloise): Astrolabe is in the kitchen.
- (5) Abelard (to Eloise): *Perhaps* Astrolabe is in the kitchen. He could also be in the living room, though.

In Incurvati and Schlöder’s (2019) terminology (4) is a strong assertion, and (5) a weak one. By strongly asserting a sentence *A* speakers again bring upon themselves some obligations. For instance, they must allow their interlocutor to assert *A* in the course of an argument, and reject  $\neg A$  if asked to take a

---

<sup>3</sup>Examples (2)-(4) come from (Restall forthcoming).

stance on whether  $\neg A$ ; we will say they are *committed to the strong assertion of A*. In contrast, speakers who weakly assert a sentence  $A$  may object to their interlocutors asserting  $A$  in the course of an argument; they can coherently take the question of whether  $A$  to be unsettled.

According to many bilateralists (e.g. Restall 2005; Ripley 2017; Rumfitt 2000; Smiley 1996) the meaning of negation can be explained in terms of rules governing the (strong) assertion and rejection of certain sentences.<sup>4</sup> These rules can be concisely represented if we use signed languages, where each formula is preceded by a force indicator ‘+’ or ‘-’;  $+A$  stands for the assertion of  $A$  and  $-A$  for its rejection. Since ‘+’ and ‘-’ represent the force of speech acts they can’t be embedded or iterated, but for the rest the unsigned part of formulas is built up from atomic expressions and logical operators in the usual way. In this setting Smiley (1996) proposes the following rules for classical negation:

$$\frac{-A}{+\neg A} (+\neg I) \quad \frac{+\neg A}{-A} (+\neg E) \quad \frac{+A}{-\neg A} (-\neg I) \quad \frac{-\neg A}{+A} (-\neg E)$$

Rule  $(+\neg I)$  states that if a speaker is committed to the rejection of  $A$ , then they are committed to the assertion of  $\neg A$ . Conversely,  $(+\neg E)$  states that if a speaker is committed to the assertion of  $\neg A$  then they are committed to the rejection of  $A$ . These two rules should clarify the sense in which, in **(2)**, Eloise does something equivalent to asserting  $\neg A$ ; given  $(+\neg I)$  and  $(+\neg E)$  speakers are committed to the rejection of  $A$  exactly when they are committed to the assertion of  $\neg A$ . The second pair of rules is read analogously:  $(-\neg I)$  states that commitment to the assertion of  $+A$  entails commitment to the rejection of  $\neg A$ , and  $(-\neg E)$  that commitment to the rejection of  $\neg A$  entails commitment to the assertion of  $A$ .<sup>5</sup>

This approach to explaining the role of negation in inference (and so, for inferentialists, its meaning) can be extended to the rest of the operators of classical propositional logic. In the resulting picture, the meaning of logical constants is explained in terms of the effects that sentences containing them have for the commitments of speakers. The next section lays this out in more detail, and explains why the resulting calculi are relevant for the harmony debate.

## 4.2 Bilateralism, Classical Logic, and Harmony

In his (2000) Ian Rumfitt formalises classical propositional logic using natural deduction rules for signed formulas. His operational rules for the connectives are as follows:<sup>6</sup>

<sup>4</sup>From now on, ‘assertion’ and ‘rejection’ mean strong assertion and rejection, respectively.

<sup>5</sup>The idea of reading these rules in terms of commitment comes from (Incurvati and Schlöder 2017), who take inspiration from Brandom (1983, 1994). Rumfitt (2000, 2002) understands the rules in terms of conventions about the correct assertibility of sentences.

<sup>6</sup>Some notational conventions: unless otherwise stated, uppercase Roman letters range over unsigned formulas, uppercase Greek letters over sets of signed formulas, and lowercase Greek

### Conjunction:

$$\frac{+A_1 \quad +A_2}{+A_1 \wedge A_2} (+\wedge \text{I})$$

$$\frac{-A_i}{-A_1 \wedge A_2} (-\wedge \text{I})$$

$$\frac{+A_1 \wedge A_2}{+A_i} (+\wedge \text{E})$$

$$\frac{\Gamma_1, [-A_1]^1 \quad \Gamma_2, [-A_2]^1 \quad \mathcal{D}_1 \quad \mathcal{D}_2 \quad \varphi}{-A_1 \wedge A_2 \quad \varphi} (-\wedge \text{E})^1$$

### Disjunction:

$$\frac{+A_i}{+A_1 \vee A_2} (+\vee \text{I})$$

$$\frac{-A_1 \quad -A_2}{-A_1 \vee A_2} (-\vee \text{I})$$

$$\frac{\Gamma_1, [+A_1]^1 \quad \Gamma_2, [+A_2]^1 \quad \mathcal{D}_1 \quad \mathcal{D}_2 \quad \varphi}{+A_1 \vee A_2 \quad \varphi} (+\vee \text{E})^1$$

$$\frac{-A_1 \vee A_2}{-A_i} (-\vee \text{E})$$

### Implication:

$$\frac{\Gamma, [+A_1]^1 \quad \mathcal{D} \quad +A_2}{+A_1 \rightarrow A_2} (+\rightarrow \text{I})^1$$

$$\frac{+A_1 \quad -A_2}{-A_1 \rightarrow A_2} (-\rightarrow \text{I})$$

$$\frac{+A_1 \rightarrow A_2 \quad +A_2}{+A_2} (+\rightarrow \text{E})$$

$$\frac{-A_1 \rightarrow A_2}{+A_1 / -A_2} (-\rightarrow \text{E})$$

### Negation:

$$\frac{-A}{+\neg A} (+\neg \text{I})$$

$$\frac{+A}{-\neg A} (-\neg \text{I})$$

letters over signed formulas. Brackets indicate discharged assumptions, and both multiple and vacuous discharges are allowed. I will indicate vacuous discharges with the superscript  $v$  next to the relevant application of a rule.

$$\frac{+\neg A}{-A} (+\neg E)$$

$$\frac{-\neg A}{+A} (-\neg E)$$

In addition to operational rules, Rumfitt’s calculus contains *coordination principles*. These are rules that govern the interaction between ‘+’ and ‘−’, rather than specific connectives. They are supposed to capture general norms about the relation between assertion and rejection. Rumfitt’s first coordination principle is the rule of *Rejection*, which uses a primitive symbol  $\perp$  to indicate that a certain position is incoherent:<sup>7</sup>

$$\frac{+A \quad -A}{\perp} (\text{Re})$$

(Re), then, simply states that commitment to the assertion and the rejection of the same sentence is incoherent. In addition, his system has the rule of *Smilean Reductio*:

$$\begin{array}{ccc} \Gamma, [+A]^1 & & \Gamma, [-A]^1 \\ \mathcal{D} & & \mathcal{D} \\ \frac{\perp}{-A} (\text{SR}_1)^1 & & \frac{\perp}{+A} (\text{SR}_2)^1 \end{array}$$

The first half of *Smilean Reductio* states that if committing to the assertion of  $A$  would lead to incoherence, then one is committed to the rejection of  $A$ . The second half states that if commitment to the rejection of  $A$  would lead to incoherence, then one is committed to its assertion. *Smilean Reductio* also encodes a form of explosion: by vacuously discharging an occurrence of  $+A$  or  $-A$ , the rule allows one to derive any  $\varphi$  from  $\perp$ .

We will call the resulting system  $\mathcal{R}$ .  $\mathcal{R}$  gives rise to a derivability relation between signed formulas, just like standard (or ‘unilateral’) calculi give rise to derivability relations between unsigned ones.  $\mathcal{R}$ ’s signed consequence relation coincides with the usual formalisations of classical logic in its  $+$ -signed fragment but, unlike standard calculi, makes room for rejections too. Let’s introduce some new terminology. Given an unsigned language  $\mathcal{L}$ , where formulas are built from atoms  $p, q, \dots$  and connectives  $\wedge, \vee, \rightarrow$  and  $\neg$  in the usual way, let  $\mathcal{L}_s := \{+A \mid A \in \mathcal{L}\} \cup \{-A \mid A \in \mathcal{L}\}$  be the signed version of  $\mathcal{L}$ . Say also that, for any  $\Gamma \subseteq \mathcal{L}$ , the set  $\Gamma^+$  of signed formulas is  $\Gamma^+ := \{+B \mid B \in \Gamma\}$ . Then as (Rumfitt 2000) shows, for any  $\Gamma \cup \{A\} \subseteq \mathcal{L}$  we have that:

**Observation 4.1.**  $\Gamma \vdash_{CL} A$  if and only if  $\Gamma^+ \vdash_{\mathcal{R}} +A$ .

<sup>7</sup>Following (Tennant 1999) Rumfitt takes  $\perp$  to be a punctuation sign in derivations, rather than a formula. Since it is not a formula, it cannot be embedded or signed.

Rumfitt’s coordination principles are unusual in several ways. First, although *Rejection* and *Smilean Reductio* are distinct rules, all applications of *Smilean Reductio* require a previous occurrence of  $\perp$ , and this occurrence can only result from an application of (Re). In other words, it doesn’t make much sense to consider (SR) independently of (Re). Similarly, it is not clear that it makes sense to consider (Re) independently of (SR). Recall that for Rumfitt  $\perp$  is not the conclusion of an argument; it is a punctuation sign, a placeholder indicating that a specific proof configuration has been reached. To some extent it is (SR) that does all the deductive work, and we can see (Re) as a dispensable graphical device. In fact, the original formulation of *Smilean Reductio* (see Smiley 1996) was:

$$\frac{\begin{array}{c} \Gamma_1, [\varphi]^1 \\ \mathcal{D}_1 \\ \psi \end{array} \quad \begin{array}{c} \Gamma_2, [\varphi]^1 \\ \mathcal{D}_2 \\ \bar{\psi} \end{array}}{\bar{\varphi}} 1$$

which is easily seen to be equivalent to the combination of (SR) and (Re).<sup>8</sup> In addition, *Smilean Reductio* does two distinct jobs at the same time. First, it states that commitment to the assertion and the rejection of a formula jointly exhaust the possibilities for any speaker. And secondly, it also codifies a principle of explosion. We can get a clearer picture of the resulting relation between assertion and rejection if we swap Rumfitt’s coordination principles for the more perspicuous *Explosion* and *Bilateral Excluded Middle*:

$$\frac{+A \quad -A}{\varphi} \text{ (Ex)} \quad \frac{\begin{array}{c} \Gamma_1, [+A]^1 \\ \mathcal{D}_1 \\ \varphi \end{array} \quad \begin{array}{c} \Gamma_2, [-A]^1 \\ \mathcal{D}_2 \\ \varphi \end{array}}{\varphi} \text{ (Bem)}^1$$

It is routine to check that (Re), (SR<sub>1</sub>) and (SR<sub>2</sub>) are derivable from (Ex) and (Bem), and vice versa. Now, say  $\mathcal{R}'$  is the calculus containing Rumfitt’s operational rules plus (Ex) and (Bem). It follows from the fact that (Ex) and (Bem) are interderivable with (Re) and (SR<sub>1/2</sub>), combined with Observation 4.1, that:

**Observation 4.2.**  $\Gamma \vdash_{CL} A$  if and only if  $\Gamma^+ \vdash_{\mathcal{R}'} +A$ .

In contrast with unilateral formalisations of classical logic,  $\mathcal{R}$  and  $\mathcal{R}'$  satisfy all the usual harmony criteria. It is intuitively clear why this should be so, at least for criteria that bank on an informal notion of balance. Take the rules

<sup>8</sup>Here and in what follows  $\bar{\varphi}$  denotes the *conjugate* of  $\varphi$ , where the conjugate of  $+A$  is  $-A$  and vice versa.

for negation: in order to apply  $(+\neg I)$  one needs to derive a sentence of the form  $-A$ , which is precisely what an application of  $(+\neg E)$  yields. Similarly, in order to apply  $(-\neg I)$  one needs a derivation of  $+A$ , which is precisely what an application of  $(-\neg E)$  yields. The introductions and eliminations of the same sign straightforwardly ‘undo’ or ‘invert’ each other. Something similar happens with the rest of connectives, and the proof that Rumfitt’s operational rules are strongly conservative is straightforward as well (see Rumfitt 2000 for details).

On the face of it we have a general response to harmony-based objections to classical logic: if we take the difference between assertion and rejection seriously, classical logic seems as inferentialist-friendly as we could wish for (Rumfitt 2000, Sec. VII). On closer inspection, however, this is simply too quick. Granted, Rumfitt’s bilateral system satisfies all the harmony criteria put forward for *unilateral* calculi. But in a bilateral setting these criteria can’t to rule out all trivialising connectives. Consider, for example, *bank*:

$$\begin{array}{cc} \frac{+A \quad -A}{+\text{bank } A} (+\text{bank } I) & \frac{-A}{-\text{bank } A} (-\text{bank } I) \\ \\ \frac{+\text{bank } A}{+A/-A} (+\text{bank } E) & \frac{-\text{bank } A}{-A} (-\text{bank } E) \end{array}$$

The *bank* rules for each sign are direct inverses of each other, and *bank* is harmonious according to all the usual measures of ‘balance’. Still, in the presence of  $(SR_1)$  *bank* trivialises any consequence relation. The following derivation, for example, shows that it allows us to derive  $-A$  for any  $A$ :

$$\frac{\frac{[+\text{bank } A]^1}{+A} (+\text{bank } E) \quad \frac{[+\text{bank } A]^1}{-A} (+\text{bank } E)}{\frac{\perp}{-\text{bank } A} (SR_1)^1} \frac{}{-A} (-\text{bank } E)$$

The same point can be made with respect to the other half of *Smilean Reductio*. Consider *benk*, the mirror image of *bank*:

$$\begin{array}{cc} \frac{-A}{+\text{benk } A} (+\text{benk } I) & \frac{+A \quad -A}{-\text{benk } A} (-\text{benk } I) \\ \\ \frac{+\text{benk } A}{-A} (+\text{benk } E) & \frac{-\text{benk } A}{+A/-A} (-\text{benk } E) \end{array}$$

Just as before, these rules are balanced according to all the usual standards. Still, when combined with (SR<sub>2</sub>) they allow us to derive  $-A$  for any  $A$ :

$$\frac{\frac{[- \text{benk } A]^1}{+A} (- \text{benk } E) \quad \frac{[- \text{benk } A]^1}{-A} (- \text{benk } E)}{\frac{\perp}{+ \text{benk } A} (\text{SR}_2)^1 \quad (+ \text{benk } E)} \frac{}{-A}$$

The problem is not specific to *Smilean Reductio*. The following introduction and elimination rules, for instance, are also balanced according to the usual criteria:

$$\frac{+A}{+A \text{ blink } B} (+ \text{blink } I) \quad \frac{-B}{-A \text{ blink } B} (- \text{blink } I)$$

$$\frac{+A \text{ blink } B}{+A} (+ \text{blink } E) \quad \frac{-A \text{ blink } B}{-B} (- \text{blink } E)$$

In the presence of (Re), however, they make any assertion incompatible with any rejection:

$$\frac{\frac{+A}{+A \text{ blink } B} (+ \text{blink } I) \quad \frac{-B}{-A \text{ blink } B} (- \text{blink } I)}{\perp} (\text{Re})$$

What can we learn from these examples? The usual measures of balance between introduction and elimination rules, and more generally, the usual criteria of harmony, were all put forward with unilateral calculi in mind. In these calculi what one can do with a connective depends on operational (and perhaps structural) rules. In bilateral systems, however, coordination principles license further inferences. As a result, connectives that are acceptable according to the usual standards can become tonk-like when they interact with the wrong coordination principle. This suggests that, when we move to a bilateral setting, the notion of harmony should take into account *both* the relation between introduction and elimination rules (or whatever other feature is taken to determine ‘unilateral’ harmony), *and* the relation between operational rules and coordination principles.

One way of going about this suggests itself when we take a closer look at how blink, benk and blink interact with coordination principles. Take blink and (Re). (Re) stipulates that commitment to the assertion and the rejection of the

same sentence is incoherent. *A fortiori*, commitment to  $+(A \text{ blink } B)$  and  $-(A \text{ blink } B)$  is incoherent. When we look at the rules for blink in isolation, though, there appears to be no incompatibility between these two sentences;  $+(A \text{ blink } B)$  is equivalent to  $+A$  and  $-(A \text{ blink } B)$  is equivalent to  $-B$ . This intuitive ‘mismatch’ has a correlate at the level of derivability: if we postulate (Re) for atomic sentences only there is no way to derive  $\perp$  from  $+(A \text{ blink } B)$  and  $-(A \text{ blink } B)$ .<sup>9</sup> Contrast this with the relation between the negation rules and (Re). It is easy to see that if (Re) holds in a given calculus, then it will automatically hold for the expansion of that calculus with  $\neg$  and its rules, because we can always mimic applications of (Re) to formulas governed by  $\neg$  by applications of the (Re) to their immediate subformulas:

$$\frac{+\neg A \quad -\neg A}{\perp} \text{ (Re)} \quad \rightsquigarrow \quad \frac{\frac{+\neg A}{-A} (+\neg \text{ E}) \quad \frac{-\neg A}{+A} (+\neg \text{ E})}{\perp} \text{ (Re)}$$

The same happens with respect to blink and (SR<sub>1</sub>). Given (SR<sub>1</sub>) we are committed to  $-\text{blink } A$  for arbitrary  $A$ .<sup>10</sup> If we look at the rules for blink in isolation, however, there is no reason why this should be so:  $-\text{blink } A$  is equivalent to  $-A$ . Just as before, if we postulate (SR<sub>1</sub>) only for atomic sentences, we cannot derive it for sentences governed by blink.<sup>11</sup> In contrast, if (SR<sub>1</sub>) holds for a given calculus, then it will also hold for the extension of that calculus with  $\neg$ :

$$\frac{\frac{[\neg A]^1}{+\neg A} (+\neg \text{ I})}{\mathcal{D}} \quad \frac{\frac{\perp}{+A} (\text{SR}_1)^1}{-\neg A} (-\neg \text{ I})}{\frac{[\neg A]^1}{+\neg A} (+\neg \text{ I})}{\mathcal{D}} \quad \frac{\perp}{+A} (\text{SR}_1)^1}{-\neg A} (-\neg \text{ I})} \rightsquigarrow \frac{[\neg A]^1}{+\neg A} (+\neg \text{ I})}{\mathcal{D}} \quad \frac{\perp}{+A} (\text{SR}_1)^1}{-\neg A} (-\neg \text{ I})$$

Let’s generalise this idea. We will say that coordination principles  $\mathbf{P}_1 \dots \mathbf{P}_n$  are *preserved* by the rules for a connective  $\mathbf{c}$  if, when  $\mathbf{P}_1 \dots \mathbf{P}_n$  are restricted to atomic sentences, all their instances for sentences with  $\mathbf{c}$  as their main operator are derivable. Preservation of coordination principles is, as we have seen, a correlate of the intuitive ‘mismatch’ between blink, bink, and certain coordination principles. This suggests the following bilateral criterion of harmony:

**Bilateral harmony:** A connective  $\mathbf{c}$  is bilaterally harmonious, given coordination principles  $\mathbf{P}_1 \dots \mathbf{P}_n$ , iff (i)  $(+\mathbf{cI})$  and  $(+\mathbf{cE})$  are unilaterally harmonious; (ii)  $(-\mathbf{cI})$  and  $(-\mathbf{cE})$  are unilaterally harmonious; (iii)  $\mathbf{P}_1 \dots \mathbf{P}_n$  are preserved by the rules for  $\mathbf{c}$ .

<sup>9</sup>For a proof see the Appendix to this chapter.

<sup>10</sup>Strictly speaking, given (SR<sub>1</sub>) and (Re); as we have seen *Smilean Reductio* can’t be teased apart from *Rejection*.

<sup>11</sup>See again the Appendix for a proof.

This criterion doesn't commit us to any particular notion of unilateral harmony. This is a feature, not a bug. The debate on unilateral harmony is far from settled, and defining things this way is enough to adapt unilateral harmony criteria to bilateral settings while leaving bilateralists wiggle room to adopt their favourite unilateral criterion. The idea, simply put, is that whatever unilateral harmony consists in, bilateral harmony is *that* plus preservation of all available coordination principles.<sup>12</sup>

This criterion has, I take it, some plausibility. Coordination principles lay down general norms concerning the relation between assertion and rejection. When we define a connective by means of operational rules, we want this definition to 'fit together' with those independently given stipulations. Given the examples above, it seems that preservation of all coordination principles is at least a sufficient condition for this vague 'fitting together' to obtain. If we couple this requirement with a suitable measure of harmony between introductions and eliminations, we get a workable criterion bilateral harmony. Alternatively, we could argue as follows. In bilateral calculi the behaviour of connectives is determined by both operational rules and coordination principles, so bilateral harmony criteria should take both into account. If we ensure that coordination principles for complex formulas can be dispensed with, though, the 'bilateral problem of harmony' reduces to its unilateral part. This is essentially what the present idea of harmony ensures, so we can again take it as a sufficient condition for harmony.

Now, how does the new criterion fare? As I have already mentioned, it rules out *bink*, *benk* and *blink*, since their rules don't to preserve the coordination principles which make them trivialising. This is not because our examples have been cherry-picked. Many other 'bilateral tonks' have been proposed in the literature (see e.g. Francez 2014, Kürbis 2021 and Simonelli, forthcoming), and the criterion rules them out as well. In fact, the new criterion of harmony rules out more than we might expect. Rumfitt (2000) has shown that all his operational rules preserve the coordination principle of *Rejection*. As Ferreira (2008) points out, however, some of Rumfitt's rules don't preserve *Smilean Reductio*. Rumfitt's  $\mathcal{R}$ , therefore, is not harmonious in our new sense. It seems that we have come full circle: in order to avoid the problem of harmony we moved to a bilateral setting, where classical connectives can be given rules that are harmonious according to all the usual standards. It then became clear that, in this new setting, the usual standards of harmony aren't enough. To solve this we proposed a specifically bilateral criterion of harmony. But according to this new criterion, the bilateral rules for classical connectives aren't harmonious after all. Luckily, there is a way to solve this problem. I will present the general idea behind it in the next section, and take care of the more technical details in Sections 4 and 5.

---

<sup>12</sup>The idea of cashing out bilateral harmony (partly) in terms of preservation is not new. Rumfitt requires that all operational rules preserve (Re), and considers this "a precondition for the connectives to possess coherent bilateral senses" (Rumfitt 2008, p. 1060). As *bink* and *benk* make clear, though, preservation of (Re) isn't enough to rule out all tonk-like operators.

### 4.3 Harmony at last

Fernando Ferreira's (2008) gives a model-theoretic proof that some of Rumfitt's operational rules don't preserve *Smilean Reductio*. By looking at the matter proof-theoretically we can get a clear idea of exactly which rules these are. We know from the previous section that there is a scheme for re-writing applications of (SR<sub>1</sub>) to formulas governed by  $\neg$  as applications of (SR<sub>1</sub>) to their immediate subformulas. An analogous scheme is available for applications of (SR<sub>2</sub>):

$$\frac{[\neg\neg A]^1}{\mathcal{D}} \quad \frac{\perp}{+\neg A} \text{ (SR}_2\text{)}^1 \quad \rightsquigarrow \quad \frac{[+A]^1}{-\neg A} \text{ (}\neg\neg\text{ I)} \quad \frac{\perp}{-A} \text{ (SR}_2\text{)}^1 \quad \frac{}{+\neg A} \text{ (}\neg\text{ I)}$$

Using these two schemes we can show that, if *Smilean Reductio* is laid down for atomic formulas, then it is derivable for complex formulas where  $\neg$  is the main operator. In other words, the rules for  $\neg$  preserve *Smilean Reductio*. The same holds for Rumfitt's rules for the conditional, given the following re-writing schemes:<sup>13</sup>

$$\frac{[+A \rightarrow B]^1}{\mathcal{D}} \quad \frac{\perp}{-A \rightarrow B} \text{ (SR}_1\text{)}^1 \quad \rightsquigarrow \quad \frac{[+A]^1 \quad [A]^{-2}}{\frac{\perp}{+B} \text{ (SR}_2\text{)}^v \quad \frac{}{+A \rightarrow B} \text{ (}\rightarrow\text{ I)}^1} \quad \frac{[+B]^3}{\frac{}{+A \rightarrow B} \mathcal{D}} \quad \frac{\perp}{+A} \text{ (SR}_2\text{)}^2 \quad \frac{\perp}{-B} \text{ (SR}_1\text{)}^3 \quad \frac{}{-A \rightarrow B}$$

$$\frac{[-A \rightarrow B]^1}{\mathcal{D}} \quad \frac{\perp}{+A \rightarrow B} \text{ (SR}_2\text{)}^1 \quad \rightsquigarrow \quad \frac{[+A]^1 \quad [B]^{-2}}{\frac{}{-A \rightarrow B} \mathcal{D}} \quad \frac{\perp}{-A} \text{ (SR}_1\text{)}^1 \quad \frac{[+A]^3}{\frac{\perp}{+B} \text{ (SR}_2\text{)}^2 \quad \frac{}{+A \rightarrow B} \text{ (}\rightarrow\text{ I)}^3}$$

<sup>13</sup>Recall that the superscript  $v$  indicates the vacuous discharge of an assumption.

It is also easy to find re-writing schemes for applications of (SR<sub>1</sub>) that discharge sentences of the form  $+A \vee B$  and, dually, for applications of (SR<sub>2</sub>) that discharge sentences of the form  $-A \wedge B$ :

$$\begin{array}{ccc}
\frac{[+A \vee B]^1}{\frac{\mathcal{D}}{\frac{\perp}{-A \vee B} (\text{SR}_1)^1}} & \rightsquigarrow & \frac{\frac{[+A]^1}{+A \vee B} (+\vee \text{I})}{\mathcal{D}} \quad \frac{\frac{[+B]^2}{+A \vee B} (+\vee \text{I})}{\mathcal{D}} \\
& & \frac{\frac{\perp}{-A} (\text{SR}_1)^1}{-A \vee B} \quad \frac{\frac{\perp}{-B} (\text{SR}_1)^2}{-B} (-\vee \text{I}) \\
\\
\frac{[-A \wedge B]^1}{\frac{\mathcal{D}}{\frac{\perp}{+A \wedge B} (\text{SR}_1)^1}} & \rightsquigarrow & \frac{[-A]^1}{-A \wedge B} (-\wedge \text{I}) \quad \frac{[-B]^2}{-A \wedge B} (-\wedge \text{I}) \\
& & \frac{\mathcal{D}}{\frac{\perp}{+A} (\text{SR}_1)^1} \quad \frac{\mathcal{D}}{\frac{\perp}{+B} (\text{SR}_1)^2} \\
& & \frac{\perp}{+A \wedge B} (+\wedge \text{I})
\end{array}$$

This means that the only problematic uses of *Smilean Reductio* are those in which sentences of the form  $-A \vee B$  or  $+A \wedge B$  are discharged. They are the source of disharmony in Rumfitt's system, and if we take care of them we will get a bilaterally harmonious calculus. This is not difficult to achieve. Consider the following rules for rejected conjunctions and asserted disjunctions:

$$\begin{array}{ccc}
\frac{[+A]^1}{\frac{\mathcal{D}}{\frac{-B}{-A \wedge B} (-\wedge \text{I})^1}} & & \frac{-A \wedge B}{-B} \quad \frac{+A}{-A} (-\wedge \text{E}) \\
\\
\frac{[-A]^1}{\frac{\mathcal{D}}{\frac{+B}{+A \vee B} (+\vee \text{I})^1}} & & \frac{+A \vee B}{+B} \quad \frac{-A}{-A} (+\vee \text{E})
\end{array}$$

As we are about to see, these rules preserve all the coordination principles we have considered. More generally, let  $\mathcal{B}$  be the set of Rumfitt's operational rules, with our modifications for conjunction and disjunction  $\mathcal{B}$ . Then we have that:

**Observation 4.3.** All the rules in  $\mathcal{B}$  preserve *Smilean Reductio* and *Rejection*, as well as *Explosion* and *Bilateral Excluded Middle*.

Observation 4.3 follows trivially from the normalisation theorems to be proven shortly. Now let's define two calculi,  $\mathcal{HB}_1$  and  $\mathcal{HB}_2$ .  $\mathcal{HB}_1$  consists of  $\mathcal{B}$ ,  $(\text{SR}_{1/2})$ , and  $(\text{Re})$ , and  $\mathcal{HB}_2$  consists of  $\mathcal{B}$ ,  $(\text{Ex})$ , and  $(\text{Bem})$ . By Observation 4.3 in both these calculi all coordination principles are preserved. Moreover, their operational rules are unilaterally harmonious.<sup>14</sup> And finally, it is routine to check that:

**Observation 4.4.**  $\mathcal{HB}_1$  and  $\mathcal{HB}_2$  are equivalent to Rumfitt's calculus (i.e.  $\Gamma \vdash_{\mathcal{R}} \varphi$  iff  $\Gamma \vdash_{\mathcal{HB}_{1/2}} \varphi$ ).

$\mathcal{HB}_1$  and  $\mathcal{HB}_2$ , then, are bilaterally harmonious calculi for classical logic. In the next two sections I will prove normalisation theorems for them. As a corollary to these theorems we will get not just Observation 4.3, but also the usual separation and subformula properties, and a straightforward comparison between the shape of normal derivations in  $\mathcal{HB}_1$  and  $\mathcal{HB}_2$ , on the one hand, and those from Prawitz's (1965) on the other.

## 4.4 Harmony and Normalisation: $\mathcal{HB}_1$

All derivations in normal form, according to Dag Prawitz's (1965) original result, share a central feature: no formula occurrence in them is simultaneously the consequence of an introduction rule and the major premise of an elimination. In the terminology of the previous section, normal derivations don't contain any local peaks. This is what inspires I-harmony. If the operational rules for a connective are 'balanced', the thought goes, one should gain nothing by first introducing and then immediately eliminating a connective. Therefore, it should be possible to eliminate all such steps within a derivation.<sup>15</sup> This is the core principle behind normalisation, and so for normal derivations in  $\mathcal{HB}_1$  we will require that:

- (i) No conclusion of an I-rule is a major premise of an E-rule.

The notion of bilateral harmony from the previous section suggests a similar principle, this time regarding the interaction between operational rules and coordination principles. Speaking very loosely, we can say that a connective is 'balanced' with respect to a coordination principle if one can lay down the coordination principle for atoms and prove it for complex sentences. To reflect this at the level of derivations we will require that applications of coordination principles to complex formulas be eliminable. In normal  $\mathcal{HB}_1$  derivations, then:

<sup>14</sup>This is obvious in the case of local criteria of harmony, since the rules of each sign simply invert each other. The fact that the operational rules are strongly conservative also follows from the normalisation theorems below (see corollaries 4.14 and 4.22).

<sup>15</sup>This is not the whole story. In the presence of, for example, the usual rules for disjunction, one can introduce a connective and eliminate it a few steps below in the derivation, rather than immediately after the introduction. To cover these cases normalisation proofs usually include so-called 'permutative reductions'. This sort of problem can't arise with the modified rules in  $\mathcal{HB}_1$ , though, so we need not worry about it.

(ii) Coordination principles are applied only to atoms.

Clauses (i) and (ii) are enough to ensure that normal derivations have the separation property, and therefore, that the operational rules of  $\mathcal{HB}_1$  are strongly conservative. They are not enough to secure the stronger subformula property, however, as the following derivation shows:

$$\frac{\frac{+p \quad -p}{\frac{\perp}{+q} (\text{SR}_2)^v} \quad \frac{+p \quad -p}{\frac{\perp}{-q} (\text{SR}_1)^v}}{\frac{\perp}{+r} (\text{SR}_2)^v}$$

This derivation satisfies (i) and (ii), but contains signed formula occurrences  $+q$  and  $-q$  which are subformulas of neither the assumptions nor the conclusion. This is due to the fact that the form of explosion encoded by *Smilean Reductio* is used twice, consecutively. To obtain the subformula property for normal derivations we just have to require that:

(iii) No conclusion of *Smilean reductio* is a premise of *Rejection*.

Putting everything together, we have the following definition of normal form.

**Definition 4.5. Normal form**

A derivation in  $\mathcal{HB}_1$  is in normal form if in it: (i) No conclusion of an I-rule is a major premise of an E-rule. (ii) Coordination principles are applied only to atoms. (iii) No conclusion of *Smilean Reductio* is a premise of *Rejection*.

Formula occurrences which violate clauses (i), (ii) and (iii) are called **maximal operational formulas**, **maximal coordination formulas** and **ancillary maximal formulas**, respectively.

In the rest of this section I will prove normalisation and corollaries for  $\mathcal{HB}_1$ . The first step is laying down the appropriate re-writing schemes, or ‘reduction steps’ as I will call them from now on. Since the rules for disjunction and conjunction are analogous I won’t give reduction steps for the latter; they can be obtained from those for disjunction by switching  $+$  for  $-$  and vice versa. Finally, it may be useful to keep the overall normalisation strategy in mind when examining the reduction steps. Reductions for maximal coordination formulas may create maximal operational formulas of the same complexity. Reduction steps for maximal operational formulas, on the other hand, may create new ancillary maximal formulas only. Finally, reduction steps for *atomic* ancillary maximal formulas create no new maximal formulas of any kind. Therefore, the normalisation process reduces maximal coordination formulas first, followed by maximal operational formulas, and then ancillary maximal formulas.

### 4.4.1 Operational Reductions

**Negation:**

$$\frac{\mathcal{D}_1}{\frac{-A}{+\neg A} (+\neg I)} \quad \frac{\mathcal{D}_1}{-A} (+\neg E) \quad \rightsquigarrow \quad \frac{\mathcal{D}_1}{-A}$$

$$\frac{\mathcal{D}_1}{\frac{+A}{-\neg A} (-\neg I)} \quad \frac{\mathcal{D}_1}{+A} (-\neg E) \quad \rightsquigarrow \quad \frac{\mathcal{D}_1}{+A}$$

**Implication:**

$$\frac{[+A]^1}{\frac{\mathcal{D}_1}{+A \rightarrow B} (+\rightarrow I)^1} \quad \frac{\mathcal{D}_2}{+A} (+\rightarrow E) \quad \rightsquigarrow \quad \frac{\mathcal{D}_2}{+A} \quad \mathcal{D}_1 \quad +B$$

$$\frac{\mathcal{D}_1}{+A} \quad \frac{\mathcal{D}_2}{-B} (-\rightarrow I) \quad \frac{\mathcal{D}_2}{-A \rightarrow B} (-\rightarrow E) \quad \rightsquigarrow \quad \frac{\mathcal{D}_{1/2}}{+A / -B}$$

**Disjunction:**

$$\frac{[-A]^1}{\frac{\mathcal{D}_1}{+A \vee B} (+\vee I)^1} \quad \frac{\mathcal{D}_2}{-A} (+\vee E) \quad \rightsquigarrow \quad \frac{\mathcal{D}_2}{-A} \quad \mathcal{D}_1 \quad +B$$

$$\frac{\mathcal{D}_1}{-A} \quad \frac{\mathcal{D}_2}{-B} (-\vee I) \quad \frac{\mathcal{D}_2}{-A \vee B} (-\vee E) \quad \rightsquigarrow \quad \frac{\mathcal{D}_{1/2}}{-A / -B}$$

#### 4.4.2 Reducing (Re) to Atoms

**Negation:**

$$\frac{\frac{+\neg A}{\perp} \quad -\neg A}{\perp} \text{ (Re)} \quad \rightsquigarrow \quad \frac{\frac{+\neg A}{-A} \quad \frac{-\neg A}{+A}}{\perp} \text{ (Re)}$$

**Implication:**

$$\frac{\frac{\mathcal{D}_1}{+A \rightarrow B} \quad \frac{\mathcal{D}_2}{-A \rightarrow B}}{\perp} \text{ (Re)} \quad \rightsquigarrow \quad \frac{\frac{\mathcal{D}_1}{+A \rightarrow B} \quad \frac{\frac{\mathcal{D}_2}{-A \rightarrow B}}{+A}}{+B} \quad \frac{\mathcal{D}_2}{-A \rightarrow B}}{-B} \text{ (Re)}{\perp}$$

**Disjunction:**

$$\frac{\frac{\mathcal{D}_1}{+A \vee B} \quad \frac{\mathcal{D}_2}{-A \vee B}}{\perp} \text{ (Re)} \quad \rightsquigarrow \quad \frac{\frac{\mathcal{D}_1}{+A \vee B} \quad \frac{\frac{\mathcal{D}_2}{-A \vee B}}{-A}}{+B} \quad \frac{\mathcal{D}_2}{-A \vee B}}{-B} \text{ (Re)}{\perp}$$

#### 4.4.3 Reducing (SR) to Atoms

**Negation:** (the remaining case is analogous)

$$\frac{\frac{[\neg A]^1}{+\neg A} \quad \mathcal{D}}{\perp} \text{ (SR}_1\text{)}^1 \quad \rightsquigarrow \quad \frac{\frac{[\neg A]^1}{+\neg A} \quad \mathcal{D}}{\perp} \text{ (SR}_2\text{)}^1}{-\neg A}$$



#### 4.4.4 Ancillary Reductions

Ancillary reductions eliminate formula occurrences that are consequences of *Smilean reductio* and premises of *Rejection*. Because of the way the normalisation process takes place, we only need to consider them for atomic formulas. In what follows, then,  $\alpha$  ranges over arbitrary atoms and  $\bar{\alpha}$  denotes the conjugate of  $\alpha$ .<sup>16</sup> There are three cases to consider:

**Case 1:** One of the premises of *Rejection* is not the conclusion of *Smilean Reductio*. Suppose (without loss of generality) that it is the left one:

$$\frac{\mathcal{D}_1 \quad \frac{[\alpha]^1 \quad \mathcal{D}_2}{\perp} \text{ (SR}_{1/2})^1}{\alpha \quad \frac{\perp}{\bar{\alpha}} \text{ (Re)}} \quad \rightsquigarrow \quad \frac{\mathcal{D}_1 \quad \alpha}{\mathcal{D}_2} \quad \perp$$

Note that since  $\alpha$  is an atom, this eliminates the ancillary maximal formula in question while creating no new maximal formulas of any kind.

**Case 2:** Both premises of *Rejection* are the conclusion of *Smilean Reductio*, and at least one of the applications of *Smilean Reductio* discharges no premises of *Rejection*. Say (without loss of generality) that the rightmost application is of this type:

$$\frac{\mathcal{D}_1 \quad \frac{[\alpha]^1 \quad \mathcal{D}_2}{\perp} \text{ (SR)}^1}{\alpha \quad \frac{\perp}{\bar{\alpha}} \text{ (Re)}} \quad \rightsquigarrow \quad \frac{\mathcal{D}_1 \quad \alpha}{\mathcal{D}_2} \quad \perp$$

Note that this also reduces the number of ancillary maximal formulas without giving creating new maximal formulas of any kind.

**Case 3:** Both premises of *Rejection* are conclusions of *Smilean Reductio* that discharge some premise of *Rejection*.

---

<sup>16</sup>Recall that the conjugate of  $+A$  is  $-A$  and vice versa.

$$\begin{array}{ccc}
\frac{[\alpha]^1}{\perp} & \frac{\mathcal{D}_0}{\bar{\alpha}} & \\
\perp & & \\
\mathcal{D}_1 & & \\
\frac{\perp}{\bar{\alpha}} \text{ (SR)}^1 & \frac{\mathcal{D}_2}{\alpha} \text{ (Re)} & \\
\perp & & \\
\hline
& & \rightsquigarrow
\end{array}
\qquad
\begin{array}{ccc}
& \frac{\mathcal{D}_2}{\alpha} & \frac{\mathcal{D}_0}{\bar{\alpha}} \text{ (Re)} \\
& \perp & \\
& \mathcal{D}_1 & \\
& \perp &
\end{array}$$

Suppose we apply this reduction to an ancillary maximal formula such that there are no ancillary maximal formulas above it or above a formula side connected with it. In the original derivation there may be further occurrences of  $\alpha$  with discharge label 1 besides the one explicitly represented above. We also replace them with a copy of  $\mathcal{D}_2$  ending in  $\alpha$ . Those occurrences that were not premises of *Rejection* are unproblematic. Those that were, though, become new ancillary maximal formulas of the same complexity. By assumption they are ancillary maximal formulas of the type covered in **Case 1**. We eliminate them as part of the current reduction step, and as a result the number of ancillary maximal formulas decreases, and we create no maximal formulas of other kinds.

#### 4.4.5 Normalisation and Corollaries

**Theorem 4.6** (Normalisation). *If there is a derivation  $\mathcal{D}$  of  $\varphi$  from  $\Gamma$  then there is a normal derivation  $\mathcal{D}'$  of  $\varphi$  from  $\Gamma' \subseteq \Gamma$ .*

*Proof.* To each derivation  $\mathcal{D}$  we assign a coordination rank  $(n, m) \in \mathbb{N} \times \mathbb{N}$ , where  $n$  is the highest complexity of a maximal coordination formula, and  $m$  the number of maximal coordination formulas of maximal complexity. A derivation without maximal coordination formulas has rank  $(0, 0)$ , and coordination ranks are ordered lexicographically. Each derivation  $\mathcal{D}$  is also assigned an operational rank  $(j, k) \in \mathbb{N} \times \mathbb{N}$  defined analogously but with respect to maximal operational formulas, and we order operational ranks with their own lexicographical order. The following is an effective procedure to normalise derivations:

1. Take a maximal coordination formula of the highest complexity such that there are no maximal coordination formulas of the highest complexity above it or above a formula side-connected with it. Apply the appropriate reduction from Sections 4.2 and 4.3. As a result, the coordination rank strictly decreases. After a finite number of iterations, then, our derivation has coordination rank  $(0, 0)$ .
2. Take a maximal operational formula of the highest complexity such that there are no maximal operational formulas of the highest complexity above it or above a formula side-connected with it. Apply the appropriate reduction from Section 4.1. The operational rank strictly decreases, and the coordination rank stays  $(0, 0)$ . After a finite number of iterations, then, our derivation has coordination and operational ranks  $(0, 0)$ .

3. Take an ancillary maximal formula (note that it must be atomic) such that there are no ancillary maximal formulas above it or above a formula side-connected with it. Apply the appropriate reduction from Section 4.4. The number of ancillary maximal formulas goes down, and the coordination and operational ranks stay the same. After a finite number of steps, the derivation is in normal form.

□

From this theorem it follows that all the operational rules in  $\mathcal{B}$  preserve *Smilean Reductio* and *Rejection*; this is the first half of Observation 4.3.

**Definition 4.7. Branch**

A branch  $\pi$  in a derivation  $\mathcal{D}$  is a sequence  $\varphi_1, \dots, \varphi_n$  of occurrences of formulas or of  $\perp$  such that: **(i)**  $\varphi_1$  is a leaf (an assumption), discharged or not. **(ii)**  $\varphi_{i+1}$  stands immediately below  $\varphi_i$ . **(iii)**  $\varphi_n$  is either the conclusion of  $\mathcal{D}$  or the first formula occurrence in the sequence that is the minor premise of  $(+ \rightarrow E)$ ,  $(+\vee E)$  or  $(-\wedge E)$ .

**Lemma 4.8.** *Every formula in a derivation belongs to some branch.*

*Proof.* By induction on derivations.

□

The following theorem characterises the shape of normal derivations (see also Observation 4.10 for a comparison with Prawitz’s normal form).

**Theorem 4.9** (Shape of normal derivations). *Let  $\mathcal{D}$  be a normal derivation,  $\pi = \varphi_1, \dots, \varphi_n$  a branch in it. Then there is a minimum formula  $\varphi_i$  dividing  $\mathcal{D}$  into two (possibly empty) parts, the E part and the I-part, such that:*

- (i)** Each  $\varphi_j$  in the E-part (i.e.  $j < i$ ) is the major premise of an E-rule.
- (ii)** If  $i \neq n$  then  $\varphi_i$  is a premise of (*Rejection*) or an I-rule.
- (iii)** Each  $\varphi_k$  in the I-part (i.e.  $i < k$ ) is a premise of an I-rule, except  $\varphi_{i+1}$ , which may be a premise  $\perp$  of *Smilean Reductio*.

*Proof.* Let  $\pi = \varphi_1, \dots, \varphi_n$  be a branch in a normal derivation  $\mathcal{D}$ . Then in  $\pi$  there are (a) no applications of an E-rule after an I-rule, (b) no applications of *Smilean Reductio* after an I-rule, (c) no applications of *Rejection* after an I-rule and (d) no applications of an E-rule after *Smilean Reductio*. I will prove (a) as an example; (b)-(d) are proved analogously.

**No applications of an E-rule after an I-rule:** suppose for a contradiction that there are such applications, let  $\varphi_k$  be the first consequence of an E-rule applied after an I-rule, and consider  $\varphi_{k-1}$ . Since the derivation is normal,  $\varphi_{k-1}$  is not the consequence of an I-rule. It cannot be the consequence of *Smilean Reductio* or *Rejection* either, as then we couldn’t obtain  $\varphi_k$  from it through an E-rule. Therefore  $\varphi_{k-1}$  must be the consequence of an E-rule, contradicting the assumption that  $\varphi_k$  was first.

Once (a)-(d) are secured the rest of the theorem follows easily. Consider the last rule applied in  $\pi$ : if it is an E-rule, let  $\varphi_i = \varphi_n$ . If it is *Smilean Reductio*,

let  $\varphi_i = \varphi_{n-2}$ . If it is *Rejection*, let  $\varphi_i = \varphi_1$ . Finally, if the last rule is an I-rule, let  $\varphi_i$  be the only formula occurrence in  $\pi$  that is a premise of *Rejection* –if there is one–, or else let  $\varphi_i$  be the first premise of an I-rule.  $\square$

**Observation 4.10.** An alternative way of phrasing Theorem 4.9 is to say that a branch in a normal derivation consists of three (possibly empty) parts: an E-part, where every formula occurrence is the major premise of an E-rule, a C-part, where every formula occurrence is atomic and a premise of a coordination principle, and an I-part, where every formula occurrence is a premise of an I-rule. Branches in Prawitz’s classical normal derivations (see Prawitz 1965) consist of a (possibly empty) E-part and a (possibly empty) I-part, joined together by a (possibly empty) part where classical reductio is applied to an atom.

We can now obtain the subformula and separation properties as corollaries.

**Definition 4.11. Subformula**

Signed formula  $\psi$  is a subformula of signed formula  $\varphi$  if the unsigned part of  $\psi$  is a subformula (in the standard sense) of the unsigned part of  $\varphi$ . For example, all of  $+p$ ,  $-p$ ,  $+q$ ,  $-q$  are signed subformulas of  $+p \rightarrow q$ . Note that  $\perp$  is not a formula but a punctuation sign.

**Definition 4.12. Order of a branch** A branch  $\pi = \varphi_1, \dots, \varphi_n$  in a derivation  $\mathcal{D}$  is of order 0 if  $\varphi_n$  is the conclusion of  $\mathcal{D}$ , and of order  $k + 1$  if it ends on the minor premise of an E-rule the major premise of which belongs to a branch  $\pi'$  order  $k$ .

**Corollary 4.13** (Subformula property). *All the formulas that occur in a normal derivation of  $\varphi$  from  $\Gamma$  are subformulas of some  $\gamma \in \Gamma$  or of  $\varphi$ .*

*Proof.* By induction on the order of branches. Let  $\pi = \varphi_1, \dots, \varphi_n$  be a branch of order  $k$  and assume the result for branches of order  $j < k$ . We will think of the E, C and I-parts of a branch as defined in Observation 4.10. The result is obvious for the I-part: if  $k = 0$  all formulas in it are subformulas of  $\varphi_n = \varphi$ . Similarly, if  $k > 0$  then all formulas in the I-part are subformulas of  $\varphi_n$ , which is in its turn a subformula of the major premise  $\psi$  of an elimination rule that belongs to a branch of lower order. By inductive hypothesis  $\psi$  is itself subformula of some  $\gamma \in \Gamma$  or of  $\varphi$ , and therefore so are all the formulas in the I-part.

It remains to show the result for the E and C-parts. Note that all remaining formulas are subformulas of  $\varphi_1$ , the first formula of the branch. Now, if  $\varphi_1$  is an undischarged assumption the result follows trivially. If  $\varphi_1$  is a discharged assumption, there are two cases to consider:

**Case 1:** If  $\varphi_1$  is discharged by *Smilean Reductio* then  $\varphi_1$  must be an atom, and so the E-part of our branch  $\pi$  is empty. Moreover, the application of *Smilean Reductio* in question concludes  $\bar{\varphi}_1$ , the conjugate of  $\varphi_1$ . Note that  $\varphi_1$  is a subformula of  $\bar{\varphi}_1$ , and that  $\bar{\varphi}_1$  must be a subformula of  $\varphi_n$ , the last formula of the branch. Therefore,  $\varphi_1$  is a subformula of  $\varphi_n$ . If the branch  $\pi$  is of order 0 this means that  $\varphi_1$  is a subformula of the conclusion, and if  $\pi$  is of order  $> 0$  then the result follows by inductive hypothesis.

**Case 2:** If  $\varphi_1$  is discharged by an I-rule, then it is a subformula of the consequence  $\varphi_k$  of that application, and  $\varphi_k$  is in its turn a subformula of  $\varphi_n$ , the last formula in the branch. Once again, if the branch  $\pi$  is of order 0 this means that  $\varphi_1$  is a subformula of the conclusion, and if  $\pi$  is of order  $> 0$  then the result follows by inductive hypothesis.  $\square$

**Corollary 4.14** (Separation property). *In a normal derivation of  $\varphi$  from  $\Gamma$  only operational rules for connectives in  $\varphi$  and  $\Gamma$  (and perhaps coordination principles) are used.*

*Proof.* Follows immediately from Corollary 4.13.  $\square$

## 4.5 Harmony and Normalisation: $\mathcal{HB}_2$

The first two clauses of the definition of normal form for  $\mathcal{HB}_2$  are identical to those of  $\mathcal{HB}_1$ . In other words, in all normal  $\mathcal{HB}_2$  derivations:

- (i) No conclusion of an I-rule is a major premise of an E-rule.
- (ii) Coordination principles are applied only to atoms.

The motivation behind them remains the same: (i) is taken from Prawitz, and (ii) is suggested by the notion of bilateral harmony from Section 3. Just as in the case of  $\mathcal{HB}_1$ , this is enough to ensure that normal derivations satisfy the separation property, but not enough to obtain the subformula property, as the following derivation shows:

$$\frac{\frac{+p \quad -p}{+q} \text{ (Ex)} \quad \frac{+p \quad -p}{-q} \text{ (Ex)}}{+r} \text{ (Ex)}$$

In order to make sure that normal derivations have the subformula property we will follow the same strategy as before, that is, imposing constraints on the way coordination principles interact with each other. These constraints are given by clause (iii) of Definition 4.15.

### Definition 4.15. Normal form

A derivation in  $\mathcal{HB}_2$  is in normal form if in it: (i) No conclusion of an I-rule is a major premise of an E-rule. (ii) Coordination principles are applied only to atoms. (iii) (a) No conclusion of Ex) is a premise of (Ex), (b) no application of (Ex) has both premises discharged by (Bem) and, (c) no conclusion of (Bem) is a premise of (Ex).

Formula occurrences that infringe clauses (i) and (ii) are called **maximal operational formulas** and **maximal coordination formulas**, respectively.

Formula occurrences that infringe clause (iii) are called **ancillary maximal formulas**.

$\mathcal{HB}_1$  and  $\mathcal{HB}_2$  share operational rules, so the reduction steps for operational maximal formulas are as before. The obvious similarity between *Rejection* and *Explosion* means that the reduction steps which restrict (Ex) to atomic premises are identical to the reductions steps which restrict (Re) to atomic premises; I will omit them as well. The remaining reduction steps are as follows.

#### 4.5.1 Reducing Conclusions to Atoms in (Ex)

**Negation:**

$$\frac{\mathcal{D}_1 \quad \mathcal{D}_2}{\frac{+A \quad -A}{+\neg B} \text{ (Ex)}} \rightsquigarrow \frac{\mathcal{D}_1 \quad \mathcal{D}_2}{\frac{+A \quad -A}{-B} \text{ (Ex)}} \frac{-B}{+\neg B} \text{ (-}\neg \text{I)}$$

$$\frac{\mathcal{D}_1 \quad \mathcal{D}_2}{\frac{+A \quad -A}{-\neg B} \text{ (Ex)}} \rightsquigarrow \frac{\mathcal{D}_1 \quad \mathcal{D}_2}{\frac{+A \quad -A}{+B} \text{ (Ex)}} \frac{-B}{-\neg B} \text{ (-}\neg \text{I)}$$

**Implication:**

$$\frac{\mathcal{D}_1 \quad \mathcal{D}_2}{\frac{+A \quad -A}{+B \rightarrow C} \text{ (Ex)}} \rightsquigarrow \frac{\mathcal{D}_1 \quad \mathcal{D}_2}{\frac{+A \quad -A}{+C} \text{ (Ex)}} \frac{+C}{+B \rightarrow C} \text{ (+}\rightarrow \text{I)}^v$$

$$\frac{\mathcal{D}_1 \quad \mathcal{D}_2}{\frac{+A \quad -A}{-B \rightarrow C} \text{ (Ex)}} \rightsquigarrow \frac{\frac{\mathcal{D}_1 \quad \mathcal{D}_2}{\frac{+A \quad -A}{+B} \text{ (Ex)}} \quad \frac{\mathcal{D}_1 \quad \mathcal{D}_2}{\frac{+A \quad -A}{-C} \text{ (Ex)}}}{-B \rightarrow C} \text{ (-}\rightarrow \text{I)}$$

**Disjunction:**

$$\frac{\mathcal{D}_1 \quad \mathcal{D}_2}{\frac{+A \quad -A}{+B \vee C} \text{ (Ex)}} \rightsquigarrow \frac{\mathcal{D}_1 \quad \mathcal{D}_2}{\frac{+A \quad -A}{+C} \text{ (Ex)}} \frac{+C}{+B \vee C} \text{ (+}\vee \text{I)}^v$$

$$\frac{\mathcal{D}_1 \quad \mathcal{D}_2}{\frac{+A \quad -A}{-B \vee C} \text{ (Ex)}} \rightsquigarrow \frac{\frac{\mathcal{D}_1 \quad \mathcal{D}_2}{+A \quad -A} \text{ (ex)} \quad \frac{\mathcal{D}_1 \quad \mathcal{D}_2}{+A \quad -A} \text{ (Ex)}}{\frac{-B \quad -C}{-B \vee C} \text{ (-}\vee \text{ I)}}$$

#### 4.5.2 Reducing Assumptions to Atoms in (Bem)

**Negation:**

$$\frac{\frac{[+\neg A]^1 \quad \mathcal{D}_1}{\varphi} \quad \frac{[-\neg A]^1 \quad \mathcal{D}_2}{\varphi} \text{ (Bem)}^1}{\varphi} \rightsquigarrow \frac{\frac{[-A]^1}{+\neg A} \quad \mathcal{D}_1}{\varphi} \quad \frac{[+A]^1}{-\neg A} \quad \mathcal{D}_2}{\varphi} \text{ (Bem)}^1$$

**Implication:**

$$\frac{\frac{[+A \rightarrow B]^1 \quad \mathcal{D}_1}{\varphi} \quad \frac{[-A \rightarrow B]^1 \quad \mathcal{D}_2}{\varphi} \text{ (Bem)}^1}{\varphi} \rightsquigarrow \frac{\frac{[+B]^3}{+A \rightarrow B} \quad \frac{\frac{[-B]^3}{-A \rightarrow B} \quad [A]^2}{\mathcal{D}_2} \quad \frac{\frac{[-A]^2}{+B} \quad [A]^1}{+A \rightarrow B} \text{ (Ex)}}{\varphi} \quad \frac{\mathcal{D}_1}{\varphi} \text{ (Bem)}^2}{\varphi} \text{ (Bem)}^3$$

**Disjunction:**

$$\frac{\frac{[+A \vee B]^1 \quad \mathcal{D}_1}{\varphi} \quad \frac{[-A \vee B]^1 \quad \mathcal{D}_2}{\varphi} \text{ (Bem)}^1}{\varphi} \rightsquigarrow \frac{\frac{[+B]^3}{+A \vee B} \quad \frac{\frac{[-B]^3}{-A \vee B} \quad [-A]^2}{\mathcal{D}_2} \quad \frac{\frac{[-A]^1}{+B} \quad [A]^2}{+A \vee B} \text{ (Ex)}}{\varphi} \quad \frac{\mathcal{D}_1}{\varphi} \text{ (Bem)}^2}{\varphi} \text{ (Bem)}^3$$

#### 4.5.3 Reducing Conclusions to Atoms in (Bem)

**Negation:** (the case where  $\varphi = -\neg B$  is analogous)

$$\frac{\begin{array}{c} [+A]^1 \quad [-A]^1 \\ \mathcal{D}_1 \quad \mathcal{D}_2 \\ +\neg B \quad +\neg B \end{array}}{+\neg B} (\text{Bem})^1 \quad \rightsquigarrow \quad \frac{\begin{array}{c} [+A]^1 \quad [-A]^1 \\ \mathcal{D}_1 \quad \mathcal{D}_2 \\ +\neg B \quad +\neg B \\ \hline -B \quad -B \\ \hline -B \\ \hline +\neg B \end{array}}{+\neg B} (\text{Bem})^1$$

**Implication:**

$$\frac{\begin{array}{c} [+A]^1 \quad [-A]^1 \\ \mathcal{D}_1 \quad \mathcal{D}_2 \\ +B \rightarrow C \quad +B \rightarrow C \end{array}}{+B \rightarrow C} (\text{Bem})^1 \quad \rightsquigarrow \quad \frac{\begin{array}{c} [+A]^1 \quad [-A]^1 \\ \mathcal{D}_1 \quad \mathcal{D}_2 \\ +B \rightarrow C \quad +B \rightarrow C \\ \hline +C \quad +C \\ \hline +C \\ \hline +B \rightarrow C \end{array}}{+B \rightarrow C} (\text{Bem})^1 \quad (+ \rightarrow \text{I})^2$$

$$\frac{\begin{array}{c} [+A]^1 \quad [-A]^1 \\ \mathcal{D}_1 \quad \mathcal{D}_2 \\ -B \rightarrow C \quad -B \rightarrow C \end{array}}{-B \rightarrow C} (\text{Bem})^1 \quad \rightsquigarrow \quad \frac{\begin{array}{c} [+A]^1 \quad [-A]^1 \quad [+A]^2 \quad [-A]^2 \\ \mathcal{D}_1 \quad \mathcal{D}_2 \quad \mathcal{D}_1 \quad \mathcal{D}_2 \\ -B \rightarrow C \quad -B \rightarrow C \quad -B \rightarrow C \quad -B \rightarrow C \\ \hline +B \quad +B \quad -C \quad -C \\ \hline +B \quad -C \\ \hline -B \rightarrow C \end{array}}{-B \rightarrow C} (\text{Bem})^1 \quad (\text{Bem})^2$$

**Disjunction:** Analogous to implication.

#### 4.5.4 Ancillary reductions

As before, the  $\alpha_i$  range over arbitrary *atoms* and  $\bar{\alpha}_i$  is the conjugate of  $\alpha_i$ .

**Clause (iii)(a):**

$$\frac{\begin{array}{c} \mathcal{D}_1 \quad \mathcal{D}_2 \\ +p \quad -p \\ \hline \alpha_1 \end{array} (\text{Ex}) \quad \frac{\mathcal{D}_3}{\bar{\alpha}_1} (\text{Ex})}{\alpha_2} \quad \rightsquigarrow \quad \frac{\begin{array}{c} \mathcal{D}_1 \quad \mathcal{D}_2 \\ +p \quad -p \\ \hline \alpha_2 \end{array} (\text{Ex})$$

**Clause (iii)(b):**

$$\begin{array}{c}
\frac{[\alpha_1]^1 \quad [\bar{\alpha}_1]^n}{\alpha_2} \text{ (Ex)} \quad \frac{[\bar{\alpha}_1]^1}{\mathcal{D}_2} \\
\mathcal{D}_1 \quad \alpha_3 \text{ (Bem)}^1 \quad [\alpha_1]^n \\
\frac{\alpha_3}{\mathcal{D}_3} \quad \frac{\alpha_4}{\mathcal{D}_4} \text{ (Bem)}^n \\
\frac{\alpha_4}{\alpha_4} \text{ (Bem)}^n \rightsquigarrow \frac{[\bar{\alpha}_1]^1}{\mathcal{D}_2} \quad \frac{[\alpha_1]^1}{\mathcal{D}_4} \\
\frac{\alpha_4}{\alpha_4} \text{ (Bem)}^1
\end{array}$$

Note that in this last reduction we have assumed that the left premise of the application of (Ex) is discharged before the right one. This doesn't matter; if it is the other way around, the appropriate reduction is analogous.

**Clause (iii)(c):**

$$\begin{array}{c}
\frac{[+p]^1 \quad [-p]^1}{\mathcal{D}_1 \quad \mathcal{D}_2} \quad \frac{\mathcal{D}_3}{\alpha_1} \text{ (bem)}^1 \quad \frac{\mathcal{D}_3}{\bar{\alpha}_1} \\
\frac{\alpha_1}{\text{(ex)}} \quad \frac{\alpha_1}{\alpha_2} \text{ (bem)}^1 \rightsquigarrow \frac{[+p]^1 \quad \mathcal{D}_3}{\alpha_1} \quad \frac{[-p]^1 \quad \mathcal{D}_3}{\bar{\alpha}_1} \text{ (ex)} \\
\frac{\alpha_2}{\alpha_2} \text{ (bem)}^1
\end{array}$$

Applications of (ex) like the one above on the left, where at least one of the premises is a conclusion of (Bem), are called **peaks**. The **size** of a peak is the sum of the length of the maximal segments that the premises of (Ex) belong to (if a premise is not part of a maximal segment, we assign it length 0).<sup>17</sup> In the normalisation process we will assign to each derivation a **peak rank**  $(j, k)$ , where  $j$  is the greatest size of a peak in the derivation, and  $k$  the number of peaks of greatest size. It is easy to check that the reduction above, when applied to a maximal segment such that there are no longer maximal segments above it, side connected with it, or above a formula side connected with it, strictly reduces the peak rank of a derivation.

#### 4.5.5 Normalisation and corollaries

**Theorem 4.16** (Normalisation). *If there is a derivation  $\mathcal{D}$  of  $\varphi$  from  $\Gamma$  then there is a normal derivation  $\mathcal{D}'$  of  $\varphi$  from  $\Gamma' \subseteq \Gamma$ .*

*Proof.* Analogous to the previous proof of normalisation. Derivations are assigned a coordination and an operational rank, defined as before. We apply

<sup>17</sup>Segments are defined in Def. 4.17 below.

coordination reductions first (Sections 5.1-5.3) and then operational reductions (Section 4.1), starting always from maximal formulas of maximal complexity such that there are no maximal formulas of maximal complexity above them or above a formula side connected with them. Once a derivation has no coordination or operational maximal formulas, we assign it a peak rank, as defined at the end of Section 5.4, and apply the reduction for ancillary formulas of type **(iii)**(c) as indicated there. Once there are no peaks left, the only remaining maximal formulas are those that infringe clauses **(iii)**(a) and **(iii)**(b). They can be eliminated in any order using the appropriate reductions from Section 5.4.  $\square$

From this theorem it follows that all the operational rules in  $\mathcal{B}$  preserve *Bilateral Excluded Middle* and *Explosion*; this is the second half of Observation 4.3.

**Definition 4.17. Segment** A segment  $\sigma$  in a branch  $\pi$  is a sequence of formula occurrences  $\sigma_1, \dots, \sigma_n$  in  $\pi$  such that: **(i)**  $\sigma_1$  is *not* the conclusion of an application of (Bem). **(ii)** Each  $\sigma_i$  for  $i < n$  is a premise of (Bem), and  $\sigma_{i+1}$  stands immediately below  $\sigma_i$ . **(iii)**  $\sigma_n$  is *not* a premise of an application of (Bem).

Definition 4.17 entails that all the elements of a segment are occurrences of the same formula. The **length of a segment** is the number of formula occurrences in it. A segment is called **maximal** if it ends in an application of (Ex). This means that maximal coordination formulas that violate clause **(iii)**(c) of Definition 4.15 are always final formula occurrences in maximal segments of length  $\geq 1$ , and maximal coordination formulas that violate clauses **(iii)**(a) and **(iii)**(b) are always maximal segments of length 1. There are no maximal segments of other types.

**Lemma 4.18.** *Every branch can be uniquely divided into consecutive segments.*

*Proof.* By induction on the length of branches.  $\square$

**Theorem 4.19** (Shape of normal derivations). *Let  $\mathcal{D}$  be a derivation in normal form,  $\pi$  a branch in  $\mathcal{D}$ , and let  $\sigma_1, \dots, \sigma_n$  be the segments in  $\pi$ . Then there is a segment  $\sigma_i$  in  $\pi$ , called the *minimum segment*, which separates two (possibly empty) parts, the *E-part* and the *I-part*, with the properties:*

1. *For each  $\sigma_j$  in the E-part (i.e.  $j < i$ ),  $\sigma_j$  is a major premise of an E-rule, except possibly  $\sigma_{i-1}$ , which may be a premise of (ex).*
2. *If  $i \neq n$ , then each formula in the segment  $\sigma_i$  is a premise of (bem) except the last one, which may be a premise of an I-rule.*
3. *For each  $\sigma_j$  in the I-part (i.e.  $i < j < n$ ),  $\sigma_j$  is a premise of an I-rule.*

*Proof.* It is easy to see that, in a branch  $\pi = \varphi_1, \dots, \varphi_n$  of a normal derivation, no formula occurrences that are premises of an Introduction rule precede formula occurrences that are major premises of an Elimination, (Bem) or (Ex).

Similarly, no formula occurrences that are premises of (Bem) precede formula occurrences that are premises of (Ex) or major premises of an E-rule, and no formula occurrences that are premises of (Ex) precede formula occurrences that are major premises of an E-rule or (Ex). Now:

If there is no formula occurrence that is a premise of an I-rule or (Bem), let  $\sigma_i = \varphi_n$ . If there is a formula occurrence that is a premise (Bem), let  $\varphi_i$  be the first such formula, and let  $\sigma_i$  be the segment starting at  $\varphi_i$ . Finally, if there is no formula occurrence that is a premise (Bem), but there is a formula occurrence that is a premise of an I-rule, let  $\varphi_i$  be the first such formula, and let  $\sigma_i = \varphi_i$ .  $\square$

**Observation 4.20.** An alternative way of phrasing Theorem 4.19 is to say that a branch in a normal derivation consists of three (possibly empty) parts: an E-part, where every formula occurrence is the major premise of an E-rule, a C-part, where every formula occurrence is a premise of a coordination principle –and within which (Ex) is applied before (Bem)– and an I-part, where every formula occurrence is a premise of an I-rule.

**Corollary 4.21** (Subformula property). *All the formulas that occur in a normal derivation of  $\varphi$  from  $\Gamma$  are subformulas of some  $\gamma \in \Gamma$  or of  $\varphi$ .*

*Proof.* By induction on the order of branches. Let  $\pi = \sigma_1, \dots, \sigma_n$  be a branch of order  $p$ , let  $\sigma_i$  be its minimum segment, and assume the result for branches of lower order. Consider first all  $\sigma_j$  with  $i \leq j \leq n$ . All such formulae are subformulas of  $\varphi_n$ , the formula in the last segment  $\sigma_n$  of the branch. If the branch in question is of order 0 the result immediately follows. If the branch is of order  $> 0$  then  $\varphi_n$  is the minor premise of an application of an E-rule, the major premise  $\psi$  of which belongs to a branch of order  $p - 1$ . But by induction hypothesis the result holds for  $\psi$ , and  $\varphi_n$  is a subformula of  $\psi$ , so the result follows.

It remains to account for the  $\sigma_j$  with  $j < i$ . Now, all such formulae are subformulas of  $\varphi_1$ , the first formula of the branch. If  $\varphi_1$  is an undischarged assumption, the result immediately follows. Similarly, if  $\varphi_1$  is discharged by an application of an I-rule, then it is a subformula of some formula in an I-part, and the result follows by the above. Finally, suppose that  $\varphi_1$  is discharged by an application of (Bem).  $\varphi_1$  cannot be the major premise of an elimination rule, since it is an atom. If it is the minor premise of an E-rule, or a premise of an I-rule or (Bem), then there are no  $\sigma_j$  with  $j < i$  and we are done. The only remaining possibility is that  $\varphi_1$  is a premise of (Ex). Then  $\varphi_1$  is the only formula before the minimum segment  $\sigma_i$  (in other words,  $\varphi_1$  is the only formula we still need to account for).  $\varphi_1$  is a subformula of the other premise  $\bar{\varphi}_1$  of the application of (Ex) in question, and  $\bar{\varphi}_1$  cannot be discharged by (Bem). Moreover,  $\bar{\varphi}_1$  belongs to a branch of the same order as  $\pi$ . If  $\bar{\varphi}_1$  is undischarged, or discharged by a I-rule, the result immediately follows. If it is a consequence of an E-rule, then it is a subformula of the initial formula  $\psi$  of its branch. But then  $\psi$  is not atomic, and so can only be undischarged or discharged by an I-rule. In either case, the result follows.  $\square$

**Corollary 4.22** (Separation property). *In a normal derivation of  $\varphi$  from  $\Gamma$  only operational rules for connectives in  $\varphi$  and  $\Gamma$  (and perhaps coordination principles) are applied.*

*Proof.* Follows immediately from the previous corollary. □

## Appendix I

This Appendix shows that the rules for *bink* and *blink* don't preserve *Smilean Reductio* and *Rejection*, using the Kripke models introduced in (Ferreira 2008). An **F-Model** is a tuple  $M = (W, R, v_+, v_-)$  where  $(W, R)$  is a non-empty poset and  $v_+, v_-$  are valuation functions from the set of propositional atoms to  $\mathcal{P}(W)$  such that:

- (i)  $\forall w, w' \in W, wRw' \text{ and } w \in v_+(p) \implies w' \in v_+(p)$ .
- (ii)  $\forall w, w' \in W, wRw' \text{ and } w \in v_-(p) \implies w' \in v_-(p)$ .
- (iii)  $v_-(p) \cap v_+(p) = \emptyset$  for all atoms  $p$ .

**Satisfaction** in an F-model is defined as follows (see Ferreira 2008, p. 1055 for the satisfaction conditions of the Boolean connectives):

$$M, w \models_+ p \text{ iff } w \in v_+(p).$$

$$M, w \models_- p \text{ iff } w \in v_-(p).$$

$$M, w \models_+ \text{bink } A \text{ iff } M, w \models_+ A \text{ and } M, w \models_- A.$$

$$M, w \models_- \text{bink } A \text{ iff } M, w \models_- A.$$

We use  $M, w \models +A$  as shorthand for  $M, w \models_+ A$  and  $M, w \models -A$  as a shorthand for  $M, w \models_- A$ . We also stipulate that  $M, w \not\models \perp$ .

Now, it is obvious that the rules for *bink* are sound for the class of F-models. To see that they don't preserve *Smilean Reductio*, assume for a contradiction that they do and consider the F-Model  $\mathcal{M}$  with  $W = \{w_0, w_1, w_2\}$ ,  $R = \{(w_0, w_0), (w_0, w_1), (w_1, w_1), (w_0, w_2), (w_2, w_2)\}$  and where  $v_+(p) = \{w_1\}$  for all atoms  $p$ ,  $v_-(q) = \{w_2\}$  for all atoms  $q$ . As shown by Ferreira (2008, p. 1056)  $\mathcal{M}$  satisfies *Smileian Reductio* for atoms. By assumption and the soundness of *bink*, it follows that  $\mathcal{M}$  satisfies *Smileian Reductio* for all sentences. Using *Smileian Reductio* and the rules for *bink*, as we saw in Section 4.2, we have  $\vdash -A$  for any  $A$ . By soundness and because  $\mathcal{M}$  satisfies (SR) it follows that  $\mathcal{M}, w \models -p$  for all  $w$ . But by construction,  $\mathcal{M}, w_1 \not\models -p$ .

The proof that the rules for *blink* don't preserve *Rejection* is analogous.

## Chapter 5

# Logic, Translation, and Minimalism

While a stretto diminution  
is an obvious solution

---

Glenn Gould, *So You Want to  
Write a Fugue*

In this chapter we will examine the Meaning-Variance Problem. The best way to introduce this problem is to consider different ways in which philosophers disagree about logic. They disagree, first of all, about how it constrains the way we think and argue. They disagree about how we come to know logical truths, and about the relation between logic and mathematics. And perhaps perhaps more surprisingly, they also disagree about which arguments are logically valid. Let's focus on this last type of disagreement.

According to a popular idea, partisans of different logics are in fact 'partisans of different languages'. Whenever two logics validate different arguments, the thought goes, they use different logical vocabulary; we will call this the Meaning-Variance claim. It is often thought that, if the Meaning-Variance claim is true, then there can be no 'real' or 'substantial' disagreement between parties who disagree about the validity of arguments: they use different languages, so they are *merely* talking past each other. The goal of this chapter is to examine arguments for and against the Meaning-Variance claim, and to clarify its consequences.

The best-known defence of the Meaning-Variance thesis comes from Quine (1954, 1960, 1986), and we will begin by looking at his work. Unfortunately, philosophers *also* disagree about what Quine's case for the Meaning-Variance claim consists in. Here I will focus on the most plausible interpretations: the Translation Argument, discussed in Section 1, and the Use Argument, discussed in Section 2.

This separation is somewhat artificial; in Quine's work the arguments are

intermingled, and Quine jumps from one to the other without notice. Still, this way of organizing what Quine says can help us make better sense of his position. Ultimately, in any case, my goal is not to weigh in on what the right interpretation of Quine is, but to assess different arguments for Meaning-Variance.

A common response to the Meaning-Variance claim, going back to Putnam (1957) and Morton (1973), has it that sides who agree on ‘enough’ logical principles use the same logical vocabulary. The obvious difficulty with this idea is spelling out what counts as ‘enough agreement’. Morton and the later Putnam (see his 1969) thought no precise criterion could be given. More recently, though, some inferentialists have tried to provide one. They are often called *minimalists*, following Hjortland’s (2014) terminology. In Sections 3-6 I will present minimalist proposals by Francesco Paoli (2003, 2014), Greg Restall (2002, 2014), and Bogdan Dicher (2016), and argue that they should be rejected. Finally, in Section 7 I will take stock and reassess the importance of the Meaning-Variance claim.

## 5.1 Logic and Translation

Let’s start with a birds-eye view of Quine’s position. Quine thought that partisans of different logics use different logical vocabulary. He also thought that the ordinary meaning of logical terms is captured by classical logic. Therefore, he concluded, non-classical logicians simply “change the subject” (Quine 1986, p. 81), that is, they speak novel languages.

Quine typically put his point in terms of translation, or rather, in terms of a loose sense of ‘translation’ that includes the interpretation of our native language. If a speaker endorses a non-classical logic, Quine said, no acceptable translation can map their logical terms to ours (where again, ‘we’ are allegedly classical logicians). He gave a relatively straightforward argument for this idea in *Philosophy of Logic* (1986); we will call it the Translation Argument. Very roughly put, the argument goes as follows: first, classical logical truths are obvious. But secondly, translation must ‘save the obvious’. Therefore, no acceptable translation identifies the classical connectives and quantifiers with those of non-classical logicians.

Our first job in this section is to flesh out the Translation Argument. We will start with the first premise, and in particular with the word ‘obvious’. Quine uses it as a term of art, stripped of the usual epistemological connotations:

I am using the word ‘obvious’ in a [...] behavioral sense, with no epistemological overtones. When I call ‘ $1 + 1 = 2$ ’ obvious to a community I mean only that everyone, nearly enough, will unhesitatingly assent to it, for whatever reason; and when I call ‘It is raining’ obvious in particular circumstances I mean that everyone will assent to it in those circumstances. (Quine 1986, p. 82)

Understood this way, the claim that all classical logical truths are obvious seems trivially false. Speakers aren't logically omniscient, and there is no guarantee that a long and complex classical logical truth will be immediately recognised as such. As a result, it need not be assented to at all, even by classical logicians. To accommodate these cases Quine speaks of *potential* obviousness:

Every [classical] logical truth is either obvious as it stands or can be reached from obvious truths by a sequence of individually obvious steps. To say this is in effect just to repeat [...] that the [classical] logic of quantification and identity admits of complete proof procedures, and some of these are procedures that generate sentences purely from visibly true sentences by steps that visibly preserve truth. (Quine 1986, pp. 82-83)

Quine doesn't spell out what an 'individually obvious step' is, but it's safe to assume he intends the metaphor about visible (preservation of) truth to be understood in behaviouristic terms, like his earlier gloss on obvious sentences.<sup>1</sup> With this in mind, we can take an inferential step to be obvious for a given community if it would ordinarily go unchallenged by its members.

Now, it would be easy to extend the notion of potential obviousness from sentences to inferential steps; for instance, we could say that an inferential step is potentially obvious if its conclusion can be obtained from its premises by means of obvious inferential steps. Quine, however, doesn't go as far. His point is just that there are enough obvious classical logical truths, and enough obvious classically valid inferential steps, to make all classical logical truths (potentially) obvious. Let's make this official:

**(Obv)** All classical logical truths are (potentially) obvious.

The second premise is harder to pin down. Quine insists that translation must 'save the obvious', but it is not entirely clear what this means. At a first pass, we can say that it consists in never translating our interlocutors in a way that makes them disagree with us over (potentially) obvious sentences:

**(Sto)** Correct translation never represents our interlocutors as disagreeing with us over (potentially) obvious sentences.

Quine's examples can help us put a finger on the sort of thing he had in mind. Suppose, for instance, that the speakers of a language so far unknown to us will not assent to a certain sentence in the rain. Then, Quine says:

We have reason not to translate the sentence as 'It is raining'. Naturally the native's unreadiness to assent to a certain sentence gives us reason not to construe the sentence as saying something whose truth should be obvious to the native at the time. (Quine 1986, p. 81)

---

<sup>1</sup>Recall also that Quine held a disquotational theory of truth, see e.g. (Quine 1986, Ch. 3).

The general maxim behind this seems to be that:<sup>2</sup>

(i) If an English sentence is (potentially) obvious, it must be translated by a foreign sentence that is assented to.

Something like the converse of (i) is also supposed to constrain translation:

It behooves us, in translating a strange language, to make the obvious sentences go over into English sentences that are true and, preferably, also obvious. (Quine 1986, p. 82)

Put in the obviousness-and-assent terminology of the rest of the argument, this is to say that:

(ii) If a foreign sentence is obvious, it must be translated by an English sentence that is assented to.

And lastly, to take another famous example, suppose that our fictional natives assent to a certain sentence ' $q$  ka bu  $q$ ' which would be straightforwardly translated by an English sentence of the form ' $q$  and not  $q$ '. Then:

Just how good a translation is this, and what may the lexicographer's method have been? If any evidence can count against a lexicographer's adoption of 'and' and 'not' as translations of 'ka' and 'bu', certainly the natives' acceptance of ' $q$  ka bu  $q$ ' [...] counts overwhelmingly. (Quine 1954, p. 352)

What matters in this case, as far as Quine is concerned, is not that sentences of the form ' $q$  and not  $q$ ' are logical falsehoods; it is that that they are *obvious* falsehoods. The general idea seems to be that:

(iii) If a foreign sentence is assented to, it must not be translated by an obviously false English sentence.<sup>3</sup>

(**Obv**) and (**Sto**) rule out interpreting our interlocutors as asserting classical logical falsehoods, and as refusing to assent to classical logical truths. Whenever a speaker seems to do either of those things, the maxims will prompt us to change the translation of the sentences the speaker has asserted or refused to assent to; we will have to 'translate away' the disagreement, so to speak.

The resulting account is subtler than it looks, and there are some niceties that we should keep in mind. First, Quine does not deny that some speakers endorse non-classical logics. What he rejects is a certain *picture* of the situation, according to which non-classical logicians have abandoned some classical

---

<sup>2</sup>Clauses (i)-(iii) below come from (Morton 1973).

<sup>3</sup>Some care must be put into what we mean by 'obviously false' here. If we don't mind begging some questions against dialethists, we can take an obviously false sentence to be one whose negation is (potentially) obvious. Alternatively, we can define obvious falsehood much like Quine defines obviousness: an obviously false sentence, in this sense, would be one from which 'everyone, nearly enough, would unhesitatingly *dissent*'.

principle but retained the classical logical vocabulary. Disagreement is ‘translated away’ only in the sense that by abandoning classical logic one abandons classical logical constants with it. In this sense, for example, the intuitionist:

should not be viewed as controverting us as to the true laws of certain fixed logical operations, namely, negation and alternation [disjunction]. He should be viewed rather as opposing our negation and alternation as unscientific ideas, and propounding certain other ideas, somewhat analogous, of his own. (Quine 1986, p. 87)

Secondly, there is no reason to assume that every aspect of classical (or any other) logic can be recovered under translation. Logical truth, Quine says, is guaranteed under translation “only in a negative sense”:

What is negative about this guarantee is that it does not assure that all our logically true sentences carry over into truths of the foreign language; some of them might resist translation altogether. (Quine 1986, p. 83)

And finally, whether asserting classical logical falsehoods and refusing to assent to classical logical truths covers all cases of deviation from classical logic depends on how we tell logics apart. For Quine logics were sets of theorems, and disagreement about logic was disagreement about logical truth (Quine 1954, p. 354, 1986, p. 80). It has now become the norm to use more fine-grained identity criteria: logics are typically seen as consequence relations between formulas or sequents.<sup>4</sup> As a result, distinct logics can agree in their respective sets of theorems, but validate different inferences or meta-inferences. If we want Quine’s argument to cover these cases, we need to understand **(Obv)** and **(Sto)** more broadly. In the case of **(Obv)** this is more or less straightforward. If we extend the notion of potential obviousness to inferences (say, along the lines I sketched above) we can strengthen **(Obv)** to:

**(Obv)<sup>+</sup>** All classical logical truths and valid inferences are (potentially) obvious.

In the case of **(Sto)** we would need to add further clauses along the lines of:

**(iv)** If the inferential step from  $\Gamma$  to  $A$  is (potentially) obvious, then the inferential step from the translation of  $\Gamma$  to the translation of  $A$  must be (potentially) obvious.

Whether these changes are faithful to Quine’s position is unclear. On the one hand, as I said, he officially understands disagreement between partisans of different logics as disagreement about logical truth. But on the other hand, he also turns to disagreement about logical consequence in some of his examples:

---

<sup>4</sup>See e.g. (Avron 1991) and (Ripley 2021) for discussion, and (Dicher and Paoli 2018) for a more abstract notion of a consequence relation.

[Paraconsistent logicians] think they are talking about negation, ‘ $\neg$ ’, ‘not’; but surely the notation ceased to be recognizable as negation when they took to regarding some conjunctions of the form  $p \wedge \neg p$  as true, *and stopped regarding such sentences as implying all others*. (Quine 1986, p. 81, my emphasis)

In what follows I will largely ignore this issue. What matters for our purposes is that Quine’s argument, or something like it, is compatible with fine-grained ways of individuating logics. In the spirit of good book-keeping, however, here is the general form of the Translation Argument: all classical logical truths and valid inferences are (potentially) obvious. Translation must not construe our interlocutors as disagreeing with us over (potentially) obvious matters — where this includes sentences and inferences. Therefore, whoever endorses a non-classical logic is best seen as using ‘non-classical vocabulary’.

On the face of it the Translation Argument only deals with the relation between classical and non-classical logics. However, it is easy to see that if the argument works at all, then it has consequences more or less across the board. **(Sto)** is presented as general constraint on translation. As long as some (not necessarily classical) logic is obvious for a community, it rules out all translations that identify its logical vocabulary with that of partisans of other logics. In this sense, then, Quine is committed to a general form of the Meaning-Variance claim.

We now have a better idea of what the Translation Argument looks like. The obvious next question is: how good an argument is it? As we will see, there is reason to be wary. Let’s begin with another look at **(Obv)**, the claim that all classical logical truths are (potentially) obvious. Here it will be useful to distinguish, following (Burgess 2005), between descriptive and prescriptive criticisms of classical logic. According to descriptive critics, classical logic endorses principles that are (a) invalid and (b) not operative in actual practice. According to prescriptive critics, on the other hand, classical logic correctly describes our standard inferential practice, but this practice is defective for one reason or other.

Now, descriptive critics of classical logic will just flatly deny **(Obv)**. Priest, for example, dismisses it as “mind-numbingly false” (2003, p. 462), and charges Quine with confusing pure and applied logic. The idea, roughly put, is that while it may be clear that a certain principle holds according to a given logic, it is often unclear whether it holds according to the (largely implicit) canons of inference of a certain community.

More generally, the fact that there are critics of classical logic at all may be a problem for **(Obv)**. Take once again (English-speaking) intuitionists, who refuse to assent to some sentences of the form ‘ $A$  or not  $A$ ’. If **(Obv)** is to be true, then they can’t be full-blown citizens of ‘our’ linguistic community. More precisely, whenever they refuse to assent to a sentence of the form ‘ $A$  or not  $A$ ’, we must assume that the words ‘or’ and ‘not’, as they use them, don’t correspond to our ‘or’ and ‘not’. But of course, that is just what the Translation Argument supposed to show. In this sense, **(Obv)** seems either false or question-begging.

Ultimately, though, it is hard to take any conclusive stance on **(Obv)**. According to Quine a sentence is obvious for a community if “everyone, *nearly enough*, will unhesitatingly assent to it” (Quine 1986, p. 82, my emphasis). Whether Quine’s argument works depends, to a large extent, on how we interpret this ‘nearly’. If it is understood quite broadly, then **(Obv)** may well be true—at least if we put descriptive criticisms of classical logic aside. At the same time, the broader this ‘nearly enough’ is, the more things will count as obvious, which makes **(Sto)** more and more implausible. The fact is that **(Obv)** is formulated with a lot of built in slack, while Quine’s conclusion isn’t. This, however, makes the Translation Argument both harder to refute and harder to endorse.

Something similar happens with the claim that translation must save the obvious. At times Quine presents it as a defeasible rule of thumb:

The maxim of translation underlying [**(Sto)**] is that assertions startlingly false on the face of them are *likely* to turn on hidden differences of language. [...] The *common sense* behind the maxim is that one’s interlocutor’s silliness, beyond a certain point, is *less likely than* bad translation or, in the domestic case, linguistic divergence. (Quine 1960, p. 59, my emphases)

If that is all there is to **(Sto)**, the Translation Argument can’t get off the ground. What follows from the argument, understood this way, is only that disagreements about logic are sometimes, or perhaps often, a sign that the parties involved use different logical vocabulary. This is more or less uncontroversial, but also clearly weaker than what Quine takes his argument to show:

The logics of two cultures will be, we saw, incommensurable at worst and *never* in conflict, since *conflict would simply discredit our translation*. The same attitude would apply, we noted, even to a logical deviant within our linguistic community: we would account his deviation a difference of dialect. (Quine 1986, p. 96, my emphasis)

A tempting way to argue for the full-fat version of **(Sto)**, understood to apply always and everywhere, is to appeal to the Principle of Charity. Quine does this only in passing (Quine 1960, p. 59), so it will be useful to draw on Davidson’s closely related work to fill in the gaps.<sup>5</sup>

According to Davidson our interpretation of other speakers’ utterances should be as charitable as possible, in two ways. First, we must make the interpreter right (by our standards) whenever this is plausible (Davidson 1974, p. 320). And secondly, we must read “a certain degree of logical consistency” in the speech

---

<sup>5</sup>There is an important difference between Quine and Davidson’s approaches. Quine looks at the topic of interpretation in terms of translation, and Davidson in terms of assignments of truth-conditions to sentences. In the rest of this section I will put aside Davidson’s talk of truth-conditions; my goal is to build the best possible argument for **(Sto)**, and not so much to analyse Davidson’s views.

and thought of our interlocutors (Davidson 2001b, p. 211). Both maxims fall under the label ‘Principle of Charity’.

Davidson’s expression “a certain degree of logical consistency” is bound to raise some questions. For example, what degree of logical consistency? And consistency according to which logic? Davidson himself seems undecided. At times he states that “we have no choice but to project our own logic” onto our interlocutors (Davidson 2004, p. 157). In other writings it is only “a modicum” (Davidson 2001b, p. 211) of logic that we project, or as much of it as is required to get a Tarskian theory of truth for the language we are interpreting (Davidson 1973, p. 323). And other times still, the goal is to “maximize the self-consistency we attribute to [the interpretee], on pain of not understanding *him*” (Davidson 1967, p. 313), which suggests that we should take our interlocutor’s position to be consistent—or at least, not obviously inconsistent—according to the standards we take *them* to have.

In what follows I will understand “a certain degree of logical consistency” in this last (and weakest) sense. This is for two reasons. The first is that if the Principle of Charity already involves reading our logic into other speakers, there is no point in arguing from Charity to **(Sto)**: whoever rejects **(Sto)** will automatically reject the Principle of Charity. The second reason is that, as I will argue below, a case can be made that even if we grant that some degree of charity, broadly understood, is a necessary requirement for interpretation, charity need not extend as far as taking our interlocutors to share our logic.

With this out of the way we can move on to the next important question: *why* should we adopt the Principle of Charity? Davidson insists that it is far from well-meaning assumption; it is, he says, a necessary requirement for interpretation. According to Davidson, unless we take our interlocutors to be ‘sufficiently right’ and ‘sufficiently coherent’ we will not be able to interpret their utterances at all (and more generally, we will not be able to make sense of their actions as rational, intentional behaviour):

The aim is not the absurd one of making disagreement and error disappear. The point is rather that widespread agreement is the only possible background against which disputes and mistakes can be interpreted. Making sense of the utterances and behaviours of others, even their most aberrant behaviour, requires us to find a great deal of reason and truth in them. (Davidson 1974, p. 321)

The general idea, then, is that being ‘too wrong’ is tantamount to being unintelligible:

If we cannot find a way to interpret the utterances and other behaviour of a creature as revealing a set of beliefs largely consistent and true by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything. (Davidson 1973, p. 193)

Ultimately these claims are grounded in Davidson’s views on the relation between the interpretation of speech, belief ascriptions, and belief. Davidson

begins by noting that we can't interpret another speaker's utterances unless we simultaneously ascribe them some appropriate beliefs:

Beliefs and meanings conspire to account for utterances. A speaker who holds a sentence true on a certain occasion does so in part because of what he means, or would mean, by an utterance of that sentence, and in part because of what he believes. If all we have to go on is the fact of honest utterance, we cannot infer the beliefs without knowing the meaning, and have no chance of inferring the meaning without the belief. (Davidson 1974, p. 310)

But crucially, Davidson says, beliefs can only be identified against the background of other beliefs that are related to them. I can believe that there is a cloud passing before the sun —to use one of his examples— only if I believe that there is a sun, that water can turn into gas, and so on. There may not be a list of necessary and sufficient conditions for having any one belief, Davidson says, but some related beliefs must be ascribed along with any given one if the ascription is to make sense at all.<sup>6</sup> According to Davidson, this entails that a speaker can't be too wrong about something if we are to sensibly interpret her as talking or thinking about *it*: “too much mistake simply blurs the focus”. (Davidson 1984c, p. 168).

At first sight it is tempting to argue from the Principle of Charity to **(Sto)**. The argument in question could go something as follows. To take our interlocutors to assert logical falsehoods, or dissent from logical truths, is to take them to be wrong. Of course, Charity doesn't require us to make our interlocutors infallible —“the aim is not the absurd one of making disagreement and error disappear”. Nevertheless, logical principles seem to have a special status. They are obvious (or at least let's grant that they are, for the sake of argument). They are also, as Quine puts it, “germane to all topics” (1960, p. 13). And they are crucial to the explanation of rational behaviour.<sup>7</sup> Perhaps the point can be extended to obvious sentences in general. As Quine says in a slightly different context:

Somehow we feel that if our interlocutor will not agree with us on these platitudes there is no depending on him in most of the further contexts containing the terms in question. (Quine 1960, p. 67)

The idea is that taking someone to be wrong about obvious matters is incompatible with them being ‘right enough’ and ‘coherent enough’ to be interpretable. Charity, therefore, should push us to translate away their apparently deviant views. Let's call this the Argument from Charity.

---

<sup>6</sup>See (Davidson 1984b) for the example, and (Glüer 2011, Ch. 3) for more on the topic.

<sup>7</sup>This is perhaps clearer in the case of inferences than in the case of logical truths. Think, for example, of explaining someone's action of ringing a doorbell in terms of their desire to make their presence known and their belief that *if* they ring the doorbell, *then* their presence will be known.

The Argument from Charity, I think, looks better from a distance than it does from up close. To make matters simpler I will focus on logic, and in particular on the debate between classical and non-classical logicians. In this context the crux of the argument is the claim that if we don't translate away our interlocutor's logical deviance, they will seem unintelligible to us. But is this true? Or at least, is this true in all cases, as **(Sto)** requires?

A first reason to be sceptical is that the Argument from Charity overestimates how much non-classical logicians need to deviate from classical logic. In many cases the departure from classical standards is limited to highly specialised domains. Intuitionists (some intuitionists, at least) refrain from using classical logic in the context of mathematical discourse that quantifies over infinite domains, and no further. Paraconsistentists (some paraconsistentists, anyway) abandon classical logic only when it comes to unusual sentences that make classical logicians hesitate and puzzle, like the semantic paradoxes. And more generally, non-classical logicians of all breeds appeal to recapture results in order to show that one can reason classically in most contexts.<sup>8</sup> The idea that, since logic is germane to all topics and crucial to the explanation of rational behaviour, taking our interlocutors to be wrong about logic makes their overall behaviour unintelligible, loses much of its bite. Generally speaking, taking someone to be wrong about logic need not involve taking them to be wrong that often, nor about vital day-to-day matters.

A second reason to be weary is perhaps obscured by Quine's writing. His examples of disagreement about logic are often artificial thought experiments. They deal with fictional speakers who flatly assert contradictions without further explanation, or who neatly switch up the classical principles that govern conjunction and disjunction. Real-life examples are always more sophisticated. Relevantly for us, flesh and bone non-classical logicians give detailed arguments as to *why* they reject classical logic. They typically fail to convince classical logicians, but the arguments at least explain how one would come to be wrong (wrong, that is, from the classical logician's point of view). In practice, taking our interlocutors to be wrong about logic is often to impute them a reasonable, perhaps even compelling mistake, and this is not the type of mistake that threatens to make them unintelligible. Moreover, although non-classical logicians reject classical principles, their standards of correct argumentation are as clear as those of classicists. The feeling that "if our interlocutor will not agree with us on platitudes there is no depending on him in most of the further contexts containing the terms in question" tends to arise only in fictional cases, and just for lack of detail. Classical mathematicians may take intuitionists to be wrong, but they are perfectly capable of reasoning constructively.

And lastly, a third reason to doubt the Argument from Charity is that partisans of different logics sometimes *do* take each other to be wrong, rather than to speak a different language. That doesn't mean that they are right in doing so, of course. But it does suggest, against the Argument from Charity,

---

<sup>8</sup>See e.g. (Fiore and Rosenblatt 2023) for an overview of recapture results and the problems they involve.

that we can make sense of the verbal and non-verbal behaviour of someone who we take to be mistaken about logic without assuming that they speak a different language.

Here we should consider one last line of defence for the Argument from Charity. If the objections above are correct, they show (at best) that we don't *need* to translate away disagreement about logic, but the Principle of Charity demands more; it requires us to make our interlocutors 'optimally right':

A good theory of interpretation maximizes agreement. Or, given that sentences are infinite in number [...] a better word might be *optimize*. (Davidson 1984c, p. 169)

With this in mind, one could argue that translating away logical deviance makes for 'optimal interpretations', and so Charity does, in the end, push us to save the obvious. To see whether this reply can work we have to ask what optimizing agreement consists in. For Davidson, it is a matter of getting the right sort of agreement, where:

The 'right sort' [...] is no easier to specify than to say what constitutes a good reason for holding a particular belief. (Davidson 1984a, p. xvii)

At times Davidson gives general descriptions of what he has in mind:

Disagreement about theoretical matters may (in some cases) be more tolerable than disagreement about what is more evident: disagreement about how things look or appear is less tolerable than disagreement about how they are; disagreement about the truth of attributions of certain attitudes to a speaker by that same speaker may not be tolerable at all, or barely. It is impossible to simplify the considerations that are relevant, for everything we know or believe about the way evidence supports belief can be put to work in deciding where the theory can best allow error, and what errors are least destructive of understanding. (Davidson 1984c, p. 169)

But here we run into two problems, one old and one new. First, even if we take this sort of characterisation on board, it is still (and by design) a collection of defeasible guidelines. When all is said and done, we face the same problem as with Quine's characterisation of **(Sto)** as an approximate rule of thumb. The Translation Argument requires **(Sto)** to apply always and everywhere, but the appeal to Charity can't deliver as sweeping a conclusion.

Secondly, it isn't clear that Davidson's talk of optimization should be accepted. Perhaps attributing too much error to our interlocutors makes them unintelligible, and therefore some degree of charity is necessary for interpretation. But once there is enough agreement with our interlocutor to make them interpretable, why should we (always and everywhere, as a matter of methodological requirement) take them to be in 'optimal in agreement' with us? The

fact —if it is a fact— that “too much mistake blurs the focus” doesn’t entail that less mistake will always make for a better interpretation. Seen this way, Davidson’s insistence on optimal agreement outstrips the justification of the Principle of Charity.

All things considered, then, the Translation Argument doesn’t hold up. The claim that classical logical truths are obvious, insofar as it is more or less precise, is questionable, and the claim that translation must save the obvious (always and everywhere) seems ungrounded, even when we try to back it up with an appeal to the Principle of Charity. Now the question is: are there other ways of arguing for the Meaning-Variance claim?

## 5.2 Logic and Use

In this section we will look into a second argument Quine offers in favour of the Meaning-Variance thesis. The best way to introduce it is to test the reader’s patience and turn to **(Obv)** yet another time. In the previous section we understood **(Obv)** as a descriptive claim: the claim that, as a matter of fact, everyone in ‘our’ linguistic community unhesitatingly assents to classical logical truths.<sup>9</sup> We also saw that there seems to be a problem with this claim. If our linguistic community includes, say, all English-speakers, then **(Obv)** is plainly false; there are English-speaking non-classical logicians. If, on the other hand, we exclude non-classical logicians from our linguistic community, we seem to be begging the question. The Translation Argument itself was supposed to show that non-classical logicians are linguistic deviants, so we can’t assume it from the outset.

Alternatively, we can read **(Obv)** as a sort of stipulation. On this reading, Quine’s point is simply that belonging to the linguistic community of classical logicians *consists in* assenting to all classical logical truths, and refusing to assent to classical logical falsehoods. To mean what classical logicians mean by ‘or’, ‘not’, and so on, is just to have a certain disposition to assent and dissent. Let’s call this the constitutive reading of **(Obv)**.

As before, the constitutive reading of **(Obv)** can be generalised in two ways. First, if we want a notion of logic that isn’t solely focused on logical truth, we can say that using classical logical vocabulary also consists in accepting all classically valid inference steps. And secondly, we can extend the constitutive reading from classical to any other logic: to use the connectives of a certain logic  $\mathcal{L}$ , on this view, is just to be disposed to assent to and dissent from certain sentences, and to accept some inferences.

This way of understanding **(Obv)** has immediate consequences for the Meaning-Variance claim. We can illustrate them with an example from the previous section: Quine’s imaginary natives, who earnestly assent to sentences of the form ‘ $q$  ka bu  $q$ ’. On the constitutive reading of **(Obv)**, there can be no evidence that ‘ka’ should be translated by classical conjunction, and ‘bu’ by classical negation.

---

<sup>9</sup>At least once a suitable classically valid argument for it has been produced, recall the notion of potential obviousness.

To use classical conjunction and negation *consists in* not earnestly assenting to sentences of the form ‘ $q$  and not  $q$ ’ (among other things). Translating ‘ka’ and ‘bu’ by classical conjunction and negation, therefore, can never be correct. If we also assume, with Quine, that English speakers use ‘and’ and ‘not’ in a way that is faithfully described by classical logic, we should conclude that ‘ka’ and ‘bu’ don’t correspond to ‘and’ and ‘not’, respectively. The apparent disagreement with the natives could only be a mistranslation.

The point is as general as we want the constitutive reading of **(Obv)** to be. If using the vocabulary of a given logic  $\mathcal{L}$  consists in having a certain disposition to assent and dissent from sentences, and to accept some inferences, then whoever doesn’t display the disposition in question doesn’t use the vocabulary of  $\mathcal{L}$ . Partisans of different logics, therefore, use different logical vocabulary. We can call this the Use Argument.

Authors like Berger (1990) and Gustafsson (2017) take the Use Argument to be at the core of Quine’s case for Meaning-Variance. This raises two questions. First, how good is the Use Argument as an interpretation of Quine? And second, how good is it as an argument? The rest of this section is focused on the first question, and the rest of the chapter on the second.

In §13 of *Word and Object* Quine puts forward the Use Argument more or less explicitly, although only for propositional connectives. He begins by stating that there are behavioural criteria which describe the use of classical negation, conjunction, and disjunction in terms of assent and dissent:

The criterion of negation is that it turns any [...] sentence to which one will assent into a sentence from which one will dissent, and vice versa. That of conjunction is that it produces a compound to which [...] one is prepared to assent always and only when one is prepared to assent to each component. That of alternation is similar with assent changed twice to dissent. (Quine 1960, pp. 57-8)<sup>10</sup>

Fulfilling these semantic criteria, Quine says, is all that matters when it comes to deciding whether to translate a native idiom by the corresponding classical connective:

When we find that a native construction fulfills one or another of these three semantic criteria, we can ask no more toward an understanding of it. (Quine 1960, pp. 58)

We have settled a people’s logical laws completely, so far as the truth-functional part of logic goes, once we have fixed our translations by the above semantic criteria. (Quine 1960, pp. 60)

---

<sup>10</sup>Attentive readers will have noticed that these are the rules ( $+¬$  I/E), ( $+∧$  I/E) and ( $-∨$  I/E) from Chapter 4. Incidentally, these rules are enough to derive the rest of Rumfitt’s operational rules with the help of his coordination principles, and they were the only primitive operational rules in (Smiley 1996).

A similar claim appears in *Philosophy of Logic*. There Quine asks us to imagine a logician whose conjunction is governed by the classical laws for disjunction, and whose disjunction is governed by the classical laws for conjunction. After noting that we would “regard his deviation [from classical logic] as merely notational and phonetic”, Quine goes on to say:

Could we be wrong in so doing? Could he really be meaning and thinking genuine conjunction in his use of ‘and’ after all, just as we do, and genuine alternation in his use of ‘or’, and merely disagreeing with us on points of logical doctrine respecting the laws of conjunction and alternation? *Clearly this is nonsense. There is no residual essence of conjunction and alternation in addition to the sounds and notations and the laws in conformity with which a man uses those sounds and notations.* (Quine 1986, p. 81, my emphasis)

Clearly that the Use Argument is at least part of what Quine had in mind when defending the Meaning-Variance claim. At the same time, some of what he says is outright incompatible with this argument. To begin with, Quine denies that his point about the translation of connectives extends to quantifiers:

Of what we think of as logic, the truth-functional part is the only part the recognition of which, in a foreign language, we seem to be able to pin down to behavioral criteria. (Quine 1960, p. 61)

And more puzzlingly, his discussion of behavioural criteria for the translation of connectives is immediately followed by an appeal to the Principle of Charity, and by a passage, quoted earlier, in which the directive to translate away logical deviance is put in terms of common sense and likelihood:

The maxim of translation underlying all this is that assertions startlingly false on the face of them are likely to turn on hidden differences of language. [...] The common sense behind the maxim is that one’s interlocutor’s silliness, beyond a certain point, is less likely than bad translation or, in the domestic case, linguistic divergence. (Quine 1960, p. 59)

Note, however, that according to the constitutive reading of **(Obv)** it makes no sense to say that a translation according to which our interlocutors share our logical vocabulary but reject one of our logical principles is ‘unlikely’: to use our logical vocabulary *is* to share our dispositions to assent, dissent, and acquiesce to inferences.

By and large, different authors have emphasised the Use or the Translation Argument at the cost of passing over some of what Quine says. Priest (2003) and Bryant (2017) focus on the Translation Argument, and leave out passages where Quine seems to adopt the constitutive reading of **(Obv)**. Berger (1990), who prefers the Use Argument, doesn’t discuss Quine’s take on quantifiers or his appeal to Charity, and Gustafsson (2017), who follows Berger, leaves out

the quantifier issue but notices a certain tension in the appeal to Charity. Ultimately, the question of which reading gives the ‘real’ Quine is not important. Both interpretations have enough textual support, and both give rise to complications. Instead, I will focus on a more important question. I have argued that the Translation Argument doesn’t hold up. But what about the Use Argument?

Clearly, what animates this argument is the constitutive reading of **(Obv)**. In its most general form, this is the claim that using the vocabulary of a certain logic  $\mathcal{L}$  consists in a disposition to assent to some sentences, dissent from others, and accept certain inferences. Now, Quine insists at several points that *any* deviation from classical logic amounts to a change in all logical vocabulary:

We had been picturing the rejection of the law of excluded middle, ‘ $p \vee \neg p$ ’, mainly as rejection of classical negation. I have now directed the intuitionist’s case rather at the alternation. Actually the distinction is unreal; once you upset the interrelations of the logical operators, you may be said to have revised any or all. (Quine 1986, p. 87, my emphasis)

This claim has come under fire more or less since Quine first put it forward. The usual response, going back to Morton (1973) and the early work of Putnam (1957), is that parties who agree on enough logical principles use the same logical vocabulary. This response comes with a burden: one has to explain what ‘enough agreement’ is. Morton and the later Putnam (see his 1969) thought no precise criterion could be given. More recently, however, some philosophers have tried to provide one. We will call them *minimalists*, and examine their views in the following three sections.

### 5.3 Minimalism (and some bureaucracy)

Minimalism is best seen as part of the broader tradition of (strict) inferentialism, according to which the meaning of logical vocabulary is given by the rules of inference governing its use. For minimalists not all inference rules contribute to the meaning of logical constants. They draw a boundary roughly along the line between the operational and structural rules of sequent calculi: operational rules determine meanings, structural rules don’t. Disagreements about structural rules, therefore, need not involve a change of meaning.

It is useful to distinguish between two types of minimalist proposal, which tackle slightly different questions. The first one is: when do different *logics* have the same vocabulary? Here we are looking for identity criteria for connectives and quantifiers across logics, where a logic is something different to the various calculi for it. This is the problem Francesco Paoli (2003, 2014) addresses; let’s call his approach ‘logic-first’ minimalism.

Since minimalists hold that inference rules define the vocabulary they govern, they can also ask a second, related question: when do different *sets of inference rules* define the same logical constant? Here we are looking for identity criteria across calculi, rather than logics. Greg Restall (2002, 2014) and

Bogdan Dicher (2016) focus on this second problem; let's call their approach 'rule-first' minimalism.

The distinction should not be blown out of proportion. Ultimately all minimalists have to address both questions, and answers to one cannot be independent of answers to the other. Still, it is important to note that these *are* different questions, and that they bring about different types of identity criterion.

In the next two sections I will argue against both logic-first and rule-first minimalism. The problem with them is easy to explain: it is possible to give different operational rules for the same connective, and we can exploit this fact to show that minimalist criteria of identity yield clearly wrong verdicts about which connectives are the same. More precisely, I will argue that Paoli's logic-first criterion is too coarse-grained, whereas Restall and Dicher's rule-first accounts draw too many distinctions.

Before we get into the details we have to fix some terminology and notation. From now on, capital Greek letters range over multisets of sentences, and Roman capital letters from the beginning of the alphabet range over sentences. A *sequent* (in some language  $\mathbf{L}$ ) is a pair of finite multisets of  $\mathbf{L}$ -sentences, where the pair  $(\Gamma, \Delta)$  is represented as ' $\Gamma \Rightarrow \Delta$ '. A sequent-to-sequent *inference* is a pair  $(X, \sigma)$ , where  $X$  is a set of sequents (the premises) and  $\sigma$  is a sequent (the conclusion). Lastly, a *rule* (in a language  $\mathbf{L}$ ) is a set of  $\mathbf{L}$ -sequent to  $\mathbf{L}$ -sequent inferences. Rules will be presented in the usual notation, for example:

$$\frac{\Gamma, A, B \Rightarrow \Delta}{\Gamma, A \wedge B \Rightarrow \Delta}$$

We will also need a notion of 'same rule over different languages'. There is a plain, bread-and-butter sense in which this is one rule formulated using different symbols  $\wedge$  and  $\&$ :

$$\frac{\Gamma, A, B \Rightarrow \Delta}{\Gamma, A \wedge B \Rightarrow \Delta} \quad \frac{\Gamma, A, B \Rightarrow \Delta}{\Gamma, A \& B \Rightarrow \Delta}$$

This sense can be made precise with the usual appeal to rule schemata, and it is enough for our purposes. I will assume a meta-language with dummy formulas  $(\varphi, \psi, \dots)$ , dummy connectives  $(\circ, \circ', \dots)$  and dummy context symbols  $(\nabla, \nabla', \dots)$ . Using them we can formulate rule schemata such as:

$$\frac{\nabla, \varphi, \psi \Rightarrow \nabla'}{\nabla, \varphi \circ \psi \Rightarrow \nabla'}$$

An *application* (in  $\mathbf{L}$ ) of a rule schema is obtained by uniformly replacing dummy context symbols by multisets of  $\mathbf{L}$ -formulas, dummy connectives by connectives of  $\mathbf{L}$ , and dummy formulas by  $\mathbf{L}$ -formulas. An *instance* of a rule schema (in  $\mathbf{L}$ ) is the set of all its applications (in  $\mathbf{L}$ ), and rules in languages  $\mathbf{L}$  and  $\mathbf{L}'$  are 'the same rule' when they are instances of the same schema.

Discussing logic-first minimalism will also require us to distinguish between logics and calculi. Since we will have to accommodate examples of substructurality, in what follows logics are *multiset* consequence relations. In other words, a logic with underlying language  $\mathbf{L}$  is a relation between finite multisets of  $\mathbf{L}$ -formulas and  $\mathbf{L}$ -formulas, or between finite multisets of  $\mathbf{L}$ -formulas (i.e. multiple conclusions are allowed). I will use ‘ $\mathcal{L}$ ’ and ‘ $\vdash$ ’, possibly with subscripts, to denote logics, and use ‘logic’ and ‘consequence relation’ interchangeably.

Calculi are just sets of rules, and a rule  $R$  is *admissible* in a calculus  $\mathcal{S}$  if its addition to  $\mathcal{S}$  yields no new provable sequents. A single-conclusion sequent calculus  $\mathcal{S}$  is *cut-free* if the following rule is admissible in it:

$$\frac{\Gamma \Rightarrow A \quad \Gamma', A \Rightarrow C}{\Gamma, \Gamma' \Rightarrow C} \text{ (Cut)}$$

and the same goes for multiple-conclusion calculi and the rule:

$$\frac{\Gamma \Rightarrow A, \Delta \quad \Gamma', A \Rightarrow \Delta'}{\Gamma, \Gamma' \Rightarrow \Delta, \Delta'} \text{ (Cut)}^+$$

With this out of the way, we can finally discuss minimalism. I will deal with logic-first minimalism in the next section, and rule-first minimalism in Section 5.

## 5.4 Minimalism and Logics

Logic-first minimalists give identity criteria for connectives across different logics. Roughly put, they hold that the connectives of two logics are identical if they share operational rules in certain calculi. The underlying idea is that:

If we identify meaning *tout court* with operational meaning [the meaning determined by operational rules alone], we are in a position to claim that although the classes of provable sequents are different in each case [...] the connectives’ meanings do not change. (Paoli 2014, pp. 441-442)

Francesco Paoli (2003, 2014) has put forward a minimalist proposal along these lines.<sup>11</sup> His goal is to determine the conditions under which logics are ‘genuine rivals’. In order for this to happen, he says, the logics in question must (1) share their logical vocabulary, and (2) validate different arguments.

To make this idea more precise he uses the auxiliary notions of a *presentation* of a logic and a *homophonic translation*. A presentation  $\mathbf{P} = (\mathbf{L}, \mathcal{S})$  of a logic  $\mathcal{L}$  is a pair with  $\mathbf{L}$  a propositional language and  $\mathcal{S}$  a cut-free sequent calculus

---

<sup>11</sup>Here I focus on Paoli’s views as presented in his (2014). His position has since changed in important ways, as indicated in (Dicher and Paoli 2018).

(composed of rules in  $\mathbf{L}$ ) for  $\mathcal{L}$ . Two presentations  $\mathbf{P} = (\mathbf{L}, \mathcal{S})$  and  $\mathbf{P}' = (\mathbf{L}', \mathcal{S}')$  whose languages contain the connectives  $c_1, \dots, c_n$  and  $c'_1, \dots, c'_n$ , respectively, are *similar* if the arity of  $c_i$  and  $c'_i$  is the same for all  $i \leq n$ . Finally, the homophonic translation between languages  $\mathbf{L}$  and  $\mathbf{L}'$  is defined as follows:

**Homophonic translation:** Let  $\mathbf{L}$  and  $\mathbf{L}'$  be languages with connectives  $c_1, \dots, c_n$  and  $c'_1, \dots, c'_n$ , respectively, and suppose the arity of  $c_i$  and  $c'_i$  is the same for all  $i \leq n$ . The homophonic translation  $t$  of  $\mathbf{L}$ -sentences into  $\mathbf{L}'$ -sentences is inductively defined as:

$$t(p) = p \text{ for every atom } p;$$

$$t(c_1(A_1, \dots, A_n)) = c'_1(t(A_1), \dots, t(A_n)).$$

By extension, the homophonic translation of an  $\mathbf{L}$ -sequent  $A_1, \dots, A_n \Rightarrow B_1, \dots, B_m$  into an  $\mathbf{L}'$ -sequent is  $t(A_1), \dots, t(A_n) \Rightarrow t(B_1), \dots, t(B_m)$ .

With this under his belt, Paoli transforms conditions (1) and (2) above into a formal criterion of genuine rivalry for presentations, in terms of which he proposes a criterion of rivalry for logics. The criterion of rivalry for presentations goes as follows:

**(CGR):** Let  $\mathbf{P} = (\mathbf{L}, \mathcal{S})$  and  $\mathbf{P}' = (\mathbf{L}', \mathcal{S}')$  be similar presentations containing connectives  $c_1, \dots, c_n$  and  $c'_1, \dots, c'_n$ , respectively. Then  $\mathbf{P}$  and  $\mathbf{P}'$  are genuine rivals if:

- (i) For all  $i \leq n$ ,  $c_i$  has the same operational rules in  $\mathcal{S}$  as  $c'_i$  in  $\mathcal{S}'$ .
- (ii) There is an  $\mathcal{S}$ -provable sequent containing some connective  $c_k$  whose homophonic translation is not  $\mathcal{S}'$ -provable, or vice versa.

Lastly, two logics  $\mathcal{L}$  and  $\mathcal{L}'$  are genuine rivals if they have genuinely rival presentations.

As Paoli is careful to stress:

We do not require of some putatively competing logics that *all* of their presentations be genuine rivals; it suffices that *there exists* a pair of genuinely rival presentations for the criterion to apply. (Paoli 2014, p. 442)

This comment is important for two reasons. First, it emphasises that we are dealing with a criterion of identity and rivalry for logics, although the criterion itself is put in terms of presentations. And secondly, it underlines how permissive the criterion is. In the rest of this section I will explain why this gives rise to problems, but first let me narrow the debate. Here we are not really interested in the ‘rivalry’ part of the criterion —that is, in clause **(ii)**— but only in the

criterion of identity for connectives, clause **(i)**. Let's isolate this part for future reference:<sup>12</sup>

**(CIC):** Let  $\mathcal{L}$  and  $\mathcal{L}'$  be logics containing connectives  $c_1, \dots, c_n$  and  $c'_1, \dots, c'_n$ , respectively. Then  $\mathcal{L}$  and  $\mathcal{L}'$  share the same connectives if there are similar presentations  $\mathbf{P} = (\mathbf{L}, \mathcal{S})$  and  $\mathbf{P}' = (\mathbf{L}', \mathcal{S}')$  of  $\mathcal{L}$  and  $\mathcal{L}'$ , respectively, such that:

**(i)** For all  $i \leq n$ ,  $c_i$  has the same operational rules in  $\mathcal{S}$  as  $c'_i$  in  $\mathcal{S}'$ .

At first sight **(CIC)** works as intended. According to it classical logic shares its connectives with several substructural logics, including certain subsystems of linear logic. In addition **(CIC)** is also able to deal with notational variants of a logic, since notational variants can always be given calculi that share operational rules. On closer inspection, though, **(CIC)** is too coarse grained: it forces us to identify connectives that should be kept apart.

We will start with some contrived examples, and later move on to more natural ones. To begin with, take the logics  $\mathcal{L}_\wedge$  and  $\mathcal{L}_*$ .  $\mathcal{L}_\wedge$  is the conjunction-only fragment of classical logic, and  $\mathcal{L}_*$  is a trivial consequence relation over a language with a single binary connective  $*$ ; more precisely,  $\mathcal{L}_*$  consists of all pairs  $(\Gamma, A)$  with non-empty  $\Gamma$ . Quite surprisingly, **(CIC)** yields that  $\wedge$  and  $*$  are the same connective. To show this we just need to give appropriate presentations  $\mathbf{P}_{\mathcal{L}_\wedge} = (\mathbf{L}_\wedge, \mathcal{S}_\wedge)$  and  $\mathbf{P}_{\mathcal{L}_*} = (\mathbf{L}_*, \mathcal{S}_*)$ . Let  $\mathbf{L}_\wedge$  and  $\mathbf{L}_*$  be the propositional languages with (only) the binary connectives  $\wedge$  and  $*$ , respectively. As for the cut-free sequent calculi, say that  $\mathcal{S}_\wedge$ , the calculus for  $\mathcal{L}_\wedge$ , has the usual stock of single-conclusion structural rules:

$$\frac{}{A \Rightarrow A} \text{ (Id)} \quad \frac{\Gamma, A \Rightarrow C}{\Gamma, A, B \Rightarrow C} \text{ (W}_l\text{)} \quad \frac{\Gamma, A, A \Rightarrow C}{\Gamma, A \Rightarrow C} \text{ (Ctr}_l\text{)}$$

and the following (also standard) operational rules:

$$\frac{\Gamma, A \Rightarrow C}{\Gamma, A \wedge B \Rightarrow C} \quad \frac{\Gamma, B \Rightarrow C}{\Gamma, A \wedge B \Rightarrow C} \quad \frac{\Gamma \Rightarrow A \quad \Gamma \Rightarrow B}{\Gamma \Rightarrow A \wedge B}$$

Next we need a calculus  $\mathcal{S}_\wedge^*$  for the logic  $\mathcal{L}_*$ . We will let it have all the structural rules of  $\mathcal{S}_\wedge$  and, in addition, the following rule (Triv):

$$\frac{\Gamma \Rightarrow A}{\Gamma \Rightarrow B} \text{ (Triv)}$$

<sup>12</sup>Here I'm just following Paoli: "Recall that, according to **(CGR)**, whenever two connectives have the same operational properties in appropriate sequent calculus presentations, they also have identical meanings" (Paoli 2014, p. 452).

Other than that, the rules for  $*$  are the same as the rules for  $\wedge$ :

$$\frac{\Gamma, A \Rightarrow C}{\Gamma, A * B \Rightarrow C, \Delta} \quad \frac{\Gamma, B \Rightarrow C}{\Gamma, A * B \Rightarrow C} \quad \frac{\Gamma \Rightarrow A \quad \Gamma \Rightarrow B}{\Gamma \Rightarrow A * B}$$

That, however, is all we need:  $\mathcal{S}_\wedge$  and  $\mathcal{S}_\wedge^*$  are cut-free and give the same operational rules for  $\wedge$  and  $*$ . According to **(CIC)**, then,  $\wedge$  and  $*$  are the same connective.<sup>13</sup>

I think this is a good indicator that something is wrong with **(CIC)**. At the same time it probably isn't a definitive objection. Paoli's criterion is meant to be a precise version of the fuzzy, intuitive notion of 'sharing enough logical principles'.  $\mathcal{L}_*$  is an artificial limit case, though; it is (arguably) one of those instances where the notion we are trying to do justice to is not fully settled one way or another. And so, perhaps we shouldn't find it so concerning if **(CIC)** yields unexpected results in unexpected cases.

Unfortunately, the problems with **(CIC)** don't end here. To see why we just need to take another look at the calculus  $\mathcal{S}_\wedge^*$  for  $\mathcal{L}_*$ . It should be clear that the structural rule (Triv) makes the operational rules for  $*$  more or less irrelevant; we could have chosen almost any other set of operational rules and we would have ended up with a calculus for  $\mathcal{L}_*$  anyway.<sup>14</sup> Since **(CIC)** is blind to structural rules, this means we can identify  $*$  with (almost) any binary connective. Earlier we showed that  $*$  was identical to classical conjunction. Now we will show, by essentially the same argument, that  $*$  is identical to classical *disjunction*.

Take  $\mathcal{L}_\vee$ , the disjunction-only fragment of classical logic, and consider a presentation  $\mathbf{P}_{\mathcal{L}_\vee} = (\mathbf{L}_\vee, \mathcal{S}_\vee)$  where  $\mathbf{L}_\vee$  is the language containing only  $\vee$  and  $\mathcal{S}_\vee$  is the following cut-free sequent calculus:

$$\frac{}{A \Rightarrow A} \text{ (Id)} \quad \frac{\Gamma, A \Rightarrow C}{\Gamma, A, B \Rightarrow C} \text{ (W}_l\text{)} \quad \frac{\Gamma, A, A \Rightarrow C}{\Gamma, A \Rightarrow C} \text{ (Ctr}_l\text{)}$$

$$\frac{\Gamma, A \Rightarrow C \quad \Gamma, B \Rightarrow C}{\Gamma, A \vee B \Rightarrow C} \quad \frac{\Gamma \Rightarrow A}{\Gamma \Rightarrow A \vee B} \quad \frac{\Gamma \Rightarrow B}{\Gamma \Rightarrow A \vee B}$$

Now consider the presentation  $\mathbf{P}'_{\mathcal{L}_*} = (\mathbf{L}_*, \mathcal{S}_\vee^*)$  of  $\mathcal{L}_*$ , where  $\mathbf{L}_*$  is as before and  $\mathcal{S}_\vee^*$  has the structural rules (Id), (W<sub>l</sub>), (Ctr<sub>l</sub>), and (Triv), and the same operational rules for  $\vee$  that we have just seen:

<sup>13</sup>A proof that  $\mathcal{S}_\wedge$  is cut-free can be found in e.g. (Negri and Von Plato 2008, Ch. 5). To see that  $\mathcal{S}_\wedge^*$  is cut-free just note that we can derive any of its provable sequents using only (Id), (W<sub>l</sub>), and (Triv).

<sup>14</sup>More concretely, any set of operational rules such that the resulting calculus (a) was cut-free and (b) couldn't prove any sequent with an empty antecedent, would have done.

$$\frac{\Gamma, A \Rightarrow C \quad \Gamma, B \Rightarrow C}{\Gamma, A * B \Rightarrow C} \quad \frac{\Gamma \Rightarrow A}{\Gamma \Rightarrow A * B} \quad \frac{\Gamma \Rightarrow B}{\Gamma \Rightarrow A * B}$$

At this point a *rule*-first minimalist could claim that this new calculus,  $\mathcal{S}_\vee^*$ , and the previous one,  $\mathcal{S}_\wedge^*$ , define different connectives. That response is not available to Paoli, though. He puts the identity question in terms of *logics*, and  $\mathcal{S}_\vee^*$  and  $\mathcal{S}_\wedge^*$  are calculi for the same logic. Therefore, the argument can go through as before:  $\mathcal{S}_\vee^*$  and  $\mathcal{S}_{*\vee}$  are cut-free and assign the same operational rules to  $\vee$  and  $*$ , so according to **(CIC)**,  $\vee$  and  $*$  are the same connective.<sup>15</sup>

When we put everything together, we are in trouble. We have shown that  $\wedge$  is the same connective as  $*$ , and also that  $*$  is the same connective as  $\vee$ . We are then forced to say that  $\wedge$  is the same connective as  $\vee$ . But of course, classical conjunction and disjunction are not the same connective, and any criterion of identity should keep them apart. In other words, **(CIC)** will not do.

In the course of answering an objection by Hjortland Paoli presents a second, more fine-grained version of **(CIC)**. This second version uses Avron's (1988) distinction between the internal and external consequence relations of sequent calculi:

**Internal consequence relation:**  $B$  is an internal consequence of  $A_1, \dots, A_n$  in calculus  $\mathcal{S}$  (formally,  $A_1, \dots, A_n \vdash_{\mathcal{S}}^I B$ ) iff the sequent  $A_1, \dots, A_n \Rightarrow B$  is derivable in  $\mathcal{S}$ .

**External consequence relation:**  $B$  is an external consequence of  $A_1, \dots, A_n$  in calculus  $\mathcal{S}$  (formally,  $A_1, \dots, A_n \vdash_{\mathcal{S}}^E B$ ) iff the sequent  $\Rightarrow B$  is derivable in  $\mathcal{S}$  with the addition of the sequents  $\Rightarrow A_1, \dots, \Rightarrow A_n$  as axioms and (Cut) as a primitive rule.

The internal consequence relation of a calculus reflects which structural rules are admissible in it:  $\vdash_{\mathcal{S}}^I$  is monotonic if  $(W_l)$  is admissible in  $\mathcal{S}$ , transitive if  $\mathcal{S}$  is cut-free, and so on. External consequence relations, on the other hand, are more independent of admissibility facts. For instance,  $\vdash_{\mathcal{S}}^E$  is reflexive, transitive, and monotonic for any calculus  $\mathcal{S}$ .

Now, say the structural fragment of a consequence relation is its restriction to atomic sentences. The idea behind Paoli's second pass at **(CIC)** is that, in order for connectives of different logics to have the same meaning, it is not enough that there be cut-free sequent calculi that give them the same operational rules. In addition, the structural fragments of the external consequence relations of those calculi must coincide. Let's make this official:

**(CIC)<sup>+</sup>:** Let  $\mathcal{L}$  and  $\mathcal{L}'$  be logics containing connectives  $c_1, \dots, c_n$  and  $c'_1, \dots, c'_n$ , respectively. Then  $\mathcal{L}$  and  $\mathcal{L}'$  share the same connectives

<sup>15</sup>For a proof that  $\mathcal{S}_\vee$  is cut-free see (Negri and Von Plato 2001, Ch. 5). To see that  $\mathcal{S}_\vee^*$  is cut-free note again that we can derive any of its provable sequents using only (Id),  $(W_l)$ , and (Triv).

if there are similar presentations  $\mathbf{P} = (\mathbf{L}, \mathcal{S})$  and  $\mathbf{P}' = (\mathbf{L}', \mathcal{S}')$  of  $\mathcal{L}$  and  $\mathcal{L}'$ , respectively, such that:

- (i) For all  $i \leq n$ ,  $c_i$  has the same operational rules in  $\mathcal{S}$  as  $c'_i$  in  $\mathcal{S}'$ .
- (ii) The structural fragments of the external consequence relations of  $\mathcal{S}$  and  $\mathcal{S}'$  coincide.

Here a word of warning is important. Paoli explicitly formulates **(CIC)**, but **(CIC)**<sup>+</sup> is only implicitly found in his (2014). There he argues that linear logic without exponentials and lattice bounds, **LL**, and classical logic, **CL**, have the same connectives because, in addition to satisfying **(CIC)**:

Classical logic (seen as a multiset consequence relation) and the external consequence relation of **LL** coincide over the purely structural fragment: for any multiset  $\Gamma$  of propositional variables,  $\Gamma \vdash_{CL} p$  iff  $\Gamma \vdash_{LL}^E p$ . (Paoli 2014, p. 459)

Strictly speaking Paoli's talk of the external consequence relation *of a logic* doesn't make sense. Logics don't have external consequence relations, only calculi do. Still, it seems clear that in the passage above he is speaking somewhat loosely, and that the intended reading is **(CIC)**<sup>+</sup>.

The improved criterion **(CIC)**<sup>+</sup> is more sensitive to the presence of structural rules than its predecessor **(CIC)**. Relevantly for us, it blocks the previous collapse argument concerning  $\wedge$ ,  $*$ , and  $\vee$ . This is because the structural fragment of the external consequence relation of  $\mathcal{S}_\wedge$  is not the same as the structural fragment of the external consequence relation of  $\mathcal{S}_\wedge^*$ . Therefore, under the new criterion these calculi aren't enough to identify  $\wedge$  with  $*$ . The same goes, of course, for  $*$  and  $\vee$ .

This is just a stopgap, however, and **(CIC)**<sup>+</sup> runs into the same type of 'collapse argument' as **(CIC)**. This time instead of building our own counterexample we will take one off the shelf. It is well known that intuitionistic linear logic has two distinct conjunctions, the additive conjunction 'with' ( $\&$ ) and the multiplicative conjunction 'tensor' ( $\otimes$ ). According to **(CIC)**<sup>+</sup>, though, these connectives are the identical.

To show that  $\&$  and  $\wedge$  are the same connective consider the presentation  $\mathbf{P}_{\mathcal{L}_\wedge} = (\mathbf{L}_\wedge, \mathcal{S}_\wedge)$  given earlier, and compare it to the following presentation  $\mathbf{P}_{\mathcal{L}_\&} = (\mathbf{L}_\&, \mathcal{S}_\&)$  for  $\mathcal{L}_\&$ , the with-only fragment of intuitionistic linear logic.<sup>16</sup>  $\mathcal{L}_\&$  is, as expected, the language containing only  $\&$ , and  $\mathcal{S}_\&$  is just like  $\mathcal{S}_\wedge$  but without the structural rules ( $W_l$ ) and ( $Ctrl$ ). Since  $\mathcal{S}_\wedge$  and  $\mathcal{S}_\&$  are cut-free and assign the same operational rules to  $\wedge$  and  $\&$ , we have that  $\&$  and  $\wedge$  are the same connective.

Next, to show that  $\otimes$  and  $\wedge$  are also identical, consider the alternative presentation  $\mathbf{P}'_{\mathcal{L}_\wedge} = (\mathbf{L}_\wedge, \mathcal{S}_\wedge^m)$  for  $\mathcal{L}_\wedge$ , where  $\mathbf{L}_\wedge$  is as before and  $\mathcal{S}_\wedge^m$  is as follows:

<sup>16</sup>I'm restricting attention to single-connective fragments for the sake of simplicity, but the argument goes through if we take the full-stock of connectives too.

$$\frac{}{A \Rightarrow A} \text{ (Id)} \quad \frac{\Gamma, A \Rightarrow C}{\Gamma, A, B \Rightarrow C} \text{ (W}_l\text{)} \quad \frac{\Gamma, A, A \Rightarrow C}{\Gamma, A \Rightarrow C} \text{ (Ctr}_l\text{)}$$

$$\frac{\Gamma, A, B \Rightarrow C}{\Gamma, A \wedge B \Rightarrow C} \quad \frac{\Gamma \Rightarrow A \quad \Gamma' \Rightarrow B}{\Gamma, \Gamma' \Rightarrow A \wedge B}$$

To get a similar presentation  $\mathbf{P}_{\mathcal{L}_{\otimes}} = (\mathbf{L}_{\otimes}, \mathcal{S}_{\otimes})$  for  $\mathcal{L}_{\otimes}$ , the tensor-only fragment of intuitionistic linear logic, let  $\mathbf{L}_{\otimes}$  be the language containing only  $\otimes$ , and let  $\mathcal{S}_{\otimes}$  be just like  $\mathcal{S}_{\wedge}^m$  except without the rules (W<sub>l</sub>) and (Ctr<sub>l</sub>).

Once again, since  $\mathcal{S}_{\wedge}^m$  and  $\mathcal{S}_{\otimes}$  are cut-free and share operational rules,  $\wedge$  and  $\otimes$  are the same connective. And once again, when we put everything together we are in trouble:  $\&$  is identical to  $\wedge$  which is identical to  $\otimes$ , so we are forced to say that  $\&$  is identical to  $\otimes$ .

The problem with Paoli’s logic-first minimalism can be summarised as follows. Identity is a transitive notion, but the relation of sharing operational rules in some cut-free calculus is not. If we want a criterion of identity for connectives put in terms of sharing operational rules —call this relation  $\mathcal{R}$ — we are faced with a dilemma. If we don’t close  $\mathcal{R}$  under transitivity, we have a non-transitive notion of identity; in other words, we don’t have a notion of identity at all. And if we do close  $\mathcal{R}$  under transitivity, it becomes easy to chain calculi in such a way that we identify connectives that should be kept apart.

At a slightly more technical level, it is clear that the sort of collapse argument I have used only goes through because Paoli poses a question in terms of logics and answers it in terms of calculi. It is natural to wonder, then, whether rule-first minimalist proposals, which don’t have this built in ‘mismatch’ of levels, fare any better. That will be the topic of the next section.

## 5.5 Minimalism and Rules

Rule-first minimalists give criteria to identify connectives across different calculi. In this section I will look at proposals of this kind by Greg Restall (2002, 2014) and Bogdan Dicher (2016), and argue that they run into problems that are, in a sense, dual to the problems of logic-first minimalism.

Let’s begin with Restall’s proposal. According to Restall only operational rules determine the meaning of logical constants. Calculi with the same operational rules, therefore, have the same logical vocabulary. He allows for further wiggle room, though. On his account we can define identical connectives by means of rules that are only *similar*. Loosely speaking, rules are similar if they highlight the same pattern of inference.<sup>17</sup> Take these rules for negation in classical, intuitionistic, and dual-intuitionistic logic:<sup>18</sup>

<sup>17</sup>This way of glossing Restall’s position comes from (Dicher 2016).

<sup>18</sup>The example is Restall’s (2014, pp. 281-2).

$$\frac{\Gamma \Rightarrow A, \Delta}{\Gamma \neg A \Rightarrow \Delta} \quad \frac{\Gamma \Rightarrow A}{\Gamma, \neg A \Rightarrow} \quad \frac{\Rightarrow A, \Delta}{\neg A \Rightarrow \Delta}$$

If we ignore the multisets  $\Gamma$  and  $\Delta$ , all of them instantiate a common pattern:

$$\frac{\Rightarrow A}{\neg A \Rightarrow}$$

and in that sense they are ‘similar’. The idea can be made more precise as follows: operational rules are similar if they have the same active and principal formulas, and according to Restall similar rules define identical connectives.

A second minimalist proposal is put forward in (Dicher 2016). Dicher agrees with Restall that structural rules don’t influence the meaning of connectives. But he holds that the *structure of the sequents* in an operational rule does affect the meaning of the connectives it governs. Consider once more the previous batch of rules for negation:

$$\frac{\Gamma \Rightarrow A, \Delta}{\Gamma \neg A \Rightarrow \Delta} \quad \frac{\Gamma \Rightarrow A}{\Gamma, \neg A \Rightarrow} \quad \frac{\Rightarrow A, \Delta}{\neg A \Rightarrow \Delta}$$

The rule on the left allows for there to be a multiset on each side of the sequent arrow ‘ $\Rightarrow$ ’; it prescribes a multiset-multiset structure for sequents. The rule on the middle, on the other hand, allows for a multiset on the left of ‘ $\Rightarrow$ ’ and a multiset with at most one formula on the right; let’s call this a multiset-formula structure. Finally, the rule on the right allows for a multiset with at most one formula on the left of ‘ $\Rightarrow$ ’ and a multiset on its right; we will say this is a formula-multiset structure. Dicher holds that some of this structural information is inseparable from the definition of connectives, and should be taken into account when we set out to give identity criteria. To see why, let’s have another look at a standard rule for  $\otimes$ :

$$\frac{\Gamma, A, B \Rightarrow \Delta}{\Gamma, A \otimes B \Rightarrow \Delta}$$

In order to formulate this rule it must be possible to have at least two formulas on the left of ‘ $\Rightarrow$ ’. This is a non-trivial requirement, and some calculi —e.g. for dual-intuitionistic logic, see (Urbas 1996)— only allow single formulas on the left of the sequent arrow. According to Dicher, the stipulation that the collection to the left of ‘ $\Rightarrow$ ’ may have more than one member is inseparable from the definition of  $\otimes$ , and should be counted as part of its meaning:

Rules for the connectives carry structural information. [...] Some of that structural information is a *sine qua non*: it codifies structural properties which are essential for the definability of the connective. [...] I will call them connective-intrinsic or, for short, intrinsic. A connective's intrinsic properties are an integral part of its meaning. (Dicher 2016, p. 745)

More precisely, Dicher holds that connectives are defined by the *structurally minimal* version of their operational rules, that is, by the formulation of their operational rules that imposes the least structural demands.<sup>19</sup> The following rule, for instance, would play the same part in the definition of a connective as the previous one:

$$\frac{\Gamma, A, B \Rightarrow C}{\Gamma, A \otimes B \Rightarrow C}$$

I think Restall and Dicher's proposals, despite their differences, run into the same problem. I will try to explain it in general terms first, and then give some examples. We can pin down the same logic by means of different sets of inference rules. Put more plainly: there are lots of calculi for any given logic. Rule-first minimalists give criteria about which sets of inference rules define the same connectives, and this seems to come with a burden: to make sure that calculi for the same logic define the same connectives. This is where I think Restall and Dicher's proposals go wrong. By focusing on the fine-grain of rules, they end up drawing too many distinctions.

I will use classical propositional logic over a language with conjunction and negation as a case study. We are going to define the connectives by means of rules, as strict inferentialists would have it. Let's say we settle on the following structural rules (the choice is more or less irrelevant):

$$\frac{}{\Gamma, A \Rightarrow A, \Delta} (\text{Id})^+ \quad \frac{\Gamma \Rightarrow A, \Delta \quad \Gamma', A \Rightarrow \Delta'}{\Gamma, \Gamma' \Rightarrow \Delta, \Delta'} (\text{Cut})^+$$

$$\frac{\Gamma, A, A \Rightarrow \Delta}{\Gamma, A \Rightarrow \Delta} (\text{Ctr})_l^+ \quad \frac{\Gamma \Rightarrow A, A, \Delta}{\Gamma, A \Rightarrow A, \Delta} (\text{Ctr})_r^+$$

and add the following rules for negation:

$$\frac{\Gamma \Rightarrow A, \Delta}{\Gamma, \neg A \Rightarrow \Delta} (\neg \text{L}) \quad \frac{\Gamma, A \Rightarrow \Delta}{\Gamma \Rightarrow \neg A, \Delta} (\neg \text{R})$$

<sup>19</sup>He also requires that rules should be conservative and unique in the sense of (Belnap 1962), but this will play no part in what follows.

Now all we need to do is add some rules for classical conjunction. How should we do it? With the following set of (additive) rules, we get a calculus for classical logic:

$$\frac{\Gamma, A \Rightarrow \Delta}{\Gamma, A \wedge_1 B \Rightarrow \Delta} \quad \frac{\Gamma, B \Rightarrow \Delta}{\Gamma, A \wedge_1 B \Rightarrow \Delta} \quad \frac{\Gamma \Rightarrow A, \Delta \quad \Gamma \Rightarrow B, \Delta}{\Gamma \Rightarrow A \wedge_1 B, \Delta}$$

But we are spoilt for choice. If we add these (multiplicative) rules instead, we still get a calculus for **CL**:

$$\frac{\Gamma, A, B \Rightarrow \Delta}{\Gamma, A \wedge_2 B \Rightarrow \Delta} \quad \frac{\Gamma \Rightarrow A, \Delta \quad \Gamma' \Rightarrow B, \Delta'}{\Gamma, \Gamma' \Rightarrow A \wedge_2 B, \Delta, \Delta'}$$

Here is the problem: according to Restall and Dicher's proposals, these two sets of rules define two different connectives.<sup>20</sup> And of course this raises the question: which of them (if any) is classical conjunction?

The problem goes beyond  $\wedge_1$  and  $\wedge_2$ . For example, if we use this set of rules:

$$\frac{\Gamma, A, B \Rightarrow \Delta}{\Gamma, A \wedge_3 B \Rightarrow \Delta} \quad \frac{\Gamma, A \wedge_3 B \Rightarrow \Delta}{\Gamma, A, B \Rightarrow \Delta}$$

we also get a calculus for classical logic, and it is easy to check that, according to Dicher and Restall's criteria,  $\wedge_3$  is different from both  $\wedge_1$  and  $\wedge_2$ .

The same holds if we add these rules:

$$\frac{\Gamma, A, B \Rightarrow \Delta}{\Gamma, A \wedge_4 B \Rightarrow \Delta} \quad \frac{}{\Gamma, A, B \Rightarrow A \wedge_4 B, \Delta}$$

More generally, we can define infinitely many 'different classical conjunctions' by exploiting the fact that, given our structural rules, demanding several occurrences of a formula in a premise does not change a rule's yield. For example, this set of rules:

$$\frac{\Gamma, A, A, B \Rightarrow \Delta}{\Gamma, A \wedge_5 B \Rightarrow \Delta} \quad \frac{\Gamma \Rightarrow A, \Delta \quad \Gamma' \Rightarrow B, \Delta'}{\Gamma, \Gamma' \Rightarrow A \wedge_5 B, \Delta, \Delta'}$$

<sup>20</sup>In Restall's case this is because neither of the left introduction rules in the first set is similar to the left introduction rule of the second set. In Dicher's case, it is because —amongst other things— the structurally minimal formulation of the rules in the first set calls for a formula-formula structure, whereas the second needs at least a multiset-formula structure, due to the left introduction rule.

is just like the set defining  $\wedge_1$ , except that now we require *two* occurrences of  $A$  in the premiss-sequent of the left rule. Once more,  $\wedge_5$  counts as a new connective in both Restall and Dicher’s accounts, but the resulting calculus is still a calculus for classical logic. The same goes for the versions of these rules that require three, four, or any finite number of occurrences of  $A$ .

There seem to be two possible responses. The first is to appeal to harmony constraints. If only one set of rules for conjunction is harmonious, and the others are ‘defective’ or ‘unacceptable’ definitions, then the problem disappears. I don’t think there is much to this response. Every harmony criterion I am aware of validates two or more of the sets of rules we have considered, and it is easy to come up with more examples.<sup>21</sup> Even if we bring some form of harmony constraint into the picture, we still have to account for ‘different’ classical conjunctions.

The second, more radical response is to bite the bullet and insist that there are several classical conjunctions. Dicher (2020) follows this route. He argues that despite the usual assumption that classical logic has only one conjunction,  $\wedge_1$  and  $\wedge_2$  are different connectives. More generally, he holds that there are two versions (additive and multiplicative) of each of the usual classical connectives. He puts the point as follows:

Classical logic is usually presented over a language that does not explicitly contain Sheffer’s stroke. Even the banal ‘if and only if’ rarely makes it into the primitive language. It simply isn’t technically convenient. But if one wishes one can add them [...] safe in the knowledge that they will not modify the consequence relation, *while nevertheless leaving their mark at the level of proofs*. Consequence-wise, the biconditional’s behaviour can be mimicked with the help of the conditional and of the conjunction. [...]. Rule-wise, ‘if and only if’ is a proper citizen of the classical consequence relation. [...] Ditto for multiplicative disjunction and for every other connectives which are additive/multiplicative relations of each other. One of each pair is (classically) enough; more make for better representativity—to stick to political metaphors. (Dicher 2020, p. 89, my emphasis)

I don’t think this response works either. A first reason to worry is that, as I have argued above, both Restall and Dicher’s accounts force us to distinguish between much more than  $\wedge_1$  and  $\wedge_2$ . Perhaps there is some initial plausibility to the idea of importing the additive/multiplicative distinction from linear to classical logic, but the distinctions between  $\wedge_1$ – $\wedge_5$  (and all their variants) are of a different kind. A second reason to worry is that holding that there are several classical conjunctions—or disjunctions, or whatnot—puts the minimalist in an awkward spot. The goal, at the outset, was to come up with a criterion of identity for logical vocabulary according to which partisans of different logics

---

<sup>21</sup>(Avron 1991) and (Došen 1989), for instance, contain a different set of invertible rules for conjunction, and more generally, two sets of invertible rules for each of the usual connectives. All of them would do as examples in the discussion above.

speak the same language. But now we are committed to something harder to swallow: that partisans of the same logic speak different languages.

The third, and most important reason why I don't think Dicher's strategy works is the following. From an inferentialist point of view (which is shared by all minimalists) differences in meaning turn on differences in use. In the case of logical constants, differences in meaning turn on differences in the way those constants are used in proofs. Now, suppose there are two classical mathematicians, one of which uses  $\wedge_1$  and the other  $\wedge_3$ . What differences between their inferential practices would there be? They would agree on which sentences are theorems. They would also agree on which sentences follow from which (that is, on which arguments are valid). They would even agree on which 'meta-inferences' are correct, in the sense that the calculi with  $\wedge_1$  and  $\wedge_3$  above have the same sequent-to-sequent derivability relation. What else is there to look at? In the passage just quoted Dicher speaks of a difference 'at the level of proofs'. What he is pointing out is that, depending on which of  $\wedge_1$  and  $\wedge_3$  we take as primitive in a calculus, we end up with different sets of derivation trees. The question, then, is whether this difference is enough for an inferentialist to draw a distinction in meaning. The answer, I would argue, is 'no'.

There is always a good amount of slack between the informal proofs found in mathematics journals and textbooks, on the one hand, and the derivation trees in formal calculi set up by logicians, on the other. This is something that happens by design. Formal derivations are meant to 'fill in the gaps' —to use Frege's metaphor— of informal proofs; they substitute the actual inferential steps made in proofs as they occur in the wild by putatively simpler, 'smaller' ones. Crucially, this filling in of the gaps always includes an irreducible amount of choice by the logician. Logicians have not, and could not, 'discover' that the rules governing the use of conjunction in classical mathematics are those of  $\wedge_1$ , as opposed to those of  $\wedge_3$ . Logicians set out to account for the many patterns of inference involving conjunctions in classical mathematics in terms of a few elementary rules. And ultimately, they *decide* which rules to use for this task. The range of choice is not unlimited, and perhaps some rules are more natural than others. But in a good number of cases there is no reason to prefer some set of primitive rules over another. Granted, different choices determine different sets of derivation trees in the resulting calculi. But these differences depend on choices on the part of the logician, not on the actual use of the connectives. They do not, in general, stem from a difference in use, and so they are irrelevant —from an inferentialist point of view— when it comes to drawing distinctions in meaning.

Let's summarise the point. Rule-first minimalists identify the connectives defined by different calculi when their operational rules resemble each other. Calculi for the same logic, however, can have very different rules, and so minimalist criteria turn out to be too narrow: standard calculi for the same logic define, according to them, different connectives.<sup>22</sup> Moreover, insisting that most

---

<sup>22</sup>I have used classical logic as an example, but only for simplicity. The same kind of argument goes through for any structural logic, and several substructural ones.

logics have distinct versions of the same connective will not do. This idea — besides being implausible— forces us to (a) construe partisans of the same logic as speaking different languages and (b) draw distinctions in meaning that don't correspond to differences in use, which seems unacceptable from an inferentialist point of view.

## 5.6 Another Look at Structural Rules

In this section I will go through some further observations that weigh against minimalism, and sketch an alternative way to look at structural rules. I won't attempt to give a detailed theory of how structural rules affect the meaning of connectives, or an alternative criterion of identity for logical vocabulary. The goal is, rather, to make sense of the failure of the minimalist proposals we have looked at.

According to inferentialists logical vocabulary is defined by the role it plays in inferences. Minimalists add that there is a 'core' part of the inferential behaviour of each logical constant. They characterise this part in terms of operational rules, and argue that logical constants which fulfil the same 'core' inferential role are identical. The objections to minimalism I have raised so far stem from the same observation: that it is misleading to talk of *the* operational rules of a connective. The behaviour of (what we normally think of as) a single connective can be brought about by means of very different operational rules, even when we fix a stable background of structural assumptions. This makes logic-first minimalism clump too many connectives together, and rule-first minimalism draw too many distinctions.

There is a second, related worry about minimalism. Minimalists assume a substantial distinction between the core operational meaning of a connective, on the one hand, and the non-essential part of its behaviour that depends on structural rules, on the other. But the distinction between structural and operational rules, and between inferences validated by the former and inferences validated by the latter, is to some extent a matter of preference. Let's stick to classical conjunction as an example. In a calculus for classical conjunction with multiplicative rules, like  $\mathcal{S}_{\wedge}^m$ , we can prove the sequent  $p \Rightarrow p \wedge p$  as follows:

$$\frac{\frac{p \Rightarrow p \quad p \Rightarrow p}{p, p \Rightarrow p \wedge p}}{p \Rightarrow p \wedge p} (\text{Ctr}_l)$$

In fact, it is easy to show that any derivation of  $p \Rightarrow p \wedge p$  in this calculus must contain at least one application of  $(\text{Ctr}_l)$ .<sup>23</sup> Still, it would be a mistake to conclude that the validity of  $p \vdash p \wedge p$  is inseparable from the rule of Contraction.

<sup>23</sup>The quickest way is by backwards proof search: the calculus only proves classically valid sequents, so the last rule applied to obtain  $p \Rightarrow p \wedge p$  can only be  $(\text{Ctr}_l)$ .

If we use additive rules, as in the calculus  $\mathcal{S}_\wedge$ , we can derive the same sequent using operational rules alone:

$$\frac{p \Rightarrow p \quad p \Rightarrow p}{p \Rightarrow p \wedge p}$$

and if we use a mix of additive and multiplicative rules:

$$\frac{\Gamma, A, B \Rightarrow C}{\Gamma, A \wedge B \Rightarrow C} \quad \frac{\Gamma \Rightarrow A \quad \Gamma \Rightarrow B}{\Gamma \Rightarrow A \wedge B}$$

we can get rid of  $(W_l)$  and  $(Ctr_l)$  altogether.<sup>24</sup> There is no fact of the matter as to whether the validity of  $p \vdash p \wedge p$  depends on the core operational meaning of classical conjunction. All these calculi give an equally good account of the behaviour of ‘and’ in classical logic, but they draw the structural/operational boundary in different places.

The situation does not change when we add more connectives to the mix. We can isolate aspects that are common to the behaviour of all the vocabulary of a logic (or calculus) and chalk them up to structural rules; but we can also build them into operational rules one connective at a time. When we look at things this way, it is misleading to think of structural and operational rules as different in kind. Instead, the former become a limit case of the latter: structural rules are operational rules shared by all connectives (and atoms). This makes structural rules useless when it comes to distinguishing between connectives *of the same logic* (or calculus). But it stops short of the minimalist claim that structural rules are irrelevant for distinguishing connectives *of different logics* (or calculi).

There is one last observation to keep in mind. Structural rules are usually laid down for arbitrary formulas. For example, in the standard formulation of left weakening:

$$\frac{\Gamma \Rightarrow C}{\Gamma, A \Rightarrow C} (W_l)$$

the active formula  $A$  can be of any complexity. In these cases structural rules (partly) govern the behaviour of formulas containing logical vocabulary, just like operational rules. But what if we postulate structural rules only for atoms? For instance, what if we opt for the following rule of ‘atomic weakening’?

$$\frac{\Gamma \Rightarrow C}{\Gamma, p \Rightarrow C} (W_l)'$$

---

<sup>24</sup>See e.g. (Negri and von Plato 2001, Ch. 2) for a proof.

One may think that when structural rules are formulated this way, they shouldn't play a role in the identity conditions of connectives. After all, they don't explicitly license any inferences involving them.

In his (1979) Hacking defends a version of this view. According to Hacking the operational rules of a calculus are definitions of the connectives they govern provided that (1) if we postulate structural rules, they must only apply to atoms and (2) if we postulate a structural rule for atoms, any application of it to complex formulas must be admissible:

I shall presently qualify the suggestion that Gentzen's 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. Cut-elimination, dilation-elimination, and identity-elimination (for complex formulas) are necessary conditions for this. (Hacking 1979, p. 298)

Now, this idea seems related to to the bilateral harmony criterion of the previous chapter. But how exactly? And more relevantly for this section, can Hacking's proposal be used to defend minimalism against the objections I have considered? Let's take each question in turn.

Although Hacking's idea is similar in spirit to the harmony criterion of Chapter 4, accepting the latter doesn't commit us to accepting the former. This is, first of all, because the harmony criterion of the previous chapter has to do with coordination principles, while Hacking's idea relates to structural rules. Granted, on bilateralist readings of sequent calculi (e.g. Restall 2005) structural rules and coordination principles come to the same thing. But there are many other ways of reading sequent calculi, ways that don't force us to identify the two types of rule.

More importantly, the harmony criterion I suggested earlier is only a sufficient, but not necessary, condition for the acceptability of sets of rules. The point was only that if the interaction between complex formulas and coordination principles is severed, then the bilateral problem of harmony reduces to its unilateral part. That doesn't commit us to any further views about whether coordination principles play a role in determining meanings, or about whether they should be kept in mind when thinking of the identity conditions of logical vocabulary. And of course, it also doesn't commit us to Hacking's claim that the admissibility of structural rules is a necessary condition for the acceptability of connectives, even if we take structural rules to be coordination principles in different notation.

Let us now turn to the second question: can Hacking's idea save the minimalist proposals we have looked at? As far as I can tell, it can't: the type of objection I have pointed out seems to go through even if we enforce Hacking's restrictions.

Take the multiplicative calculus  $\mathcal{S}_\wedge^m$ . As it stands, it is a calculus for  $\mathcal{L}_\wedge$ , the  $\wedge$ -only fragment of classical logic. If we get rid of  $(W_l)$  and  $(Ctr_l)$  altogether, we have a calculus for the  $\otimes$ -only fragment of intuitionistic linear logic. But

if we only *restrict* its structural rules to atoms, we still have a calculus for  $\mathcal{L}_\wedge$  and, moreover, a calculus where all applications of structural rules to complex formulas are admissible. What’s more, it is enough to add a rule of trivialisation for atoms:

$$\frac{\Gamma \Rightarrow p}{\Gamma \Rightarrow q} \text{ (Triv)'}$$

to get a calculus for the trivial logic  $\mathcal{L}_*$ , and a calculus in which all applications of structural rules to complex formulas are admissible.<sup>25</sup>

Something similar holds for  $\mathcal{S}_\wedge$ , the additive calculus for  $\mathcal{L}_\wedge$ . Removing weakening and contraction yields a calculus for the &-only fragment of intuitionistic linear logic, but restricting structural rules to atoms yields a calculus for  $\mathcal{L}_\wedge$  in which non-atomic applications of structural rules are admissible. And once again, adding the atomic rule (Triv)’ results in a calculus for  $\mathcal{L}_*$  where applications of structural rules to complex formulas are still admissible.<sup>26</sup>

What this means is that the ‘collapse argument’ of Section 4, and the ‘too-many-distinctions argument’ of Section 5, go through even if we adopt Hacking’s restrictions. In other words, it seems that, when it comes to giving identity conditions for connectives, we simply can’t ignore structural rules, even when they are formulated just for atoms.

## 5.7 A final look at Meaning-Variance

Let’s take a step back. We begun by distinguishing between two arguments for the Meaning-Variance claim, the Translation Argument and the Use Argument. We concluded (or at any rate, I did) that the Translation Argument is not successful: it establishes at best that Meaning-Variance is often plausible, but this much nobody denies.

The Use Argument, on the other hand, has more legs. If logical vocabulary is individuated by the principles that govern its use, then indeed, changing those principles amounts to changing logical vocabulary. But as we saw, the Use Argument argument rests on the controversial claim that any change in logic, no matter how small, is enough to change the meaning of all logical constants.

As Putnam and Morton pointed out long ago, it seems plausible to say that parties who agree on enough logical principles share their logical vocabulary. We therefore looked at Dicher, Restall and Paoli’s proposals as to what ‘enough agreement’ is, and found them lacking. Granted, there may be other ways to cash out the notion of ‘enough agreement’, but the ball is in the minimalists’ court. In the rest of this section I won’t try to weigh in any further on whether the Meaning-Variance claim is true. Instead I will challenge two common assumptions that, I think, are behind much of the attention that this claim has attracted.

<sup>25</sup>For proofs of these claims see the Appendix to this chapter.

<sup>26</sup>Proofs of these claims are also in the Appendix

The first assumption is that there *must* be a fact of the matter about whether any given logical expressions are synonymous. Perhaps the clearest formulation of this worry comes from Boghossian's (1996):

How can there fail to be facts about whether any two expressions mean the same? [...] Wouldn't this have to entail that there are no facts about what each expression means individually? Putting the question the other way: Could there be a fact of the matter about what each expression means, but no fact of the matter about whether they mean the same? (Boghossian 1996, pp. 370-71)

Considered very generally, the point seems less concerning than Boghossian allows. Most philosophers would grant that vague predicates are meaningful, and also that there is no fact of the matter about some meaning-related claims involving them. When it comes to logical vocabulary, though, the worry seems more pressing. The claim that there is something vague or unsettled about the meaning of logical expressions seems to imply that there is something vague or unsettled about logical validity itself, even within a single logic. And moreover, inferentialists claim the role of logical vocabulary in inference, and hence its meaning, can be pinned down in terms of precise rules. How, then, could there fail to be a fact of the matter about whether logical expressions mean the same?

There are, I think, ways to look at the situation which can help drive away this concern. Perhaps the most straightforward is to say that, although there is nothing vague about the inferential behaviour of logical constants, what is 'unsettled' or 'up for grabs' is which part of that inferential behaviour we should count as meaning-constituting. To illustrate the point it may be useful to rehash our discussion about minimalism in terms of an analogy that inferentialists are fond of, the analogy between logic and chess. The idea, as we saw in Chapter 1, is that being a certain piece in the game of chess consists in being moved according to certain rules. Similarly, being a certain logical constant, like conjunction or a disjunction, consists in being used a certain way in inferences.

Now, just as there are non-classical logics, there are 'non-classical' variants of chess. And just as we can ask whether different logics use different logical vocabulary, we may want to ask if the many well-known chess variants use different chess pieces. If we take chess pieces to be defined by the rules that govern their use, this is tantamount to asking exactly which rules define the different pieces. And just as in the case of logic, several options are available.

One idea —the analogue of Quine's position— is to say that any change in the rules of chess amounts to changing the pieces of the game. Recall that for Quine:

Once you upset the interrelations of the logical operators, you may be said to have revised any or all. (Quine 1986, p. 87)

But this is not the only option. The distinction between structural and operational rules maps surprisingly well from logic onto chess. Here is, for example, a natural way to sort chess rules into 'structural' and 'operational':

**Structural Rules:**

**S<sub>1</sub>:** Only one piece can be moved per turn.

**S<sub>2</sub>:** No piece can be moved to a square occupied by a piece of the same colour.

⋮

**Operational Rules:**

**O<sub>b</sub>:** Bishops move diagonally.

**O<sub>r</sub>:** Rooks move in straight lines.

⋮

It is possible, then, at least in principle to be a ‘minimalist about chess’, and claim that chess pieces are defined by the operational rules governing their use. A moment’s reflection shows that minimalism about chess runs into the same problems as minimalism about logical constants. The distinction between aspects of the behaviour of chess pieces that are due to ‘operational rules of chess’, on the one hand, and aspects that are due to its ‘structural rules’, on the other, is as arbitrary in chess as it is in logic. Here is, for example, an alternative way to lay down the rules of chess:

**Structural Rules:**

**Operational Rules:**

**O’<sub>b</sub>:** (i) Bishops move diagonally and (ii) they can’t be moved to a square occupied by a piece of the same colour and (iii) if a player moves a bishop, their turn ends.

**O’<sub>r</sub>:** (i) Rooks move in straight lines and (ii) they can’t be moved to a square occupied by a piece of the same colour and (iii) if a player moves a rook, their turn ends.

⋮

It is easy to see how we could exploit this fact to translate the collapse argument from Section 5, or the too-many-distinctions-argument from Section 6, from connectives to chess pieces. But there is one important sense in which the analogy between chess pieces and logical constants breaks down. In the case of chess, there is no temptation to insist that there must be a fact of the matter about whether all chess variants use the same chess pieces; we are content to say that nothing in our use of expressions like ‘is the same chess piece as’ settles the matter one way or another. More importantly, this doesn’t commit us to any sort of (problematic) scepticism. It is unambiguously clear whether

any given move is legal within a game of chess, or within a game of any well-defined chess variant. We see at once that the behaviour of chess pieces can be captured in terms of precise rules, while their identity conditions are obscure. Ditto for logical expressions. Inferentialists can say that anything we may want to count as part of the meaning of logical vocabulary comes from its (sharply defined) behaviour in inferences, but there is no fact of the matter as to what exactly we should chalk-down to meaning and what to, say, the ‘background context of deducibility’. This does not entail that there is something vague about what counts as a classical theorem, or a classically valid inference (and similarly for any other logic). And in this light, the claim that there is no fact of the matter about whether some logical constants are synonymous appears much less threatening than it seems at first sight.<sup>27</sup>

Now let’s turn to the second, widespread assumption about the Meaning-Variance debate. It is often thought that if the Meaning-Variance claim is true, then disagreements about logic boil down to verbal disputes. The point has been put in many different ways, but the general idea is that if partisans of different logics use different languages, there can be no ‘genuine’ or ‘substantial’ disagreement between them.

This is a mistake. Admittedly, the Meaning-Variance claim is incompatible with *some* forms of disagreement. If —getting back to an earlier example—classical logicians hold that all sentences of the form ‘ $P$  or<sub>c</sub> not<sub>c</sub>  $P$ ’ are logically true, while intuitionists deny that some sentences of the form ‘ $P$  or<sub>i</sub> not<sub>i</sub>  $P$ ’ are true, then classicists and intuitionists are talking about different sentences. So be it; this still leaves many interesting forms of disagreement on the table.

A first type of disagreement that is compatible with the Meaning-Variance claim is ‘descriptive’ disagreement. Here we run once more into the distinction between pure and applied logic that Priest has emphasised throughout his work. A pure logic is an abstract theory, formulated in an artificial, uninterpreted language. To apply a pure logic we need a systematic way to match expressions in some natural language with strings of symbols in the artificial language; inferring, as Priest puts it, “is a practice carried out in the vernacular” (2003, p. 459). Often, different applied logics give opposing verdicts concerning an inference in natural language. Therefore, Priest says, there is room for (substantial) disagreement concerning which pure logic best describes some implicit standard of correct inference:

Rivalry in [...] logic occurs when different theories, each relative to the way that it is applied, deliver different, incompatible verdicts [...] concerning the validity of some inference. For example, consider the inference ‘If there is no greatest prime number then there is a greatest prime number. Hence, there is a greatest prime number’. Given the usual identifications, classical logic says that it is correct to draw this inference; intuitionistic logic that it is not. (Priest 2003, p. 459)

---

<sup>27</sup>For a related idea see Stewart Shapiro’s (2015, 2019), where he argues that ‘meaning the same’, as applied to logical constants, is both vague and context-dependent.

Other forms of disagreement are available. For instance, partisans of different logics may reply that they use different logical vocabulary *for good reasons*. Here Dummett is a clear example. He held that classical logic can only be vindicated, or rather, can only be understood, against the background of notions of truth and truth conditions that transcend verification. But those notions, he claimed, are ultimately unintelligible:

It is just this which an opponent of a realist model of meaning finds incredible: he cannot believe that a grasp of a notion of truth transcending our capacities for its recognition can be acquired, and displayed, only by the acceptance of certain forms of reasoning. He concludes, instead, that these forms of reasoning, though generally accepted, are fallacious. (Dummett 1978, p. 318)

The point is not whether Dummett was right. The point is that the sort of disagreement he had with classical logicians is both compatible with the Meaning-Variance claim and anything but unsubstantial. According to his own description of the situation, Dummett chose to speak a non-classical language on the grounds that the language of classical mathematicians was incoherent. To describe this as two parties *merely* talking past each other is a blunder.

Putnam is another excellent example. In his (1969) he proposes to adopt the ‘Quantum Logic’ of Birkhoff and Von Neumann (1936) to better accommodate experimental results of modern physics. In the course of his discussion, he considers the charge that adopting a non-classical logic would involve abandoning classical logical vocabulary.<sup>28</sup> And he immediately dismisses the thought that this could be a problem:

Even if this be true, little of interest to the philosophy of logic follows. From the fact that ‘a language which does not have a word  $V$  which obeys such-and-such patterns of inference does not contain the concept *or* (or whatever) in its customary meaning’ it does not follow that a language which is adequate for the purpose of formulating true and significant statements about physical reality must contain [...], or that it should contain, a word  $V$  which obeys such and such patterns of inference. Indeed, it does not even follow that an optimal scientific language *can* contain such a word  $V$ . (Putnam 1969, p. 188-9)

What counts as a good reason to speak a different language depends, in this context, on what counts as a good reason to adopt a given logic. For philosophers like Dummett, good reasons are semantic reasons. For anti-exceptionalists like Putnam, they are the sort of reason that guide us when adopting any other theory, like simplicity and fit with observation.<sup>29</sup> And for others still, good reasons may have to do with metaphysical assumptions (see Priest 2014; Tahko

---

<sup>28</sup>Incidentally, it is clear from his discussion that he has the Use Argument in mind.

<sup>29</sup>See (Martin and Hjortland 2024) for an recent overview of anti-exceptionalism in logic.

2021). Regardless of which of them is right —if any— the fact remains that the Meaning-Variance claim does not entail that disagreements about logic are trivial or uninteresting. As Putnam put it in his (1969):

The important fact to keep in mind [...] is that the whole ‘change of meaning’ issue is raised by philosophers only to *minimize* a conceptual revolution. But only a demonstration that a certain *kind* of change of meaning is involved, namely, *arbitrary linguistic change*, would successfully demolish the philosophical importance I am claiming for these revolutions. And *this* kind of change of meaning is certainly *not* what is involved (Putnam 1969, p. 190)

Superficially, the Meaning-Variance claim promises to do away with disagreement about logic. Some welcome this idea, and are drawn to Meaning-Variance for this reason. Others find the notion that disputes about the validity of arguments are ‘merely verbal’ unacceptable, and look for ways to resist the Meaning-Variance claim. Ultimately, though, what lies behind both reactions is a misunderstanding of the consequences of Meaning-Variance. What is actually at stake in the Meaning-Variance debate is not whether there can be disagreement about logic; it is how best to describe the disagreement between partisans of different logics. Meaning-Variance, on its own, neither does away with nor precludes substantial disagreement about logic.

## Appendix II

In this Appendix I provide the proofs for several claims in Section 6.

### Results about $\mathcal{S}_\wedge^m$

We need to prove that if we restrict all the structural rules of  $\mathcal{S}_\wedge^m$  to atoms, then in the resulting calculus (a) we can derive the same sequents as in  $\mathcal{S}_\wedge^m$ , (b) all applications of structural rules to complex formulas are admissible, and (c) if we add the atomic rule (Triv)' we get a calculus for  $\mathcal{L}_*$  where again applications of structural rules to complex formulas are admissible. Since (a) implies (b), we will prove (a) and (c). To prove (a) it is enough to show that all derivations of  $\mathcal{S}_\wedge^m$  can be transformed into derivations where structural rules are applied to atoms. Eliminating non-atomic applications of contraction requires the following lemma:

**Lemma 5.1.** *If  $\Gamma, P \wedge Q \Rightarrow C$  is derivable, so is  $\Gamma, P, Q \Rightarrow C$ . Moreover, in the derivation of  $\Gamma, P, Q \Rightarrow C$ , all new applications of  $(Ctr_l)$  are to formulas of lower complexity than  $P \wedge Q$ .*

*Proof.* See (Negri and von Plato 2001), Lemma 5.1.1, (i) □

Now we can show that:

**Lemma 5.2.** *Non-atomic applications of  $(Ctr_l)$  can be substituted for applications of  $(Ctr_l)$  to formulas of lower complexity.*

*Proof.* Take a derivation of the form:

$$\frac{\frac{\pi}{\Gamma, P \wedge Q, P \wedge Q \Rightarrow C}}{\Gamma, P \wedge Q \Rightarrow C}$$

By Lemma 5.1 (applied twice) there is a derivation of  $\Gamma, P, P, Q, Q \Rightarrow C$ . We apply  $(Ctr_l)$  twice to derive  $\Gamma, P, Q \Rightarrow C$ , and then apply the left operational rule for  $\wedge$  to get a derivation of  $\Gamma, P \wedge Q \Rightarrow C$  □

By successive applications of Lemma 5.2 we can eliminate all non-atomic applications of  $(Ctr_l)$ . Then we can eliminate non-atomic applications of (Id) and  $(W_l)$  by successive uses of the following rewriting schemes:

$\mathbf{R}_{(Id)}$ :

$$\frac{}{P \wedge Q \Rightarrow P \wedge Q} \quad \rightsquigarrow \quad \frac{\frac{P \Rightarrow P \quad Q \Rightarrow Q}{P, Q \Rightarrow P \wedge Q}}{P \wedge Q \Rightarrow P \wedge Q}$$

$\mathbf{R}_{(W_l)}$ :

$$\frac{\frac{\pi}{\Gamma \Rightarrow C}}{\Gamma, P \wedge Q \Rightarrow C} \quad \rightsquigarrow \quad \frac{\frac{\frac{\pi}{\Gamma \Rightarrow C}}{\Gamma, P \Rightarrow C}}{\Gamma, P, Q \Rightarrow C}}{\Gamma, P \wedge Q \Rightarrow C}$$

This completes the proof of claim (a). It only remains to show that (c) if we add the rule:

$$\frac{\Gamma \Rightarrow p}{\Gamma \Rightarrow q} \text{ (Triv)'}$$

we get a calculus for  $\mathcal{L}_*$  where non-atomic applications of structural rules are admissible. This is straightforward. First note that:

**Lemma 5.3.** *We can derive  $\Gamma \Rightarrow p$  for any  $\Gamma \neq \emptyset$  and any atom  $p$  using only atomic applications of structural rules.*

*Proof.* Let  $q$  be an atom occurring in a formula  $D \in \Gamma$  (there must be at least one, since  $\Gamma \neq \emptyset$ ). We can derive  $q \Rightarrow p$  using atomic applications of (Id) and (Triv), then derive  $D \Rightarrow q$  from it using atomic applications of  $(W_l)$  and operational rules, and then derive  $\Gamma \Rightarrow q$  from  $D \Rightarrow q$  by the same method.  $\square$

Finally, with the help of the previous lemma, it is easy to show that:

**Lemma 5.4.** *We can derive  $\Gamma \Rightarrow A$  for any  $\Gamma \neq \emptyset$  and any  $A$  using only atomic applications of structural rules.*

*Proof.* By induction on the complexity of  $A$ . The base case is covered by the previous lemma. If  $A = P \wedge Q$ , there are two cases. If  $\Gamma$  contains more than one formula occurrence, let  $\Gamma', \Gamma''$  be non-empty multisets such that  $\Gamma = \Gamma' \cup \Gamma''$ . By inductive hypothesis we can derive  $\Gamma' \Rightarrow P$  and  $\Gamma'' \Rightarrow Q$ , and then apply the right-introduction rule for  $\wedge$  to get  $\Gamma \Rightarrow P \wedge Q$ . If  $\Gamma$  contains only one formula occurrence, let  $r$  be an atom of that formula. Then by inductive hypothesis we can derive  $r \Rightarrow P$  and  $r \Rightarrow Q$ , and apply the right-introduction rule for  $\wedge$  to get  $r, r \Rightarrow P \wedge Q$ . After that we apply  $(\text{Ctr}_l)$  to  $r$ , getting  $r \Rightarrow P \wedge Q$ , and derive  $\Gamma \Rightarrow P \wedge Q$  from it as in the previous lemma.  $\square$

## Results about $\mathcal{S}_\wedge$

We need to prove that if we restrict all the structural rules of  $\mathcal{S}_\wedge$  to atoms, then in the resulting calculus (a) we can derive the same sequents as in  $\mathcal{S}_\wedge$ , (b) all applications of structural rules to complex formulas are admissible, and (c) if we add the atomic rule (Triv)' we get a calculus for  $\mathcal{L}_*$  where again applications of structural rules to complex formulas are admissible. Since (a) implies (b), we will, again, just prove (a) and (c). To prove (a) it is enough to show that all

derivations of  $\mathcal{S}_\wedge$  can be transformed into derivations where structural rules are applied to atoms. The only tedious case is eliminating non-atomic applications of  $(\text{Ctr}_l)$ , which requires some lemmas. As usual, the *height* of a derivation is the greatest number of successive applications of rules in it.

**Lemma 5.5.** *If  $\Gamma \Rightarrow A \wedge B$  is derivable, then so are  $\Gamma \Rightarrow A$  and  $\Gamma \Rightarrow B$ . Moreover, in the derivations of  $\Gamma \Rightarrow A$  and  $\Gamma \Rightarrow B$  all new applications of  $(\text{Ctr}_l)$  are to formulas of lower complexity than  $P \wedge Q$ .*

*Proof.* By induction on the height of derivations. □

**Lemma 5.6.** *If  $\Gamma, P \wedge Q \Rightarrow C$  is derivable, then so is  $\Gamma, P, Q \Rightarrow C$ . Moreover, in the derivation of  $\Gamma, P, Q \Rightarrow C$  all new applications of  $(\text{Ctr}_l)$  are to formulas of lower complexity than  $P \wedge Q$ .*

*Proof.* By induction on the complexity of  $C$ .

**Base case:**

If  $C = r$  for some atom  $r$  our derivation looks as follows:

$$\frac{\pi}{\Gamma, P \wedge Q \Rightarrow r}$$

But then it is clear, by direct inspection of the rules of  $\mathcal{S}_\wedge$ , that some top-sequent in  $\pi$  is of the form  $r \Rightarrow r$ , where  $r$  is a subformula of some  $D \in \Gamma$ , or  $P$ , or of  $Q$ . Regardless of which it is we can derive  $\varphi \Rightarrow r$  (where  $\varphi$  is  $D$ ,  $P$  or  $Q$ , as appropriate) using applications of  $(\text{W}_l)$  and operational rules, and then derive  $\Gamma, P, Q \Rightarrow r$  from it similarly.

**Inductive step:**

If  $C = A \wedge B$ , then by Lemma 5.5 there are suitable derivations of  $\Gamma, P \wedge Q \Rightarrow A$  and  $\Gamma, P \wedge Q \Rightarrow B$ . But then by inductive hypothesis there are suitable derivations of  $\Gamma, P, Q \Rightarrow A$  and  $\Gamma, P, Q \Rightarrow B$ , from which we can derive  $\Gamma, P, Q \Rightarrow A \wedge B$  using the right-introduction rule for  $\wedge$ . □

**Lemma 5.7.** *If  $\Gamma, P, Q \Rightarrow C$  is derivable, then so is  $\Gamma, P \wedge Q \Rightarrow C$ . Moreover, in the derivation of  $\Gamma, P \wedge Q \Rightarrow C$  all new applications of  $(\text{Ctr}_l)$  are to formulas of lower complexity than  $P \wedge Q$ .*

*Proof.* By induction on  $C$ . The base case is analogous to the base case of Lemma 5.6. For the inductive step, note that when  $C = A \wedge B$ , there are (once again by Lemma 5.5) suitable derivations of  $\Gamma, P, Q \Rightarrow A$  and  $\Gamma, P, Q \Rightarrow B$ . Therefore, by inductive hypothesis there are suitable derivations of  $\Gamma, P \wedge Q \Rightarrow A$  and  $\Gamma, P \wedge Q \Rightarrow B$ , from which we can derive  $\Gamma, P \wedge Q \Rightarrow A \wedge B$  using the right-introduction rule for  $\wedge$ . □

We can finally show that non-atomic applications of  $(\text{Ctrl}_l)$  can be eliminated. First note that:

**Lemma 5.8.** *Non-atomic applications of  $(\text{Ctrl}_l)$  can be substituted for applications of  $(\text{Ctrl}_l)$  to formulas of lower complexity.*

*Proof.* Take a derivation of the form:

$$\frac{\frac{\pi}{\Gamma, P \wedge Q, P \wedge Q \Rightarrow C}}{\Gamma, P \wedge Q \Rightarrow C}$$

By Lemma 5.6 (applied twice) there is a derivation  $\pi'$  of  $\Gamma, P, P, Q, Q \Rightarrow C$ . Then we can apply  $(\text{Ctrl}_l)$  twice to get a derivation of  $\Gamma, P, Q \Rightarrow C$ , and by Lemma 7 there is a derivation  $\pi''$  of  $\Gamma, P \wedge Q \Rightarrow C$ .  $\square$

By successive applications of Lemma 5.8 we can eliminate all non-atomic applications of  $(\text{Ctrl}_l)$ . Then we eliminate non-atomic applications of  $(\text{Id})$  and  $(\text{W}_l)$  by successive uses of the following rewriting schemes:

$\mathbf{R}_{(\text{Id})}$ :

$$\frac{}{P \wedge Q \Rightarrow P \wedge Q} \rightsquigarrow \frac{\frac{P \Rightarrow P}{P \wedge Q \Rightarrow P} \quad \frac{Q \Rightarrow Q}{P \wedge Q \Rightarrow Q}}{P \wedge Q \Rightarrow P \wedge Q}$$

$\mathbf{R}_{(\text{W}_l)}$ :

$$\frac{\frac{\pi}{\Gamma \Rightarrow C}}{\Gamma, P \wedge Q \Rightarrow C} \rightsquigarrow \frac{\frac{\pi}{\Gamma \Rightarrow C}}{\Gamma, P \Rightarrow C} \frac{}{\Gamma, P \wedge Q \Rightarrow C}$$

This completes the proof of claim (a). The proof of (c) for  $\mathcal{S}_\wedge$  is analogous to the corresponding proof for  $\mathcal{S}_\wedge^m$ .

## Chapter 6

# Rules, Models, and Meanings

Just forget all that we've told you  
and the theory that you've read

---

Glenn Gould, *So You Want to  
Write a Fugue*

So far we have looked at different problems from a strict inferentialist point of view. In this chapter we will switch gears and turn to moderate inferentialism instead. According to this type of view, recall, meaning is determined by inferential use, but the notion of meaning itself need not be cashed out in inferential terms. A popular option for moderate inferentialists is to characterise meanings using the standard machinery of reference and truth-conditions. In Chapter 2 we briefly outlined one way this could be done. The thought was that a reasonable assignment of semantic values should make certain argument patterns come out truth-preserving. This places constraints on what assignments of reference and truth-conditions are adequate for a given language, and in that sense allows us to read a semantics off of inferential practice. The example we used was quite simple. We ordinarily take arguments like those below to be correct:

$$A, B \vdash A \wedge B \quad A \wedge B \vdash A \quad A \wedge B \vdash B$$

In order for them to be truth-preserving a conjunction  $A \wedge B$  must be true if and only if both its conjuncts  $A$  and  $B$  are true. Therefore, we said, the inferential role of 'and' determines the standard truth-conditions for conjunctions. We also mentioned that matters are more complicated for other propositional connectives, and even more so when we add quantifiers to the mix. The time has come to delve into those complications.

One of the main problems for moderate inferentialism is that, on many natural ways of understanding what interpretations, inferential roles, and meaning determination consist in, inferential roles determine ‘unintended’ or ‘deviant’ meanings for logical vocabulary. Roughly put: even when we cash out the notion of meaning in the standard referential terms, the meanings pinned down by inferential roles aren’t the usual ones. We will call this Carnap’s Problem, and say more about it in Section 1.

Philosophers have come up with different workarounds to Carnap’s Problem. Here I will examine the most prominent ones: James Garson’s (Section 2), Timothy Smiley’s (Section 3), Vann McGee’s (Section 4), and a joint proposal by Dag Westerståhl and Dennis Bonnay (Section 5). After this, in Section 6, I will put forward a solution of my own, and take stock in Section 7. The chapter ends with a brief Coda that wraps up the thesis by going back to its beginning, the Problem of Harmony, seen this time from a moderate inferentialist point of view.

## 6.1 What is Carnap’s Problem?

According to moderate inferentialists the meaning of logical vocabulary is determined by its inferential role. This gives rise to an obvious question: is the meaning determined by the inferential roles of logical constants the usual, or ‘normal’ one? We will call this Carnap’s Question. In order to so much as pose this question we have to clarify three things:

- (i) What we mean by interpretations, and which of them are normal.
- (ii) What we mean by inferential roles.
- (iii) What we mean by inferential roles ‘determining’ or ‘ruling out’ interpretations.

There are many different answers to (i)-(iii), and in that sense, many different ‘Carnap Questions’. I will say a little more about which of them are sensible and interesting in latter sections, but for now let’s focus on one concrete example.

In the case of classical propositional logic we can take interpretations for a language  $\mathcal{L}$  to be valuations, that is, arbitrary functions from  $\mathcal{L}$ -sentences to  $\{1,0\}$ . It is then natural to say that valuations are  $\diamond$ -normal for some connective  $\diamond$  if they respect the usual truth-table for  $\diamond$ , and *normal* when they are  $\diamond$ -normal for all connectives  $\diamond$  of  $\mathcal{L}$ .

Next we have to specify what inferential roles are. We know from Chapter 2 that authors like Brandom and Peregrin take every inferential link to play a role in carving out the meaning of all expressions. In Chapter 5 we saw an analogue of this in a logic-centred setting: according to Quine any deviation from classical logic changes the meaning of all logical vocabulary. Now, the classical consequence relation  $\vdash_{\mathcal{L}}$  on  $\mathcal{L}$  gives full information about which sentences follow from which, alone or in combination with others. From a broadly holist point

of view, then, we can say that  $\vdash_{\mathcal{L}}$  encodes the inferential role of  $\mathcal{L}$ 's logical vocabulary.<sup>1</sup>

We are almost there; it just remains to say what it is for an interpretation to be (in)compatible with the inferential roles of classical connectives. Let's stick with our previous choice, and say that  $v$  is *consistent* with  $\vdash_{\mathcal{L}}$  when it makes all valid arguments truth-preserving: if  $\Gamma \vdash_{\mathcal{L}} A$  and  $v(\Gamma) = 1$ , then  $v(A) = 1$ .<sup>2</sup>

We now have a sharp, tractable version of Carnap's Question: we are asking if all valuations consistent with  $\vdash_{\mathcal{L}}$  are normal. The answer, as Carnap (1943) pointed out, is 'no' (for any expressively complete  $\mathcal{L}$ ). Say, for concreteness, that  $\mathcal{L}$  has connectives  $\wedge, \vee, \neg$  and  $\rightarrow$ , and consider the trivial valuation  $v_T$  such that  $v_T(A) = 1$  for all sentences  $A$ . It is easy to see that  $v_T$  is (trivially) consistent with  $\vdash_{\mathcal{L}}$ , but not  $\neg$ -normal. Or take  $v_{cl}$ , the valuation such that  $v_{cl}(A) = 1$  if and only if  $A$  is a classical tautology. Once again, it is routine to check that  $v_{cl}$  is normal only for  $\wedge$ , but still consistent with  $\vdash_{\mathcal{L}}$ . In fact, given any set of  $\mathcal{L}$ -sentences  $\Gamma$ , the valuation  $v_{\Gamma}$  such that  $v_{\Gamma}(A) = 1$  if and only if  $\Gamma \vdash_{\mathcal{L}} A$  is always consistent with  $\vdash_{\mathcal{L}}$ , but normal if and only if  $\Gamma$  is maximally consistent; this means that infinitely many non-normal interpretations are consistent with  $\vdash_{\mathcal{L}}$ .

This is a problem; it is (an instance of) Carnap's Problem. On our current way of understanding interpretations, inferential roles, and 'pinning down', the inferential roles of classical connectives don't pin down the usual interpretations. Either we have been wrong all along about the meaning of classical connectives, or else that meaning isn't determined by inferential roles. Luckily for moderate inferentialists there are other compelling (perhaps *more* compelling) ways to answer (i)-(iii), which give rise to new ways of understanding Carnap's Question. It is still possible, therefore, to formulate a version of moderate inferentialism that doesn't run into this difficulty. In the rest of this chapter we will look at some prominent options.

## 6.2 Garson: Local and Global Models

Let's begin with Garson's (2013) take on Carnap's Problem. The main difference with our previous set-up is that Garson thinks of inferential roles in terms of rules, rather than consequence relations. There are at least two ways of reading semantics off of rules: we can use 'local' or 'global' models. These alternatives are examined in subsections 6.2.1 and 6.2.2, respectively. In these first subsections the discussion is restricted to (classical) propositional logic. We will look at Garson's take on Carnap's Problem for first-order languages in 6.2.3.

---

<sup>1</sup> $\vdash_{\mathcal{L}}$  is an infinite set of pairs, so one may reasonably ask how speakers are supposed to know about its extension. Inferentialists can reply that speakers follow inference rules which are sound and complete for  $\vdash_{\mathcal{L}}$  (recall Quine's notion of potential obviousness). Note that, since on this approach we only care about the overall consequence relation, in practice we are saying that sets of rules which give rise to the same derivability relation between formulas determine the same meaning.

<sup>2</sup>Here and in what follows ' $v(\Gamma) = 1$ ' is short for  $v(A) = 1$  for all  $A \in \Gamma$ .

### 6.2.1 Local Models

From a certain point of view, as we have seen, it is possible to take the inferential role of (all) classical connectives to be (simultaneously) specified by the classical consequence relation. The caveat ‘from a certain point of view’ is important, because this idea will not sit well with everyone. To begin with, we know from Chapter 2 that authors like Tennant and Dummett insist that the meaning of each connective should be separable from the meaning of the rest. And even if we put separability aside, we may not want to ‘quotient over’ sets of rules that give rise to the same consequence relation. Perhaps we want to identify rules with the same strength, but as we saw in Chapter 3, derivability between formulas is not the only way to think about strength. Or perhaps, as some of the minimalists from Chapter 5, we think differences in the rules we accept make a difference in what we mean by logical connectives, regardless of their overall effect on the derivability relation of a calculus. What happens to Carnap’s Problem when we take the meaning of connectives to be determined by the rules that govern them? That is, in a nutshell, the question Garson sets out to answer.

The first step is making the question more precise. Let’s start with interpretations. In the previous section an interpretation for a language  $\mathcal{L}$  was a valuation from  $\mathcal{L}$ -sentences to  $\{1, 0\}$ . Here we see the first difference: Garson restricts attention to *non-trivial* valuations, that is, valuations such that  $v(A) = 0$  for some  $A$ . Garson doesn’t offer an argument for this restriction. Rather, after noting that Belnap and Massey (1990) call it an ‘innocuous assumption’, he remarks that:

Adopting the “innocuous assumption” causes some feelings of guilt, since ideally one would say absolutely nothing about valuations. However, many of the pleasing results in this book depend on valuations being consistent [non-trivial]. (Garson 2013, p. 12)

This move is common (though rarely justified) in the literature on Carnap’s Problem, and we will see analogues of it when discussing McGee’s and Bonnay and Westerståhl’s accounts. Let’s grant the point, at least temporarily and for the sake of argument; in the rest of this section interpretations are non-trivial valuations.

Next we need to define rules. A *sequent* (in some language  $\mathcal{L}$ ) is a pair  $(\Gamma, \Delta)$  of sets of  $\mathcal{L}$ -sentences, where the pair  $(\Gamma, \Delta)$  is written ‘ $\Gamma \Rightarrow \Delta$ ’, and an  $n$ -ary *rule*  $R$  is a set of  $n + 1$ -tuples of sequents. Each  $n + 1$ -tuple  $(\sigma_1, \dots, \sigma_{n+1})$  in an  $n$ -ary rule  $R$  is an *application* of  $R$ , where  $\sigma_1, \dots, \sigma_n$  are the premiss-sequents and  $\sigma_{n+1}$  the conclusion-sequent.<sup>3</sup>

---

<sup>3</sup>This is essentially the definition of a rule in Chapter 5, except that we are now using arbitrary sets of formulas instead of finite multisets. The switch to sets is harmless, since we won’t consider substructural logics. The use of arbitrary (rather than finite) sets of formulas is more problematic, and I will say more about it in Section 6.2. Note, incidentally, that this definition of a rule excludes higher-order rules in the sense of (Schroeder-Heister 1984). Whenever I make a statement about ‘all rules’ of a certain type, then, higher-order rules aren’t

It just remains to explain what it is for a rule to be compatible with some interpretation. Earlier we said that valuations were consistent with a consequence relation when they made all valid arguments truth-preserving. Rules, however, take us from sequents to sequents, and sequents aren't true or false; what should we say? Garson's first suggestion is to require preservation of satisfaction. A sequent  $\Gamma \Rightarrow \Delta$  is *satisfied by an interpretation*  $v$  if and only if  $v(A) = 0$  for some  $A \in \Gamma$  or  $v(B) = 1$  for some  $B \in \Delta$ , and  $v$  *preserves satisfaction* in a rule  $R$  when, for all applications of  $R$ , if  $v$  satisfies the premiss-sequents then it also satisfies the conclusion-sequent; if this happens we will also say that  $v$  *satisfies the rule*  $R$ . Finally, we say that  $v$  satisfies (or is a *local model* of) a set of rules  $\mathbf{R}$  when it satisfies each  $R \in \mathbf{R}$ .

Garson's idea is that a (set of) rule(s) is compatible with all and only its local models. He puts the point in terms of 'conditions on interpretations'. Say that a set of rules  $\mathbf{R}$  *locally expresses a condition*  $C$  on interpretations if, for all interpretations  $v$ ,  $v$  is a local model of  $\mathbf{R}$  if and only if it obeys  $C$ . Then Carnap's Question can again be put in formal terms: we want to know if Garson's rules for the connectives locally express the usual boolean conditions:

$$(\wedge) v(A \wedge B) = 1 \text{ if and only if } v(A) = v(B) = 1.$$

$$(\neg) v(\neg A) = 1 \text{ if and only if } v(A) = 0.$$

$$(\vee) v(A \vee B) = 1 \text{ if and only if } v(A) = 1 \text{ or } v(B) = 1.$$

$$(\rightarrow) v(A \rightarrow B) \text{ if and only if } v(A) = 0 \text{ or } v(B) = 1.$$

Now, Garson gives a structural rule and some operational rules for each connective. The structural rule in question is (Id):<sup>4</sup>

$$\frac{}{\Gamma, A \Rightarrow A} \text{ (Id)}$$

The operational rules (natural deduction rules in sequent-style notation) are as follows:

---

taken into account. Finally, single-conclusion rules are ('lower order') rules with *exactly* one formula in the succedent of sequents. Rules with at most one formula in succedents will only be considered in Section 6.6.

<sup>4</sup>Strictly speaking Garson's calculi also contain rules for left Weakening and Cut. These are redundant and I omit them for simplicity, but all results below continue to hold if we add the two further structural rules.

**Conjunction:**

$$\frac{\Gamma \Rightarrow A \quad \Gamma \Rightarrow B}{\Gamma \Rightarrow A \wedge B} (\wedge \text{I}) \quad \frac{\Gamma \Rightarrow A \wedge B}{\Gamma \Rightarrow A} (\wedge \text{E}_1) \quad \frac{\Gamma \Rightarrow A \wedge B}{\Gamma \Rightarrow B} (\wedge \text{E}_2)$$

**Negation:**

$$\frac{\Gamma, A \Rightarrow B \quad \Gamma, A \Rightarrow \neg B}{\Gamma \Rightarrow \neg A} (\neg \text{I}) \quad \frac{\Gamma, \neg A \Rightarrow B \quad \Gamma, \neg A \Rightarrow \neg B}{\Gamma \Rightarrow A} (\neg \text{E})$$

**Disjunction:**

$$\frac{\Gamma \Rightarrow A}{\Gamma \Rightarrow A \vee B} (\vee \text{I}_1) \quad \frac{\Gamma \Rightarrow B}{\Gamma \Rightarrow A \vee B} (\vee \text{I}_2) \quad \frac{\Gamma \Rightarrow A \vee B \quad \Gamma, A \Rightarrow C \quad \Gamma, A \Rightarrow C}{\Gamma \Rightarrow C} (\vee \text{E})$$

**Implication:**

$$\frac{\Gamma, A \Rightarrow B}{\Gamma \Rightarrow A \rightarrow B} (\rightarrow \text{I}) \quad \frac{\Gamma \Rightarrow A \rightarrow B \quad \Gamma \Rightarrow A}{\Gamma \Rightarrow B} (\rightarrow \text{E})$$

Say that  $\mathbf{R}_\wedge$ ,  $\mathbf{R}_\neg$ ,  $\mathbf{R}_\vee$  and  $\mathbf{R}_\rightarrow$  are the sets containing the appropriate operational rules and (Id). Then it is routine to check that:

**Observation 6.1.**  $\mathbf{R}_\wedge$ ,  $\mathbf{R}_\neg$ ,  $\mathbf{R}_\vee$  and  $\mathbf{R}_\rightarrow$  locally express  $(\wedge)$ ,  $(\neg)$ ,  $(\vee)$  and  $(\rightarrow)$ , respectively.

*Proof.* See (Garson 2013, Ch. 3). □

This may seem like good news for moderate inferentialists, but Observation 6.1 is one of the “pleasing results” that hinges on Garson’s restriction of interpretations to non-trivial valuations. The trivial valuation  $v_T$  is a local model of any single-conclusion rule (just like it is consistent with all single-conclusion consequence relations), but it does not obey  $(\neg)$ . Therefore, neither  $\mathbf{R}_\neg$ , nor any other set of single-conclusion rules locally expresses  $(\neg)$  when  $v_T$  counts as an interpretation.

The root of the problem is the fact that every single-conclusion rule contains exactly one formula in the succedent, so it is natural to wonder what happens when we turn to multiple-conclusion rules instead. The choice between natural deduction or sequents, in this case, doesn’t really matter, so I will focus on the more familiar sequent setting. Suppose we pick following rules  $\mathbf{R}_\wedge^+$  for conjunction:

$$\frac{}{\Gamma, A \Rightarrow A, \Delta} (\text{Id})^+ \quad \frac{\Gamma, A, B \Rightarrow \Delta}{\Gamma, A \wedge B \Rightarrow \Delta} (\wedge \text{L})^+ \quad \frac{\Gamma \Rightarrow A, \Delta \quad \Gamma \Rightarrow B, \Delta}{\Gamma \Rightarrow A \wedge B, \Delta} (\wedge \text{R})^+$$

Clearly, all  $\wedge$ -normal interpretations (or valuations) preserve satisfaction for these rules. Conversely, any interpretation (or valuation) that preserves satisfaction for all rules in  $\mathbf{R}_\wedge^+$  must be  $\wedge$ -normal. To see this, just note that all valuations satisfy  $(\text{Id})^+$ , and that any valuation which preserves satisfaction for  $\mathbf{R}_\wedge^+$  will also satisfy all sequents provable from it, such as:

$$A, B \Rightarrow A \wedge B \quad A \wedge B \Rightarrow A \quad A \wedge B \Rightarrow B$$

But plainly, valuations that satisfy these sequents are  $\wedge$ -normal.

Something similar happens with negation. Say we pick the following rules  $\mathbf{R}_\neg^+$ :

$$\frac{}{\Gamma, A \Rightarrow A, \Delta} (\text{Id})^+ \quad \frac{\Gamma \Rightarrow A, \Delta}{\Gamma, \neg A \Rightarrow \Delta} (\neg \text{L})^+ \quad \frac{\Gamma, A \Rightarrow \Delta}{\Gamma \Rightarrow \neg A, \Delta} (\neg \text{R})^+$$

Once more, all  $\neg$ -normal valuations preserve satisfaction for these rules, and conversely, all valuations that preserve satisfaction for these rules must be  $\neg$ -normal, since they will satisfy the provable sequents:

$$A, \neg A \Rightarrow \quad \Rightarrow A, \neg A$$

Any other standard set of multiple-conclusion rules would have done; for instance, if a set of rules is interderivable with  $\mathbf{R}_\wedge^+$  or  $\mathbf{R}_\neg^+$  then it has the same local models as them, so it locally express the same conditions. It is also straightforward to check that analogous results hold for the rest of classical connectives. Putting everything together, we have that:

**Observation 6.2.** Any standard set of multiple-conclusion rules for  $\wedge$ ,  $\neg$ ,  $\vee$  or  $\rightarrow$  locally expresses  $(\wedge)$ ,  $(\neg)$ ,  $(\vee)$  or  $(\rightarrow)$ , respectively, even if we count the trivial valuation as an interpretation.

Each provable sequent in a multiple-conclusion calculus for classical logic corresponds to a valid argument in the multiple-conclusion classical consequence relation: for example  $\Rightarrow A, \neg A$  corresponds to  $\vdash A, \neg A$ , and  $A, \neg A \Rightarrow$  corresponds to  $A, \neg A \vdash$ . This means that the proof of Observation 6.2 still goes through if we take classical logic to be a multiple-conclusion consequence relation, and require valuations to be consistent with it. Put another way, the ‘improvement’ with respect to Section 6.1 comes from the switch from single to multiple-conclusions, rather than from the switch from consequence relations to rules.

Garson ultimately decides against reading conditions on interpretations using local models. His main reason is that rules sometimes locally express a semantics for which they are incomplete. Take Garson's rules for  $\rightarrow$ , that is, (Id) and:

$$\frac{\Gamma, A \Rightarrow B}{\Gamma \Rightarrow A \rightarrow B} (\rightarrow \text{I}) \qquad \frac{\Gamma \Rightarrow A \rightarrow B \quad \Gamma \Rightarrow A}{\Gamma \Rightarrow B} (\rightarrow \text{E})$$

They locally express the classical truth-condition ( $\rightarrow$ ) on interpretations, as per Observation 6.1, but we know from Section 3.2.1 that they can't prove all classically valid formulas: Peirce's Law  $((A \rightarrow B) \rightarrow A) \rightarrow A$  is a case in point.

It isn't clear that this is a reason to write off local models altogether, though. A simple fix would be to stick with local models but strengthen the notion of local expression. Say that a set of rules  $\mathbf{R}$  locally expresses<sup>+</sup> a condition C on valuations if it locally expresses C *and* is complete for the resulting semantics. Garson's rules for  $\rightarrow$  don't locally express<sup>+</sup> ( $\rightarrow$ ), but many other sets of rules will do. For instance, it is enough to supplement  $\mathbf{R}_{\rightarrow}$  with Peirce's rule:

$$\frac{\Gamma, A \rightarrow B \Rightarrow A}{\Gamma \Rightarrow A}$$

to get a calculus that proves the full  $\rightarrow$ -fragment of classical logic and locally expresses<sup>+</sup> ( $\rightarrow$ ). Moreover,  $\mathbf{R}_{\wedge}$ ,  $\mathbf{R}_{\neg}$  and  $\mathbf{R}_{\vee}$ , that is, Garson's single-conclusion rules for the rest of connectives, also locally express<sup>+</sup> the intended semantics, as does any standard system of multiple-conclusion rules. In short, nothing would be lost with the switch to local expression<sup>+</sup>.

Alternatively, we could just shrug the problem off. Van McGee makes a related suggestion when discussing the possibility that some natural deduction system for second-order logic determines the full-second order semantics:

The rules of inference determine the truth conditions. The truth conditions, together with the mathematical facts, determine the logical consequence relation. The rules of inference are simple, but the mathematical facts are complex. So it should not be a cause of alarm that derivability is recursively enumerable but logical consequence is not. (McGee 2000, p. 72)

On this approach the fact that some rules locally express a semantics for which they are incomplete isn't concerning; it is one thing to determine a semantics and another to derive all valid inferences. In the absence of further arguments, then, one could reply that this sort of incompleteness is harmless, and perhaps (in hindsight) unsurprising.

Let's take stock. First, taking meanings to be determined by single-conclusion rules and reading semantics off of them using local models doesn't solve Carnap's Problem, or at least, it solves Carnap's Problem only with the help of

an ad hoc assumption concerning the trivial valuation  $v_T$ . On the other hand, if we turn to multiple conclusion rules the usual semantics is secured. This is moderately good news for moderate inferentialism. It is good news because it shows we can avoid Carnap’s Problem (at least for propositional logic). But it is only moderately good news because, as we know from Chapter 3, many inferentialists insist that multiple-conclusion rules are inadequate formalisations of inferential practice (and in fact Garson agrees, see Garson 2013, Ch. 11). Now let’s see if global models fare any better.

### 6.2.2 Global Models

Garson’s second suggestion is to read semantics off of rules using preservation of validity in sets of interpretations, rather than preservation of satisfaction. The details behind this are easy to explain. Interpretations, as before, are non-trivial valuations, and  $\diamond$ -normal interpretations respect the usual truth-table for  $\diamond$ . A sequent is *valid* in a set of interpretations  $V$  (‘ $V$ -valid’ for short) if it is satisfied by all  $v \in V$ . And finally, given a set  $V$  of interpretations and a rule  $R$ , we say that  $R$  *preserves  $V$ -validity* when, for all its applications, if the premiss-sequents are  $V$ -valid, so is the conclusion-sequent; if this happens we will also say that  $V$  is a *global model* of  $R$ .

The idea, this time, is that sets of rules are compatible with all and only their global models. Formally, we will say that a set of rules  $\mathbf{R}$  *globally expresses a condition  $C$*  on interpretations when, for all sets of interpretations  $V$ ,  $V$  is a global model of all  $R \in \mathbf{R}$  if and only if all  $v \in V$  obey  $C$ . Carnap’s Question then becomes: do Garson’s rules for the connectives globally express the usual boolean conditions?

The answer, as in the previous section, depends on whether we have single or multiple-conclusion rules in mind. Garson shows that his preferred multiple-conclusion rules globally express the intended classical semantics (Garson 2013, Ch. 11). The argument he uses is, I think, needlessly complicated, and his calculi are quite unusual, so we will show the result for  $\mathbf{R}_\wedge^+$  and  $\mathbf{R}_\neg^+$  instead. First, it is clear that any set of  $\wedge$ -normal or  $\neg$ -normal valuations is a global model of  $\mathbf{R}_\wedge^+$  or  $\mathbf{R}_\neg^+$ , respectively. But conversely, any global model of  $\mathbf{R}_\wedge^+$  or  $\mathbf{R}_\neg^+$  must contain only  $\wedge$ -normal or  $\neg$ -normal interpretations, respectively. To see this, let  $V$  be a global model of  $\mathbf{R}_\wedge^+$  or  $\mathbf{R}_\neg^+$ . Clearly  $(\text{Id})^+$  is  $V$ -valid, since it is valid in any set of valuations. But by assumption  $\mathbf{R}_\wedge^+$  (or  $\mathbf{R}_\neg^+$ ) preserves  $V$ -validity, so the following provable sequents are  $V$ -valid, and therefore satisfied:

$$\begin{array}{l}
 A, B \Rightarrow A \wedge B \qquad A \wedge B \Rightarrow A \qquad A \wedge B \Rightarrow B \\
 A, \neg A \Rightarrow \qquad \qquad \qquad \Rightarrow A, \neg A
 \end{array}$$

Once again, though, any valuation which satisfies these sequents must be  $\wedge$ -normal or  $\neg$ -normal, and this completes the proof. Analogous results hold for any set of rules interderivable with  $\mathbf{R}_\wedge^+$  or  $\mathbf{R}_\neg^+$ , and the same argument works for

any standard choice of rules for the remaining connectives. We can summarise the situation as follows:

**Observation 6.3.** Any standard set of multiple-conclusion rules for  $\wedge$ ,  $\neg$ ,  $\vee$  or  $\rightarrow$  globally expresses  $(\wedge)$ ,  $(\neg)$ ,  $(\vee)$  or  $(\rightarrow)$ , respectively, even if we count the trivial valuation as an interpretation.

Now let's turn to the more delicate single-conclusion case. The headline news is that the only set of rules that globally expresses the normal classical semantics is  $\mathbf{R}_\wedge$ , the set of rules for conjunction. Garson's argument is, again, needlessly involved, so here is an alternative one:

**Observation 6.4.**  $\mathbf{R}_\neg$ ,  $\mathbf{R}_\vee$  and  $\mathbf{R}_\rightarrow$  don't globally express  $(\neg)$ ,  $(\vee)$  and  $(\rightarrow)$ , respectively.

*Proof.* Fix a language  $\mathcal{L}$  with  $\neg, \vee$  and  $\rightarrow$  as logical vocabulary, and consider the set  $V$  containing all normal classical interpretations and the valuation  $v_{cl}$  such that  $v_{cl}(A) = 1$  if and only if  $A$  is a classical tautology. Clearly  $v_{cl}$  doesn't obey  $(\neg)$ ,  $(\vee)$  or  $(\rightarrow)$ , so it only remains to show that  $V$  is a global model of  $\mathbf{R}_\neg$ ,  $\mathbf{R}_\vee$  and  $\mathbf{R}_\rightarrow$ . To see this, first note that if a sequent  $\Gamma \Rightarrow A$  is  $V$ -valid, then it is also classically valid, since  $V$  contains all classical (normal) valuations. Now suppose that the premiss-sequents of an application of one of our rules are  $V$ -valid. Then these premiss-sequents are also classically valid, and so is the conclusion-sequent, since  $\mathbf{R}_\neg$ ,  $\mathbf{R}_\vee$  and  $\mathbf{R}_\rightarrow$  preserve classical validity. But then the conclusion sequent is also  $V$ -valid, since  $V$  contains only (normal) classical valuations and  $v_{cl}$ , which satisfies all classically valid sequents. Note, incidentally, that this argument goes through for any other set of rules which preserves classical validity.  $\square$

Despite this result, the single-conclusion setting has a further interesting twist: according to Garson we can read *intuitionistic* semantics from certain (intuitionistic) rules for implication and negation. To explain this we need one last bit of notation. Given a set of valuations  $V$  for language  $\mathcal{L}$ , let the relation  $\leq_V$  on  $V$  be defined by:

$$(\leq) v \leq_V v' \text{ iff for all } A \in \mathcal{L}, v(A) = 1 \text{ implies } v'(A) = 1.$$

Now, Garson's rules for intuitionistic implication are the same as his rules for classical implication,  $\mathbf{R}_\rightarrow$ . This shouldn't be surprising. We have already seen that  $\mathbf{R}_\rightarrow$  alone can't prove all classical validities, and in fact it is well-known that  $\mathbf{R}_\rightarrow$ , taken on its own, formalises the  $\rightarrow$ -fragment of intuitionistic rather than classical logic. Quite suggestively, the condition globally expressed by  $\mathbf{R}_\rightarrow$  is:

$$(\mathbf{C}_\rightarrow): v(A \rightarrow B) = 1 \text{ iff } (\forall v' \in V), \text{ if } v \leq v' \text{ and } v'(A) = 1 \text{ then } v'(B) = 1.$$

Garson's set of rules  $\mathbf{R}_\sim$  for intuitionistic negation contains (Id) and:

$$\frac{\Gamma, A \Rightarrow B \quad \Gamma, A \Rightarrow \sim B}{\Gamma \Rightarrow \sim A} (\sim I) \quad \frac{\Gamma \Rightarrow A \quad \Gamma \Rightarrow \sim A}{\Gamma \Rightarrow B} (\sim E)$$

These rules globally express the following condition:

$$(\mathbf{C}_{\sim}): v(\sim A) = 1 \text{ iff } (\forall v' \in V) \text{ if } v \leq v' \text{ then } v'(A) = 0.$$

The sense in which Garson claims to read intuitionistic semantics for  $\rightarrow$  and  $\sim$  should be clear:  $(\mathbf{C}_{\rightarrow})$  and  $(\mathbf{C}_{\sim})$  look remarkably like the clauses for the interpretation of  $\rightarrow$  and  $\sim$  in intuitionistic Kripke models. Unfortunately, as we will see, there are two problems with Garson's results, one related to his semantics and the other to his rules. Let's take each of them in turn.

If we have another look at Garson's set-up we immediately notice a peculiarity. Given a set  $V$  of valuations, condition  $(\leq)$  fully determines the corresponding relation  $\leq_V$  on  $V$ . Moreover, the extension of  $\leq_V$  depends on the values of *all* formulas in *all* valuations  $v' \in V$ . Now consider  $(\mathbf{C}_{\rightarrow})$  and  $(\mathbf{C}_{\sim})$  again. According to these clauses the values of  $A \rightarrow B$  and  $\sim C$  in  $v \in V$  depend (amongst other things) on which valuations  $v'$  are such that  $v \leq_V v'$ . Since this depends of the value of all formulas in all  $v' \in V$ , the values of  $A \rightarrow B$  and  $\sim C$  in  $v$  partly depend on the values of all formulas in all  $v' \in V$ .

Garson attempts to prove a theorem that, if true, would curb the extent of this dependence. Say that a condition  $C$  on valuations is *functional* if, given a set of atomic valuations  $V_{At}$ , there is a unique set of valuations  $V$  that obey  $C$  and agree with  $V_{At}$  on the value of all atoms.<sup>5</sup> What Garson attempts to show is that  $(\mathbf{C}_{\rightarrow})$  and  $(\mathbf{C}_{\sim})$  are functional. This, he says, would show that his semantics:

behaves as a recursive definition of truth ought to do, namely to provide a unique account of the truth-values of the complex formulas given an assignment of values to the propositional variables. (Garson 2013, p. 66)

Garson's proof, however, is incorrect, and  $(\mathbf{C}_{\rightarrow})$  and  $(\mathbf{C}_{\sim})$  aren't functional.<sup>6</sup> To show this we need some stage-setting. Fix a language  $\mathcal{L}_{\rightarrow, \sim}$  built up from an infinite set of propositional atoms  $At = \{p_1, p_2, p_3, \dots\}$ ,  $\rightarrow$  and  $\sim$ . Kripke models are triples  $\mathcal{M} = (W, R, u)$  with  $W$  a non-empty set,  $R$  a partial order on  $W$ , and  $u$  a function  $u : At \times W \rightarrow \{1, 0\}$  such that:

$$(\text{Mon}): \text{For all } w, w' \in W \text{ and } p \in At, \text{ if } wRw' \text{ and } u(w, p) = 1 \text{ then } u(w', p) = 1.$$

The forcing relation  $\Vdash$  is defined as usual:

<sup>5</sup>We say that  $V$  and  $V_{At}$  agree on the value of all atoms when there is a bijection  $f$  between  $V_{At}$  and  $V$  such that  $v(p) = f(v)(p)$  for all atoms  $p$ .

<sup>6</sup>See (Garson 2013, p. 67-8) for his attempted proof, which goes wrong in the second to last line of p. 67. Woods (2012, p. 285) also claims that  $(\mathbf{C}_{\rightarrow})$  and  $(\mathbf{C}_{\sim})$  are functional, although without proof; note that Woods uses the term 'categoricity' instead of 'functionality'.

$\mathcal{M}, w \Vdash p$  iff  $u(w, p) = 1$ .

$\mathcal{M}, w \Vdash A \rightarrow B$  iff  $(\forall w' \in W)$  if  $wRw'$  and  $\mathcal{M}, w' \Vdash A$ , then  $\mathcal{M}, w' \Vdash B$ .

$\mathcal{M}, w \Vdash \sim A$  iff  $(\forall w' \in W)$  if  $wRw'$  then  $\mathcal{M}, w' \nVdash A$ .

And finally, the following well-known lemma will be useful:

(Persistence) Let  $\mathcal{M} = (W, R, u)$  be a Kripke model, let  $w, w' \in W$  be such that  $wRw'$ . Then  $\mathcal{M}, w \Vdash A$  implies  $\mathcal{M}, w' \Vdash A$ , for all  $\varphi \in \mathcal{L}_{\rightarrow, \sim}$ .

We can now prove our result. It shows that  $(\mathbf{C}_{\rightarrow})$  and  $(\mathbf{C}_{\sim})$  are not functional and a little bit more —namely, clause (iii) below.

**Theorem 6.5.** *There is a set of atomic valuations  $V_{At}$  and two distinct sets of valuations  $V, V'$  such that:*

(i)  *$V$  and  $V'$  agree with  $V_{At}$  on the value of all atoms.*

(ii) *All valuations in  $V$  and  $V'$  obey  $(\mathbf{C}_{\rightarrow})$  and  $(\mathbf{C}_{\sim})$ .*

(iii) *There are valuations  $v \in V$  and  $v' \in V'$  that agree on the value of all (proper) subformulas of  $A \rightarrow B$  and  $\sim C$  but such that  $v(A \rightarrow B) \neq v'(A \rightarrow B)$  and  $v(\sim C) \neq v'(\sim C)$ .*

*Proof.* The structure of the proof is simple. We begin with a set of atomic valuations  $V_{At}$ , then construct two Kripke models  $\mathcal{M}$  and  $\mathcal{M}'$  from it, and then extract the required  $V$  and  $V'$  from  $\mathcal{M}$  and  $\mathcal{M}'$ , respectively. The set of atomic valuations is  $V_{At} = \{w_1, w_2, w_3\}$ , where  $w_1$  only makes atom  $p_1$  true,  $w_2$  only makes atoms  $p_1, p_2$  true and  $w_3$  only makes atoms  $p_1, p_2, p_3$  true. Next we build  $V$  and  $V'$  —incidentally, the construction of  $V$  follows (Garson 2013, p. 67).

**Building  $V$ :** Consider the Kripke model  $\mathcal{M} = (W, R, u)$  with  $W = V_{At}$ ,  $R$  the reflexive and transitive closure of  $\{(w_1, w_2), (w_2, w_3)\}$ , and  $u$  defined by  $u(w_i, s) = w_i(s)$  for all atoms  $s$ . We can extract a set  $V = \{v_1, v_2, v_3\}$  of valuations from  $\mathcal{M}$  by letting  $v_i(A) = 1$  iff  $\mathcal{M}, w_i \Vdash A$ . Then:

(i) By construction  $v_i(s) = w_i(s)$  for all atoms  $s$ , i.e.  $V$  agrees with  $V_{At}$  on the values of atoms.

(ii) All valuations in  $V$  obey  $(\mathbf{C}_{\rightarrow})$  and  $(\mathbf{C}_{\sim})$ : first we show, as a lemma, that  $v_i \leq v_j$  iff  $w_i R w_j$ . The right-to-left direction follows immediately from (Persistence) and  $(\leq)$ . For the left-to-right direction, suppose that  $w_i \not R w_j$ . Then  $i > j$ , so there is some atom  $s \in \{p_1, p_2, p_3\}$  such that  $\mathcal{M}, w_i \Vdash s$  but  $\mathcal{M}, w_j \nVdash s$ . This means that  $v_i(s) = 1$  and  $v_j(s) = 0$ , so by  $(\leq)$  we have  $v_i \not\leq v_j$ . This completes the proof of the lemma. But then:

$(\mathbf{C}_{\rightarrow})$   $v_i(A \rightarrow B) = 1$  iff  $\mathcal{M}, w_i \Vdash A \rightarrow B$  iff  $w_i R w_j$  and  $\mathcal{M}, w_j \Vdash A$  implies  $\mathcal{M}, w_j \Vdash B$  iff  $v_i \leq v_j$  and  $v_j(A) = 1$  implies  $v_j(B) = 1$ .

$(\mathbf{C}_{\sim})$   $v_i(\sim A) = 1$  iff  $\mathcal{M}, w_i \Vdash \sim A$  iff  $w_i R w_j$  implies  $\mathcal{M}, w_j \nVdash A$  iff  $v_i \leq v_j$  implies  $v_j(A) = 0$ .

**Building  $V'$ :** Consider the Kripke model  $\mathcal{M}' = (W', R', u')$  with  $W'$  and  $u'$  defined as before, but with  $R'$  the reflexive and transitive closure of  $\{(w_1, w_2), (w_1, w_3)\}$  instead. We extract a second set of valuations  $V' = \{v'_1, v'_2, v'_3\}$  from  $\mathcal{M}'$  letting  $v'_i(A) = 1$  iff  $\mathcal{M}', w_i \Vdash A$ . Just as in the previous case:

(i) By construction  $v'_i(s) = w_i(s)$  for all atoms  $s$ , i.e.  $V'$  agrees with  $V_{At}$  on the values of atoms.

(ii) All valuations in  $V'$  obey  $(\mathbf{C}_{\rightarrow})$  and  $(\mathbf{C}_{\sim})$ : we again need to show first that  $v'_i \leq v'_j$  iff  $w_i R w_j$ . The right-to-left direction still follows from (Persistence) and  $(\leq)$ . The left-to-right direction changes slightly. First note that  $w_i \nVdash w_j$  only in case either  $i > j$  or  $i = 2$  and  $j = 3$ . To see that  $v'_2 \not\leq v'_3$ , let  $p_9$  be any atom distinct from  $p_1, p_2, p_3$ . Then  $\mathcal{M}', w_2 \Vdash p_3 \rightarrow p_9$  but  $\mathcal{M}', w_3 \nVdash p_3 \rightarrow p_9$ , so  $v'_2(p_3 \rightarrow p_9) = 1$  but  $v'_3(p_3 \rightarrow p_9) = 0$ . For the remaining case note, once again, that if  $i > j$  then there is some atom  $s \in \{p_1, p_2, p_3\}$  such that  $\mathcal{M}, w_i \Vdash s$  but  $\mathcal{M}, w_j \nVdash s$ , which means that  $v'_i(s) = 1$  and  $v'_j(s) = 0$ , so by  $(\leq)$  we have that  $v'_i \not\leq v'_j$ . The proof that all valuations in  $V'$  obey  $(\mathbf{C}_{\rightarrow})$  and  $(\mathbf{C}_{\sim})$  then proceeds exactly as it did for  $V$ .

This completes the proof of parts (i) and (ii) of the theorem. To see that part (iii) also holds, note that  $v_2$  and  $v'_2$  agree on the value of  $p_3$  and  $p_9$ , and yet  $v_2(p_3 \rightarrow p_9) = 0$  but  $v'_2(p_3 \rightarrow p_9) = 1$ , and note also that  $v_2(\sim p_3) = 0$  but  $v'_2(\sim p_3) = 1$ .  $\square$

Theorem 6.5 shows that the value of  $v(\sim A)$  and  $v(A \rightarrow B)$  in some  $v \in V$  depends on the value of all formulas in every  $v' \in V$ , and that this value can differ between models that agree on the value of all atoms in all their valuations. It is useful to compare this situation with Kripke semantics. In a Kripke model  $(W, R, u)$  to know whether a formula is forced at some node  $w$  we need to know which nodes  $w$  is R-related to, and which atoms are forced in those nodes. However, we don't need to know what formulas are forced in every node of the model, while in Garson's semantics we do. This brings other aspects of the similarity between Garson's approach and Kripke semantics into question. The monotonicity constraint (Mon) of Kripke models represents the assumption that verifications can't be 'lost' when passing from one evidential stage to another, but the analogous condition  $(\leq)$  in Garson's framework is considerably stronger. First, it applies to arbitrarily complex formulas, rather than atomic ones. And secondly, it also stipulates the 'converse' of (Mon): if  $v(A) = 1$  implies  $v'(A) = 1$  (for all  $A \in \mathcal{L}$ ), then  $v \leq v'$ . It is unclear what to make of this further requirement, but the difference suggests that the analogy between Garson's conditions and Kripke semantics should be taken with a fair amount of caution.

Now let's turn to the second problem with Garson's results. His rules  $\mathbf{R}_{\rightarrow}$  for intuitionistic implication, recall, are (Id) and:

$$\frac{\Gamma, A \Rightarrow B}{\Gamma \Rightarrow A \rightarrow B} (\rightarrow \text{I}) \quad \frac{\Gamma \Rightarrow A \quad \Gamma \Rightarrow A \rightarrow B}{\Gamma \Rightarrow B} (\rightarrow \text{E})$$

and his rules  $\mathbf{R}_{\sim}$  for intuitionistic negation are (Id) and:

$$\frac{\Gamma, A \Rightarrow B \quad \Gamma, A \Rightarrow \sim B}{\Gamma \Rightarrow \sim A} (\sim \text{I}) \quad \frac{\Gamma \Rightarrow B \quad \Gamma \Rightarrow \sim B}{\Gamma \Rightarrow A} (\sim \text{E})$$

These look like standard rules, but something sets them apart from the norm: in all of them  $\Gamma$  may be an *infinite* set of formulas. This immediately raises the question of whether Garson's results hold for the usual, finitary version of his rules. Let  $\mathbf{R}_{\rightarrow}^f$  and  $\mathbf{R}_{\sim}^f$  be the versions of  $\mathbf{R}_{\rightarrow}$  and  $\mathbf{R}_{\sim}$  where  $\Gamma$  is only allowed to be finite. It is easy to check that Garson's results about *local* expression go through for  $\mathbf{R}_{\rightarrow}^f$  and  $\mathbf{R}_{\sim}^f$ ; as we are about to show, though, his results about global expression don't hold for finitary rules.<sup>7</sup>

Once again we need to lay some groundwork. First, given a Kripke model  $\mathcal{M} = (W, R, u)$  and some  $w \in W$ , the *submodel generated by  $w$*  is  $\mathcal{M}^w = (W', R', u')$ , where  $W' = \{x \in W : w R x\}$  and  $R', u'$  are the restrictions of  $R$  and  $u$  to  $W'$ . The following lemma about generated submodels is well-known:

(Gen. Sub.) Let  $\mathcal{M} = (W, R, u)$  be a Kripke model,  $\mathcal{M}^w = (W', R', u')$  the submodel generated by some  $w \in W$ . Then for all  $A \in \mathcal{L}_{\rightarrow, \sim}$  and all  $w' \in W'$ ,  $\mathcal{M}, w' \Vdash A$  iff  $\mathcal{M}^w, w' \Vdash A$ .

We will also need the following lemma:

**Lemma 6.6.** *Let  $\Gamma \Rightarrow A$  be an intuitionistically invalid sequent with  $\Gamma$  finite. Then:*

(i) *There is a Kripke model  $\mathcal{M} = (W, R, u)$  and some  $w \in W$  such that  $\mathcal{M}, w \Vdash G$  for all  $G \in \Gamma$ ,  $\mathcal{M}, w \not\Vdash A$ , and  $w$  makes only finitely many atoms true.*

(ii) *There is a Kripke model  $\mathcal{M}' = (W', R', u')$  and some  $w' \in W'$  such that  $\mathcal{M}', w' \Vdash G$  for all  $G \in \Gamma$ ,  $\mathcal{M}', w' \not\Vdash A$ , and  $\mathcal{M}', w' \not\Vdash C$  for some classical tautology  $C$ .*

*Proof.* (i) Let  $\Gamma \Rightarrow \varphi$  be an intuitionistically invalid sequent with  $\Gamma$  finite. By completeness of Kripke semantics there is some  $\mathcal{M}_0 = (W_0, R_0, u_0)$  with a  $w_0 \in W$  such that  $\mathcal{M}_0, w_0 \Vdash G$  for all  $G \in \Gamma$  and  $\mathcal{M}_0, w_0 \not\Vdash A$ . Let  $\mathcal{M}_1 = (W_1, R_1, u_1)$  be the submodel generated by  $w_0$ . By (Gen. Sub) we still have that  $\mathcal{M}_1, w_0 \Vdash G$  for all  $G \in \Gamma$  and  $\mathcal{M}_1, w_0 \not\Vdash A$ . To ensure that  $w_0$  only

<sup>7</sup>Garson says that he allows infinite antecedents “for ease of comparison with multiple conclusion sequent systems” (Garson 2013, p. 9). This is puzzling, given that (a) his results about multiple conclusion rules don't depend on them having infinite antecedents and (b) standard multiple conclusion sequent systems only allow finite antecedents anyway.

makes finitely many atoms true, we build a second model  $\mathcal{M}_2 = (W_2, R_2, u_2)$  from  $\mathcal{M}_1$ . Let  $W_2 = W_1$ , and  $R_2 = R_1$ . As for  $u_2$ , let it be defined exactly as  $u_1$  except in  $w_0$ , where we let  $u_2(w_0, s) = u_1(w_0, s)$  for the (finitely many!) atoms  $s$  that occur in  $\Gamma$  or  $A$ , and  $u_2(w_0, p) = 0$  for all remaining atoms. Routine induction on the complexity of  $\mathcal{L}_{\rightarrow, \sim}$ -formulas shows that for all  $B \in \Gamma \cup \{A\}$  and all  $w \in W_1$ , we have that  $\mathcal{M}_2, w \Vdash B$  iff  $\mathcal{M}_1, w \Vdash B$ . Therefore,  $\mathcal{M}_2, w_0 \Vdash G$  for all  $G \in \Gamma$  and  $\mathcal{M}_2, w_0 \not\Vdash A$ .

(ii) To prove (ii) we build a further model  $\mathcal{M}_3 = (W_3, R_3, u_3)$  from  $\mathcal{M}_2$ . Let  $W_3 = W_2 \cup \{x, y\}$  for some  $x, y \notin W_2$  (with  $x \neq y$ ) and let  $R_3 = R_2 \cup \{(x, x), (y, y), (w_0, x), (w_0, y)\}$ . As for  $u_3$ , let  $p_i, p_j$  be two atoms not occurring in  $\Gamma$  or  $\varphi$  (there must be two such atoms, since  $\Gamma$  is finite). Then we define  $u_3$  by:

- (a)  $u_3(w_0, s) = u_2(w_0, s)$  for all atoms  $s$ . Note that in particular  $u_3(w_0, p_i) = u_3(w_0, p_j) = 0$ , since  $p_i, p_j$  don't occur in  $\Gamma$  or  $\varphi$ .
- (b)  $u_3(x, s) = u_2(w_0, s)$  for all  $s \neq p_i, p_j$ , and  $u_3(x, p_i) = u_3(x, p_j) = 1$ .
- (c)  $u_3(y, s) = u_2(w_0, s)$  for all  $s \neq p_i, p_j$ , and  $u_3(y, p_i) = 1$  but  $u_3(y, p_j) = 0$ .
- (d) For all remaining  $w \in W_2$ , if  $s \neq p_i, p_j$  then  $u_3(w, s) = u_2(w, s)$ , and  $u_3(w, p_i) = 1, u_3(w, p_j) = 0$ .

Routine induction on the complexity of  $\mathcal{L}_{\rightarrow, \sim}$ -formulas shows that for all  $C \in \Gamma \cup \{A\}$  and all  $w \in W_1$ , we have that  $\mathcal{M}_3, w \Vdash C$  iff  $\mathcal{M}_2, w \Vdash C$ . This immediately implies that  $\mathcal{M}_3, w_0 \Vdash G$  for all  $G \in \Gamma$  and  $\mathcal{M}_3, w_0 \not\Vdash A$ . It is also easy to check that  $\mathcal{M}_3, w_0 \not\Vdash ((p_i \rightarrow p_j) \rightarrow p_i) \rightarrow p_i$ . Since  $((p_i \rightarrow p_j) \rightarrow p_i) \rightarrow p_i$  is an instance of Peirce's law, and hence a classical tautology, (ii) is proven.  $\square$

Here are, at last, the promised results:

**Theorem 6.7.** *There is a global model  $V$  of  $\mathbf{R}_{\rightarrow}^f$  such that some  $v \in V$  does not obey  $(\mathbf{C}_{\rightarrow})$ .*

*Proof.* Fix a language  $\mathcal{L}_{\rightarrow}$  with  $\rightarrow$  as the sole connective. An  $\mathcal{L}_{\rightarrow}$ -interpretation  $v$  is intuitionistic if there is an  $\mathcal{M} = (W, R, u)$  with a  $w \in W$  such that for all  $A \in \mathcal{L}_{\rightarrow}$ ,  $v(A) = 1$  iff  $\mathcal{M}, w \Vdash A$ . Let  $I$  be the set of intuitionistic interpretations that make some classical tautology false. Our global model is  $V = I \cup \{v_{cl}\}$ , where  $v_{cl}$  is the interpretation such that  $v_{cl}(A) = 1$  iff  $A$  is a classical tautology. Note that:

- (a)  $v_{cl} \leq_V v$  iff  $v = v_{cl}$ , since all valuations in  $I$  make some classical tautology false, and
- (b)  $v_{cl}$  does not obey  $(\mathbf{C}_{\rightarrow})$ , since  $v_{cl}(p_1) = v_{cl}(p_2) = v_{cl}(p_1 \rightarrow p_2) = 0$ .

It remains to show that the rules in  $\mathbf{R}_{\rightarrow}^f$  preserve V-validity. To prove this, we first have to show that a sequent  $\Gamma \Rightarrow A$  with finite  $\Gamma$  is intuitionistically valid if and only if it is V-valid. The left-to-right direction is shown by contraposition:

suppose  $\Gamma \Rightarrow A$  is not  $V$ -valid. Then for some  $v \in V$  we have  $v(\Gamma) = 1$  for all  $G \in \Gamma$  but  $v(A) = 0$ . If  $v \in I$  the result immediately follows. If  $v = v_{cl}$  the result follows as well, since all classically valid sequents hold in  $v_{cl}$ , so  $\Gamma \Rightarrow A$  must be classically, and hence intuitionistically, invalid. The right-to-left direction is also shown by contraposition: suppose  $\Gamma \Rightarrow \varphi$  is intuitionistically invalid. Then by Lemma 6.6, part (ii), there is a Kripke model  $\mathcal{M} = (W, R, u)$  and some  $w \in W$  such that  $\mathcal{M}, w \Vdash G$  for all  $G \in \Gamma$ ,  $\mathcal{M}, w \not\Vdash A$ , and  $\mathcal{M}, w \not\Vdash C$  for some classical tautology  $C$ . But then there is some  $v \in I \subseteq V$  that invalidates  $\Gamma \Rightarrow A$ .

The theorem immediately follows from this. All applications of rules in  $\mathbf{R}_{\rightarrow}$  preserve intuitionistic validity (that is, if the premiss-sequents are intuitionistically valid, so is the conclusion-sequent). But since we are dealing only with sequents with finite antecedents, the rules in  $\mathbf{R}_{\rightarrow}^f$  also preserve  $V$ -validity.  $\square$

**Theorem 6.8.** *There is a global model  $V$  of  $\mathbf{R}_{\sim}^f$  such that some  $v \in V$  does not obey  $(\mathbf{C}_{\sim})$ .*

*Proof.* Fix a language  $\mathcal{L}_{\sim}$  built up from atoms  $At = \{p_1, p_2, p_3, \dots\}$  and with  $\sim$  as the sole connective. Let  $I^f$  be the set of intuitionistic valuations that make only finitely many atoms true. Now we define a further valuation  $v_e$  as follows: say that an  $\mathcal{L}_{\sim}$ -sentence is even iff it is an atom with an even index or one such atom preceded by an even number of occurrences of ‘ $\sim$ ’; otherwise the sentence is odd. We let  $v_e(A) = 1$  iff  $A$  is even. Our global model is  $V = I^f \cup \{v_e\}$ . Note that:

(a)  $v_e \leq_V v$  iff  $v = v_e$ , since all valuations in  $I^f$  make only finitely many atoms true.

(b)  $v_e$  does not obey  $(\mathbf{C}_{\sim})$ , since, for any odd atom  $s$  we have that  $v_e \leq v'$  implies  $v'(s) = 0$ , but  $v_e(\sim s) = 0$ .

It remains to show that the rules in  $\mathbf{R}_{\sim}^f$  preserve  $V$ -validity. To show this we first need to prove that a sequent  $\Gamma \Rightarrow A$  with finite  $\Gamma$  is intuitionistically valid if and only if it is  $V$ -valid. The left-to-right direction is shown by contraposition: if  $\Gamma \Rightarrow A$  is not  $V$ -valid, then for some  $v \in V$  we have  $v(\gamma) = 1$  for all  $\gamma \in \Gamma$  but  $v(A) = 0$ . If  $v \in I^f$  we are done. If, on the other hand,  $v = v_e$ , then  $\Gamma$  contains only even sentences and  $A$  is odd, and it is easy to find a classical valuation in which  $\Gamma \Rightarrow A$  does not hold: if  $A$  is an atom  $s$  preceded by an even number of negations, the classical valuation that makes all atoms in  $\Gamma$  true and  $s$  false will do (since  $s$  itself must be an odd atom), and if  $A$  is an atom  $s$  preceded by an odd number of negations, the classical valuation that makes all atoms in  $\Gamma$  and  $s$  will do. But since  $\Gamma \Rightarrow A$  is classically invalid, it is also intuitionistically invalid. For the right-to-left direction, note that if  $\Gamma \Rightarrow A$  is intuitionistically invalid, then by Lemma 6.6 part (i) there must be some  $v \in I^f \subseteq V$  that invalidates it. The rest of the proof proceeds as in the previous theorem.  $\square$

The upshot so far is that Garson’s results don’t hold for the finitary version of his rules. But in fact we can strengthen the result, and show that no set of finitary, single conclusion rules for the connectives Garson considers expresses

his semantics. Let's take  $\rightarrow$  as an example. Since the global model  $V$  defined in Theorem 6.7 doesn't obey  $(\mathbf{C}_{\rightarrow})$ , it is enough to show that  $V$  is a global model of any set of finitary, single-conclusion rules for  $\rightarrow$ . Now, any set of rules for  $\rightarrow$  must preserve intuitionistic validity from premiss-sequents to conclusion-sequent. Suppose that some set of rules  $\mathbf{R}$  for  $\rightarrow$  doesn't preserve  $V$ -validity. Then there is an application of some  $R \in \mathbf{R}$  with  $V$ -valid premiss-sequents and a  $V$ -invalid conclusion-sequent. But in Theorem 6.7 we showed that a sequent  $\Gamma \Rightarrow \varphi$  with finite  $\Gamma$  is  $V$ -valid if and only if it is intuitionistically valid. Therefore, the premiss-sequents are intuitionistically valid and the conclusion-sequent isn't, and the rule in question doesn't preserve intuitionistic validity; in other words,  $\mathbf{R}$  isn't a set of rules for  $\rightarrow$ . A similar argument works for  $\sim$ , and therefore:

**Corollary 6.9.**

- (i) *No set of finitary rules that proves all and only the valid sequents of the  $\sim$ -fragment of intuitionistic logic expresses  $(\mathbf{C}_{\sim})$ .*
- (ii) *No set of finitary rules that proves all and only the valid sequents of the  $\rightarrow$ -fragment of intuitionistic logic expresses  $(\mathbf{C}_{\rightarrow})$ .*

This is a problem for Garson's project. On the face of it our inferential practice is finitary: we reason from finitely many assumptions to finitely many conclusions, using rules with finitely many premisses. If this is correct, though, Garson's semantic conditions  $\mathbf{C}_{\rightarrow}$  and  $\mathbf{C}_{\sim}$  are disconnected from our use of logical vocabulary, since no set of rules with finite antecedents can express them. Note, incidentally, that the claim that our use of logical vocabulary is finitary doesn't commit us, in and of itself, to any particular view about compactness. Paseau and Griffiths (2021), for example, have recently argued that 'the classical consequence relation of English' is non-compact, but even they grant that "we understand [a valid argument with infinitely many premisses] via a finite description, and convince ourselves of [its] validity by finite reasoning" (Paseau and Griffiths 2021, p. 194). Similarly, as we know, McGee (2000) has argued that rules can express a semantics for which they are not complete, so finitary rules can determine non-compact consequence relations.

Let's summarise our results. First, taking meanings to be determined by single-conclusion rules and reading semantics off of them using global models doesn't solve Carnap's Problem: no single-conclusion rules for the classical connectives globally expresses classical semantics, and no set of finitary rules globally expresses Garson's 'intuitionistic' semantics either. On the other hand, and much like in the previous section, if we use multiple-conclusion rules the intended meaning of classical connectives is pinned down. The obvious next question is: what about the quantifiers?

### 6.2.3 But what about the Quantifiers?

In this section we will look at Garson's take on Carnap's Problem for first-order languages. Now, Garson takes interpretations for quantified languages to

be (non-trivial) valuations, as in the propositional case. The problems start when he tries to define  $\forall$ -normal interpretations. He distinguishes between a ‘substitutional’ and an ‘objectual’ reading of quantifiers. On the substitutional reading an interpretation  $v$  is  $\forall$ -normal when:<sup>8</sup>

$$(s\forall) v(\forall x\varphi) = 1 \text{ iff } v(\varphi[y/x]) = 1 \text{ for all variables } y \text{ of the language.}$$

The question that immediately arises is whether  $(s\forall)$  is a plausible notion of *normal* interpretation. And arguably, it isn’t. A first-order semantics in terms of  $(s\forall)$  doesn’t so much as mention a domain of quantification, which puts it quite far from the usual way(s) of interpreting quantifiers. But the worry goes deeper. First,  $(s\forall)$  doesn’t really match the usual truth-conditions of quantified sentences: in a (normal) classical structure we may have that  $\mathcal{M}, \sigma \models Py$  for all variables  $y$  but  $\mathcal{M}, \sigma \not\models \forall xPx$ . And secondly, as Garson himself (2013, p. 215) points out,  $(s\forall)$  induces a non-compact consequence relation. To see this, fix a language with a unary predicate  $P$  and no individual constants or function symbols, and consider the argument from all sentences of the form  $Py$ , for  $y$  a variable of the language, to  $\forall xPx$ :

$$\{P[y/x] \mid y \in \text{Vars}\} \vdash \forall xPx$$

This argument is truth-preserving on all interpretations that satisfy  $(s\forall)$ , and therefore valid on Garson’s ‘normal’ substitutional semantics. The result of taking any finite subset of  $\{P[y/x] \mid y \in \text{Vars}\}$  as premises, on the other hand, is not. Needless to say, no standard proof-system for classical logic is strongly complete for this substitutional semantics.

The objectual reading of quantifiers is more involved. Garson begins by introducing the notion of a ‘hybrid formula’. Given a domain  $D$ , hybrid formulas are recursively defined so that any well-formed-formula is a hybrid formula, and the result  $\varphi[d/x]$  of substituting an object  $d \in D$  for all free occurrences of  $x$  in a hybrid formula  $\varphi$  is a hybrid formula itself. Garson then considers interpretations that range over hybrid formulas, and takes  $v$  to be  $\forall$ -normal, given a domain  $D$ , when:

$$(o\forall) v(\forall x\varphi) = 1 \text{ iff } v(\varphi[d/x]) = 1 \text{ for all } d \in D.$$

Although  $(o\forall)$  is, perhaps, more natural than  $(s\forall)$ , there is still reason to doubt its adequacy. We were after a definition of normal interpretations over a language  $\mathcal{L}$ . What we have been given, however, is a definition of normal interpretations over a ‘hybrid’ (i.e. extended) language  $\mathcal{L}'$ . To avoid this sort of problem Brîncuş (2024) restricts attention to countable domains all of whose elements are named by a constant, which effectively collapses Garson’s objectual and substitutional readings. But this doesn’t really solve the issue: he is just left without a notion of normal interpretation applicable to all languages.

---

<sup>8</sup>From now on, when discussing first-order languages lowercase Greek letters  $\varphi, \psi, \dots$  range over formulas, and uppercase Roman letters  $P, Q, \dots$  range over non-logical predicates.

The problem we are facing is easy to explain. In order to make Carnap's Question precise enough to be worth asking we have to clarify, among other things, what we mean by (normal) interpretations. There are several ways we can go about it, but not everything will do. Some sensible minimal requirements are that normal interpretations be among the usual ways of interpreting first-order languages, and that the definition of normality be general enough to apply to languages of arbitrary signature. Garson's proposals fail on both accounts, and this seems like a good reason to reject his way of formulating Carnap's Problem altogether.

In the rest of this section I will consider an alternative way to understand Carnap's Problem for first-order languages that stays relatively close to Garson's set-up, but doesn't run into the same problems. This is partly to explore what the first-order Carnap's Problem looks like when we define interpretations as valuations, and partly because the tools we will develop will be useful in the next section.

Let's start from the basics. When we talk about semantics for first-order logic, what comes to mind is the Tarskian definition of satisfaction of a formula  $\varphi$  under a variable assignment  $\sigma$  in a structure  $\mathcal{M} = (D, I)$ , where  $D$  is a domain and  $I$  a function interpreting the non-logical vocabulary. Given such an  $\mathcal{M}$ , the usual inductive definition is:

- (i)  $\mathcal{M}, \sigma \models P(x_1, \dots, x_n)$  iff  $(\sigma(x_1), \dots, \sigma(x_n)) \in I(P)$ .
- (ii)  $\mathcal{M}, \sigma \models \varphi \wedge \psi$  iff  $\mathcal{M}, \sigma \models \varphi$  and  $\mathcal{M}, \sigma \models \psi$ .
- (iii)  $\mathcal{M}, \sigma \models \neg\varphi$  iff  $\mathcal{M}, \sigma \not\models \varphi$ .
- (iv)  $\mathcal{M}, \sigma \models \forall x\varphi$  iff  $\mathcal{M}, \sigma[a/x] \models \varphi$  for all  $a \in D$ .

where  $\sigma[a/x]$  is the variable assignment that results from setting  $\sigma(x) = a$  and leaving  $\sigma$  otherwise unchanged.

These clauses don't define assignment of semantic values; they define a ternary relation  $\models$  between structures, variable assignments, and formulas. Still, there is an obvious sense in which, given some  $\mathcal{L}$ -structure  $\mathcal{M}$ , a satisfaction relation  $\models$  determines a class of valuations: just set  $v_\sigma^\mathcal{M}(\varphi) = 1$  if  $\mathcal{M}, \sigma \models \varphi$ , else  $v_\sigma^\mathcal{M}(\varphi) = 0$ , for any formula  $\varphi$ . As the notation suggests, this yields a valuation  $v_\sigma^\mathcal{M}$  for each variable assignment  $\sigma$  from  $\mathcal{L}$  to  $\mathcal{M}$ .

Using this idea, it is easy to set-up a valuational semantics for first-order logic. An interpretation (based on an  $\mathcal{L}$ -structure  $\mathcal{M}$ ) will be a set of valuations  $V^\mathcal{M} := \{v_\sigma \mid \sigma \in A^\mathcal{M}\}$ , where  $A^\mathcal{M}$  is the set of variable assignments from  $\mathcal{L}$  to  $\mathcal{M}$ . Since the meaning of non-logical expressions is not in question here, we will assume that atomic formulas are evaluated as usual i.e.:

$$v_\sigma(Qx_1, \dots, x_n) = 1 \text{ iff } \mathcal{M}, \sigma \models Qx_1, \dots, x_n$$

for any  $v_\sigma \in V^\mathcal{M}$  and n-ary predicate  $Q$ . Now, how should we define  $\forall$ -normal interpretations? The usual satisfaction clause for  $\forall$  is:

(iv)  $\mathcal{M}, \sigma \models \forall x\varphi$  iff  $\mathcal{M}, \sigma[a/x] \models \varphi$  for all  $a \in D$ .

In ‘valuation-speak’, this is to say that  $V^{\mathcal{M}}$  is  $\forall$ -normal when, for all  $v_\sigma \in V^{\mathcal{M}}$ :

( $\forall$ )  $v_\sigma(\forall x\varphi) = 1$  iff  $v_{\sigma[a/x]}(\varphi) = 1$  for all  $a \in D$ .

Finally, we have to explain what it is for interpretations to be compatible with rules. First, say that  $\Gamma \Rightarrow \Delta$  is  $V^{\mathcal{M}}$ -valid when it is satisfied by all  $v_\sigma \in V^{\mathcal{M}}$ . And secondly, say that  $V^{\mathcal{M}}$  is *consistent with a rule R* when, for all applications of R, if the premiss-sequents are  $V^{\mathcal{M}}$ -valid, so is the conclusion-sequent. In other words,  $V^{\mathcal{M}}$  is consistent with R if it is a global model of R.<sup>9</sup>

We now have a Garson-like way of posing Carnap’s Question in first-order settings: we want to know if (given any  $\mathcal{L}$ -structure  $\mathcal{M}$ ), all interpretations  $V^{\mathcal{M}}$  that are consistent with the rules for  $\forall$  are  $\forall$ -normal.

The answer to this question is ‘no’, regardless of whether we use single or multiple-conclusion rules, natural deduction or sequent calculi. For reasons that will become clear in the next section I will focus on the multiple-conclusion natural deduction rules from (Borićić 1985), but analogous results hold for any other standard choice. Here are the rules in question:

$$\frac{\Gamma \Rightarrow \varphi(y/x), \Delta}{\Gamma \Rightarrow \forall x\varphi, \Delta} (\forall I)^+ \qquad \frac{\Gamma \Rightarrow \forall x\varphi, \Delta}{\Gamma \Rightarrow \varphi(y/x), \Delta} (\forall E)^+$$

As usual  $y$  must be free for  $x$  in  $\varphi$ ,  $\varphi(y/x)$  is the result of replacing all free occurrences of  $x$  in  $\varphi$  for  $y$ , in ( $\forall I$ )  $y$  must not occur free in  $\Gamma$ ,  $\forall x\varphi$ , or  $\Delta$ , and the context-sets are finite.

Our goal is to prove that some non  $\forall$ -normal interpretation is consistent with these rules. First, let’s fix a language  $\mathcal{L}$  with a single binary predicate  $R$  as non-logical vocabulary, and consider the  $\mathcal{L}$ -structure  $\mathcal{B} = (D, I)$  with  $D = \mathbb{Z} - \{0\}$  and  $I(R) = <$ . To get a non-normal interpretation  $V^{\mathcal{B}}$  based on  $\mathcal{B}$ , we will use a second model  $\mathcal{A} = (D', I')$  with  $D' = \mathbb{Z}$  and  $I'(R) = <$ . Just define  $V^{\mathcal{B}}$  by letting  $v_\sigma(\varphi) = 1$  if and only if  $\mathcal{A}, \sigma \models \varphi$  (for all  $\sigma \in A^{\mathcal{B}}$ , where ‘ $\models$ ’ is the usual satisfaction relation).

To see that  $V^{\mathcal{B}}$  is not  $\forall$ -normal, take any  $\sigma \in A^{\mathcal{B}}$  such that  $\sigma(x) = -1$  and  $\sigma(y) = 1$ . By definition of  $V^{\mathcal{B}}$  we have that:

$$\begin{aligned} v_{\sigma[a/z]}(\neg(Rxz \wedge Rzy)) &= 1 \text{ for all } a \in D = \mathbb{Z} - \{0\}, \\ \text{but } v_\sigma(\forall z \neg(Rxz \wedge Rzy)) &= 0. \end{aligned}$$

<sup>9</sup>Note that it wouldn’t have done to define consistency with a rule in terms of preservation of satisfaction, rather than validity, because the rules for  $\forall$  don’t preserve satisfaction in an assignment: in a normal model we may have that  $\mathcal{M}, \sigma \models P(y)$  but  $\mathcal{M}, \sigma \not\models \forall xPx$ . Garson makes a similar point in (Garson 2013, p. 43).

It remains to show that  $V^{\mathcal{B}}$  is consistent with  $(\forall I)^+$  and  $(\forall E)^+$ . To do this we need some new terminology and two quick lemmas. First, given some  $\mathcal{L}$ -structure  $\mathcal{M}$ , we will say that  $\Gamma \Rightarrow \Delta$  is  $\models$ -valid if, for any assignment  $\sigma \in A^{\mathcal{M}}$ , we have that  $\mathcal{M}, \sigma \models \Gamma$  implies that  $\mathcal{M}, \sigma \models \psi$  for some  $\psi \in \Delta$ . The required lemmas are:

**Lemma 6.10.** *For any  $\sigma \in A^{\mathcal{A}}$ , let  $\sigma^{+n}$  be the assignment such that  $\sigma^{+n}(x) = \sigma(x) + n$  for all variables  $x$ . Then for all  $\sigma \in A^{\mathcal{A}}$ ,  $n$ , and  $\varphi$ :*

$$\mathcal{A}, \sigma \models \varphi \text{ iff } \mathcal{A}, \sigma^{+n} \models \varphi$$

*Proof.* By induction on the complexity of formulas. □

**Lemma 6.11.** *Let  $\Gamma, \Delta$  be finite. If, for all  $\sigma \in A^{\mathcal{A}}$  with range in  $\mathbb{Z} - \{0\}$ , we have that  $\mathcal{A}, \sigma \models \Gamma$  implies  $\mathcal{A}, \sigma \models \psi$  for some  $\psi \in \Delta$ , then  $\Gamma \Rightarrow \Delta$  is  $\models$ -valid.*

*Proof.* Let  $\Gamma, \Delta$  be finite and suppose that, for all  $\sigma \in A^{\mathcal{A}}$  with range in  $\mathbb{Z} - \{0\}$ , we have that  $\mathcal{A}, \sigma \models \Gamma$  implies  $\mathcal{A}, \sigma \models \psi$  for some  $\psi \in \Delta$ . We want to show that  $\Gamma \Rightarrow \Delta$  is  $\models$ -valid, that is, we want to show that for all  $\sigma' \in A^{\mathcal{A}}$ , we have that  $\mathcal{A}, \sigma' \models \Gamma$  implies  $\mathcal{A}, \sigma' \models \psi$  for some  $\psi \in \Delta$ . We will do this by contraposition.

Suppose there is some  $\sigma' \in A^{\mathcal{A}}$  such that  $\mathcal{A}, \sigma' \models \Gamma$  but  $\mathcal{A}, \sigma' \not\models \psi$  for all  $\psi \in \Delta$ . We now construct a  $\sigma^* \in A^{\mathcal{A}}$  with range in  $\mathbb{Z} - \{0\}$  such that  $\mathcal{A}, \sigma^* \models \Gamma$  but  $\mathcal{A}, \sigma^* \not\models \psi$  for all  $\psi \in \Delta$ , contrary to our assumption.

Let  $x_1 \dots x_n$  be the (finitely many!) variables that occur free in  $\Gamma, \Delta$ . If  $\sigma'(x_i) \neq 0$  for all  $i \leq n$ , we just let  $\sigma^*(x_i) = \sigma'(x_i)$  for all  $i \leq n$  and  $\sigma^*(z) = 1$  for all other variables  $z$ . Then  $\mathcal{A}, \sigma^* \models \Gamma$  but  $\mathcal{A}, \sigma^* \not\models \psi$  for all  $\psi \in \Delta$ , since  $\sigma'$  and  $\sigma^*$  agree on all the free variables in  $\Gamma, \Delta$ .

If, on the other hand,  $\sigma'(x_i) = 0$  for some  $i \leq n$ , let  $k = \text{Max}\{|\sigma'(x_1)| \dots |\sigma'(x_n)|\} + 1$ , where  $|\cdot|$  is the absolute value function. Then  $\sigma'^{+k}(x_i) \neq 0$  for all  $i \leq n$ . Moreover, since we have that  $\mathcal{A}, \sigma' \models \Gamma$  but  $\mathcal{A}, \sigma' \not\models \psi$  for all  $\psi \in \Delta$ , by Lemma 6.10 we also have that  $\mathcal{A}, \sigma'^{+k} \models \Gamma$  but  $\mathcal{A}, \sigma'^{+k} \not\models \psi$  for all  $\psi \in \Delta$ . But then it is enough to let  $\sigma^*(x_i) = \sigma'^{+k}(x_i)$  for all  $i \leq n$  and  $\sigma^*(z) = 1$  for all other variables  $z$ . We then have that  $\mathcal{A}, \sigma^* \models \Gamma$  but  $\mathcal{A}, \sigma^* \not\models \psi$  for all  $\psi \in \Delta$ , since  $\sigma'^{+k}$  and  $\sigma^*$  agree on all the free variables in  $\Gamma, \Delta$  □

We can now show that  $V^{\mathcal{B}}$  is consistent with  $(\forall I)^+$  and  $(\forall E)^+$ .

**Theorem 6.12.**  *$V^{\mathcal{B}}$  is consistent with  $(\forall I)^+$  and  $(\forall E)^+$*

*Proof.*

$(\forall I)^+$ : Let  $y$  be free for  $x$  in  $\varphi$  and not occur free in  $\Gamma, \forall x\varphi$ , or  $\Delta$ , and suppose that  $\Gamma \Rightarrow \varphi(y/x), \Delta$  is  $V^{\mathcal{B}}$ -valid. We want to show that  $\Gamma \Rightarrow \forall x\varphi, \Delta$  is  $V^{\mathcal{B}}$ -valid too. Now, since  $\Gamma \Rightarrow \varphi(y/x), \Delta$  is  $V^{\mathcal{B}}$ -valid, that is, since for any  $\sigma \in A^{\mathcal{B}}$ :

$$v_{\sigma}(\Gamma) = 1 \text{ implies } v_{\sigma}(\psi) = 1 \text{ for some } \psi \in \Delta \cup \{\varphi(y/x)\}$$

we also have, by the definition of  $V^{\mathcal{B}}$ , that for any  $\sigma \in A^{\mathcal{A}}$  with range in  $\mathbb{Z} - \{0\}$ :

$A, \sigma \models \Gamma$  implies  $A, \sigma \models \psi$  for some  $\psi \in \Delta \cup \{\varphi(y/x)\}$

But then by Lemma 6.11,  $\Gamma \Rightarrow \varphi(y/x), \Delta$  is  $\models$ -valid in  $\mathcal{A}$ . Since  $(\forall I)^+$  clearly preserves  $\models$ -validity (remember that ‘ $\models$ ’ is the usual satisfaction relation), we have that  $\Gamma \Rightarrow \forall x\varphi, \Delta$  is also  $\models$ -valid in  $\mathcal{A}$ . Therefore, by the definition of  $V^{\mathcal{B}}$ , we have that for any  $\sigma' \in A^{\mathcal{B}} \subset A^{\mathcal{A}}$ :

$v_{\sigma'}(\Gamma) = 1$  implies  $v_{\sigma'}(\psi) = 1$  for some  $\psi \in \Delta \cup \{\forall x\varphi\}$

In other words,  $\Gamma \Rightarrow \forall x\varphi, \Delta$  is  $V^{\mathcal{B}}$ -valid, as was to be shown.

$(\forall E)^+$ : Analogous to the previous case. □

It is easy to see that the same reasoning applies to the single-conclusion versions of  $(\forall I)^+$  and  $(\forall E)^+$ , as well as to other standard choices of single and multiple-conclusion sequent rules for  $\forall$ .<sup>10</sup> Our discussion can therefore be summarised as follows: Garson’s way to state Carnap’s Problem for first-order settings is inadequate, since his notion of a normal interpretation is neither recognisably normal nor applicable to languages of arbitrary signature. There is a closely related way of formulating Carnap’s Problem which retains the use of valuations and global models. On this way of understanding Carnap’s Problem, however, the usual rules for  $\forall$  don’t pin down normal interpretations.

### 6.3 Smiley: Bilateral Constraints

In this section we will look at bilateral approaches to Carnap’s Problem. We will start off gently, focussing on the propositional case. Our first goal, as usual, is getting clearer about how to formulate Carnap’s Problem. We saw in Chapter 4 that the formal languages bilateralists use contain signed formulas  $+A$  and  $-B$ , where ‘ $+$ ’ and ‘ $-$ ’ are markers for assertion and rejection.<sup>11</sup> How should we define interpretations for these languages? We will take a cue from Incurvati and Smith (2010) and let interpretations for a signed language  $\mathcal{L}_s$  be valuations from the *unsigned* part of each formula to  $\{0,1\}$ . The idea, as they put it, is that “it is propositional contents that are the primary locus of evaluation” (Incurvati and Smith 2010, p. 9); roughly, it is the propositional content of an assertion or rejection that can be true or false. As usual, an interpretation is  $\diamond$ -normal for some connective  $\diamond$  if it respects the usual truth-table for  $\diamond$ , and normal if it is  $\diamond$ -normal for all of our connectives.

Next we need to explain what inferential roles are. In this case it doesn’t matter much whether we use rules or consequence relations. We will eventually

<sup>10</sup>For instance, any set of rules interderivable with  $(\forall I)^+$  and  $(\forall E)^+$  has the same global models, so  $V^{\mathcal{B}}$  will be consistent with it too.

<sup>11</sup>Here we temporarily go back to the notational conventions of Chapter 4: uppercase Roman letters range over unsigned propositional formulas, and lowercase Greek letters over signed ones.

look at both options, but let's start with the simpler one: for now the inferential role of classical connectives is given by the classical *bilateral* consequence relation of their language. To make things more concrete, fix a signed propositional language  $\mathcal{L}_s$  with the usual connectives  $\wedge, \neg, \vee$  and  $\rightarrow$ . Then the classical bilateral consequence relation  $\vdash_{\mathcal{L}_s}^b$  on  $\mathcal{L}_s$  is, as we know from Chapter 4, the derivability relation of calculi  $\mathcal{R}, \mathcal{HB}_1$  and  $\mathcal{HB}_2$ .

Finally, we need to explain which valuations are compatible with  $\vdash_{\mathcal{L}_s}^b$ . Once again we will follow (Smiley 1996) and (Incurvati and Smith 2010), though somewhat loosely. First, say that a positively signed formula  $+A$  is *correct* under a valuation  $v$  if  $v(A) = 1$ , and incorrect otherwise; dually, a negatively signed formula  $-A$  is correct under  $v$  if  $v(A) = 0$ , and otherwise incorrect. Next, a signed sequent  $\Gamma \Rightarrow \varphi$  (or inference  $\Gamma \vdash_{\mathcal{L}_s}^b \varphi$ ) is *bilaterally satisfied by*  $v$  ('satisfied<sub>b</sub> by  $v$ ' for short) when  $\varphi$  is correct under  $v$  if  $\Gamma$  is correct under  $v$ . And finally,  $v$  is *bilaterally consistent* ('consistent<sub>b</sub>') with  $\vdash_{\mathcal{L}_s}^b$  if it makes all inferences in  $\vdash_{\mathcal{L}_s}^b$  correct.

We have put together a bilateralist way of formulating Carnap's Question: we want to know if all valuations consistent<sub>b</sub> with  $\vdash_{\mathcal{L}_s}^b$  are normal. The question can be tackled directly, but it will be more illuminating to coax an answer out of results that we have already encountered.

The crucial observation is that, from a certain point of view, the bilateral single-conclusion consequence relation  $\vdash_{\mathcal{L}}^b$  is a *unilateral multiple-conclusion* consequence relation in disguise.<sup>12</sup> To better explain this let's adopt a new convention: instead of writing bilateral inferences as  $\Gamma \vdash_{\mathcal{L}}^b \varphi$ , we will indicate the positively and negatively signed premisses separately, that is, we will write things like  $+\Gamma, -\Delta \vdash_{\mathcal{L}}^b \varphi$ . Then as (Rumfitt 2000) notes:<sup>13</sup>

**Observation 6.13.** Let  $\mathcal{L}$  be the unsigned version of  $\mathcal{L}_s$ , let  $\vdash_{\mathcal{L}}$  be the classical, multiple-conclusion consequence relation on  $\mathcal{L}$ , and let  $v$  be any valuation. Then:

- (i)  $+\Gamma, -\Delta \vdash_{\mathcal{L}_s}^b +A$  iff  $\Gamma \vdash_{\mathcal{L}} A, \Delta$ , and  
 $v$  satisfies<sub>b</sub>  $+\Gamma, -\Delta \vdash_{\mathcal{L}_s}^b +A$  iff  $v$  satisfies  $\Gamma \vdash_{\mathcal{L}} A, \Delta$ .
- (ii)  $+\Gamma, -\Delta \vdash_{\mathcal{L}_s}^b -A$  iff  $\Gamma, A \vdash_{\mathcal{L}} \Delta$ , and  
 $v$  satisfies<sub>b</sub>  $+\Gamma, -\Delta \vdash_{\mathcal{L}_s}^b -A$  iff  $v$  satisfies  $\Gamma, A \vdash_{\mathcal{L}} \Delta$ .

Observation 6.13 allows us to translate back and forward between bilateral and unilateral settings to prove different results. Let's take  $\neg$  as an example. By Observation 6.13 a valuation  $v$  satisfies<sub>b</sub>  $-A \vdash_{\mathcal{L}_s}^b +\neg A$  if and only if it satisfies  $\vdash_{\mathcal{L}} \neg A, A$ , and similarly for  $+\neg A \vdash_{\mathcal{L}_s}^b -A$  and  $\neg A, A \vdash_{\mathcal{L}}$ . But clearly,  $v$  satisfies

<sup>12</sup>A word of warning: the connection between bilateral single-conclusions and unilateral multiple-conclusions I am about to explain is important at a technical level, but doesn't bear on the philosophical issue of whether assertion and rejection are distinct speech acts. Indeed, we could have just as well phrased the point by saying that unilateral multiple-conclusions are bilateral single-conclusions in disguise. For more on bilateralist readings of multiple conclusions see e.g. (Restall 2005).

<sup>13</sup>The observation goes back to at least (Shoemith and Smiley 1978).

$\vdash_{\mathcal{L}} \neg A, A$  and  $\neg A, A \vdash_{\mathcal{L}}$  if and only if it is  $\neg$ -normal. Therefore, all valuations consistent<sub>b</sub> with  $\vdash_{\mathcal{L}_s}^b$  must be  $\neg$ -normal. The point can be straightforwardly extended to the rest of connectives. We know from our discussion of Observation 6.2, that the multiple-conclusion consequence relation  $\vdash_{\mathcal{L}}$  is consistent with all and only normal valuations. But then by Observation 6.13 we have that:

**Observation 6.14.**  $\vdash_{\mathcal{L}_s}^b$  is consistent<sub>b</sub> with all and only normal valuations.

We can also replicate Garson's results about local and global models of multiple-conclusion rules. For the local case, say that  $v$  is a *bilateral local model* ('local model<sub>b</sub>' for short) of a rule  $\mathbf{R}$  when, in any application of  $\mathbf{R}$ , if the premiss-sequents are satisfied<sub>b</sub> so is the conclusion sequent, and say a set of rules  $\mathbf{R}$  *bilaterally locally expresses* ('locally expresses<sub>b</sub>') a condition  $C$  on valuations if all its local models<sub>b</sub> obey  $C$ . By Observation 6.13, given any single-conclusion bilateral rule  $\mathbf{R}$  there is a unilateral, multiple-conclusion rule  $\mathbf{R}'$  that locally expresses the same condition on valuations. The bilateral introduction rules  $(+\neg \text{I})$  and  $(-\neg \text{I})$  from Chapter 4, for instance, have exactly the same local models as these familiar multiple-conclusion sequent rules for  $\neg$ :

$$\frac{+\Gamma, -\Delta \Rightarrow -A}{\Gamma, -\Delta \Rightarrow +\neg A} (+\neg \text{I}) \quad \rightsquigarrow \quad \frac{\Gamma, A \Rightarrow \Delta}{\Gamma \Rightarrow \neg A, \Delta} (\neg \text{R})^+$$

$$\frac{+\Gamma, -\Delta \Rightarrow +A}{\Gamma, -\Delta \Rightarrow -\neg A} (-\neg \text{I}) \quad \rightsquigarrow \quad \frac{\Gamma \Rightarrow A, \Delta}{\Gamma, \neg A \Rightarrow \Delta} (\neg \text{L})^+$$

Similarly, coordination principles have the same local models as certain structural rules which are satisfied in all valuations. For example, applications of *Explosion* correspond to a combination of Cut and Weakening:

$$\frac{+\Gamma, -\Delta \Rightarrow -A \quad +\Gamma', -\Delta' \Rightarrow +A}{+\Gamma, +\Gamma', -\Delta, -\Delta' \Rightarrow +B} (\text{ex}) \quad \rightsquigarrow \quad \frac{\Gamma, A \Rightarrow \Delta \quad \Gamma', \Rightarrow A, \Delta'}{\Gamma, \Gamma' \Rightarrow B, \Delta, \Delta'}$$

$$\frac{+\Gamma, -\Delta \Rightarrow -A \quad +\Gamma', -\Delta' \Rightarrow +A}{+\Gamma, +\Gamma', -\Delta, -\Delta' \Rightarrow -B} (\text{ex}) \quad \rightsquigarrow \quad \frac{\Gamma, A \Rightarrow \Delta \quad \Gamma' \Rightarrow A, \Delta'}{\Gamma, \Gamma', B \Rightarrow \Delta, \Delta'}$$

It is routine to check that the operational rules governing each connective in  $\mathcal{R}$ ,  $\mathcal{HB}_1$  or  $\mathcal{HB}_1$ , together with the coordination principles of their respective calculi, locally express<sub>b</sub> the same conditions as any standard set of multiple-conclusion rules for the same connective. In other words, we have the following analogue of Observation 6.2:

**Observation 6.15.** The bilateral rules for  $\wedge$ ,  $\neg$ ,  $\vee$  and  $\rightarrow$  locally express<sub>b</sub>  $(\wedge)$ ,  $(\neg)$ ,  $(\vee)$  or  $(\rightarrow)$ , respectively.

A similar strategy covers Garson’s results on global expression. First, say that a sequent  $+\Gamma, -\Delta \Rightarrow \varphi$  is *bilaterally valid* in a set  $V$  of valuations (‘ $V$ -valid<sub>*b*</sub>’ for short) if it is satisfied<sub>*b*</sub> by all  $v \in V$ . Secondly, say that  $V$  is a *bilateral global model* of a rule  $R$  (‘global model<sub>*b*</sub> of  $R$ ’) when, for any application of  $R$ , if the premiss-sequents are  $V$ -valid<sub>*b*</sub> so is the conclusion-sequent. And finally, say that a set of rules  $\mathbf{R}$  *bilaterally globally expresses* (‘globally expresses<sub>*b*</sub>’) a condition  $C$  on valuations if all its global models<sub>*b*</sub> obey  $C$ .

By the same reasoning as before, given any single conclusion bilateral rule  $R$  there is a unilateral, multiple conclusion rule  $R'$  with exactly the same global models. It is again routine to check that the operational rules governing each connective in  $\mathcal{R}$ ,  $\mathcal{HB}_1$  or  $\mathcal{HB}_2$ , together with the coordination principles of their calculi, globally express<sub>*b*</sub> the same conditions as any standard set of multiple conclusion rules for the same connective. This gives us an analogue of Observation 6.3:

**Observation 6.16.** The bilateral rules for  $\wedge$ ,  $\neg$ ,  $\vee$  and  $\rightarrow$  globally express<sub>*b*</sub>  $(\wedge)$ ,  $(\neg)$ ,  $(\vee)$  or  $(\rightarrow)$ , respectively.

Now that we have a clear idea of what happens in the propositional case, we are ready to move on to quantifiers. Surprisingly, there is almost no bilateralist literature concerning the first-order version of Carnap’s Problem.<sup>14</sup> Luckily, in Section 6.2.3 we saw a way of formulating Carnap’s Problem that is easy to transpose to our new setting. Let’s get to it.

An interpretation (based on an  $\mathcal{L}$ -structure  $\mathcal{M}$ ) is a set of valuations  $V^{\mathcal{M}} := \{v_\sigma \mid \sigma \in A^{\mathcal{M}}\}$ , where  $A^{\mathcal{M}}$  is the set of variable assignments from  $\mathcal{M}$  to  $\mathcal{L}$ . We assume that atomic formulas are interpreted as usual, and say that  $V^{\mathcal{M}}$  is  $\forall$ -normal when, for all  $v_\sigma \in V^{\mathcal{M}}$ :

$$(\forall) \ v_\sigma(\forall x\varphi) = 1 \text{ iff } v_{\sigma[a/x]}(\varphi) \text{ for all } a \in D.$$

Explaining what it is for interpretations to be compatible with bilateral rules is smooth sailing. In the unilateral case we said that  $V^{\mathcal{M}}$  was consistent with a rule  $R$  when it was a global model of  $R$ . In the bilateral setting, we will say that  $V^{\mathcal{M}}$  is *bilaterally consistent* with a rule  $R$  (‘consistent<sub>*b*</sub> with  $R$ ’) if it is a global model<sub>*b*</sub> of  $R$ .<sup>15</sup>

Now we just need to settle on some bilateral rules for  $\forall$ . The obvious choice is:<sup>16</sup>

$$\frac{+\Gamma, -\Delta \Rightarrow +\varphi(y/x)}{+\Gamma, -\Delta \Rightarrow +\forall x\varphi} \text{ (+}\forall \text{ I)} \qquad \frac{+\Gamma, -\Delta \Rightarrow +\forall x\varphi}{+\Gamma, -\Delta \Rightarrow +\varphi(y/x),} \text{ (+}\forall \text{ E)}$$

<sup>14</sup>The only exception I am aware of is (Button and Walsh 2018). Button and Walsh’s interpretations are boolean-valued models (*very* roughly, many-valued valuations), and explaining their results would take us too far afield. For more details, see (Button and Walsh 2018, Ch. 13).

<sup>15</sup>Recall that  $V^{\mathcal{M}}$  is a global model<sub>*b*</sub> of  $R$  when, in any application of  $R$ , if the premiss-sequents are  $V^{\mathcal{M}}$ -valid<sub>*b*</sub>, so is the conclusion.

<sup>16</sup>Now we go back to the notational conventions of first-order languages: uppercase Roman letters range over non-logical predicates, and lowercase Greek letters over (unsigned) formulas.

where as usual,  $y$  must be free for  $x$  in  $\varphi$ ,  $\varphi(y/x)$  is the result of replacing all free occurrences of  $x$  in  $\varphi$  for  $y$ , and in  $(+\forall \text{ I})$  the variable  $y$  must not occur free in  $\Gamma$ ,  $\forall x\varphi$ , or  $\Delta$ .

Just like in the propositional case, we need to supplement  $(+\forall \text{ I})$  and  $(+\forall \text{ E})$  with coordination principles. Any of the sets of coordination principles from Chapter 4 would do, but let's settle on *Explosion* and *Bilateral Excluded Middle* for concreteness. In the new notation, they are:

$$\frac{+\Gamma, -\Delta \Rightarrow +\varphi \quad +\Gamma', -\Delta' \Rightarrow -\varphi}{+\Gamma, +\Gamma', -\Delta, -\Delta' \Rightarrow \pm\psi} \text{ (ex)}$$

$$\frac{+\Gamma, -\Delta, +\varphi \Rightarrow \pm\psi \quad +\Gamma', -\Delta', -\varphi \Rightarrow \pm\psi}{+\Gamma, +\Gamma', -\Delta, -\Delta' \Rightarrow \pm\psi} \text{ (bem)}$$

The rules  $\mathbf{R}_\forall$  for  $\forall$  are (ex), (bem),  $(+\forall \text{ I})$  and  $(+\forall \text{ E})$ . It may seem strange that we haven't given rules for negatively signed formulas. However, the following rejective rules are derivable from  $\mathbf{R}_\forall$ :<sup>17</sup>

$$\frac{+\Gamma, -\Delta \Rightarrow -\varphi(y/x)}{+\Gamma, -\Delta \Rightarrow -\forall x\varphi} \text{ } (-\forall \text{ I}) \quad \frac{+\Gamma', -\Delta' \Rightarrow -\forall x\varphi \quad +\Gamma, -\Delta, -\varphi(y/x) \Rightarrow \pm\psi}{+\Gamma, +\Gamma', -\Delta, -\Delta' \Rightarrow \pm\psi} \text{ } (-\forall \text{ E})$$

where as usual,  $y$  must be free for  $x$  in  $\varphi$ ,  $\varphi(y/x)$  is the result of replacing all free occurrences of  $x$  in  $\varphi$  for  $y$ , and in  $(-\forall \text{ E})$  the variable  $y$  must not occur free in  $\Gamma$ ,  $\forall x\varphi$ , or  $\Delta$ . Taking  $(-\forall \text{ I})$  and  $(-\forall \text{ E})$  as derived rather than primitive will simplify our proofs, but the same results apply if we include them in  $\mathbf{R}_\forall$ .

We finally have all the pieces to pose Carnap's Question: we want to know if all valuations consistent<sub>b</sub> with  $\mathbf{R}_\forall$  are  $\forall$ -normal. Once again, the question can be answered using results we have already encountered. First note that, in line with Observation 6.13, we have that:

**Observation 6.17.** For any set of valuations  $V$ :

- (i)  $+\Gamma, -\Delta \Rightarrow +\varphi$  is  $V$ -valid<sub>b</sub> iff  $\Gamma \Rightarrow \varphi$ ,  $\Delta$  is  $V$ -valid.
- (ii)  $+\Gamma, -\Delta \Rightarrow -\varphi$  is  $V$ -valid<sub>b</sub> iff  $\Gamma, A \Rightarrow \Delta$  is  $V$ -valid.

Using Observation 6.17 it is easy to find unilateral multiple-conclusion rules that have the same global models as the rules in  $\mathbf{R}_\forall$ . Coordination principles still correspond to structural rules that are valid in all sets of valuations:

<sup>17</sup>It's not obvious in our hideous notation, but  $(-\forall \text{ I})$  and  $(-\forall \text{ E})$  are analogues of the usual unilateral rules for  $\exists$ , bringing out the classical duality between the quantifiers.

$$\begin{array}{c}
\frac{+\Gamma, -\Delta \Rightarrow -\varphi \quad +\Gamma', -\Delta' \Rightarrow +\varphi}{+\Gamma, +\Gamma', -\Delta, -\Delta' \Rightarrow +\psi} \text{ (ex)} \quad \rightsquigarrow \quad \frac{\Gamma, \varphi \Rightarrow \Delta \quad \Gamma' \Rightarrow \varphi, \Delta'}{\Gamma, \Gamma' \Rightarrow \psi, \Delta, \Delta'} \\
\\
\frac{+\Gamma, -\Delta \Rightarrow -\varphi \quad +\Gamma', -\Delta' \Rightarrow +\varphi}{+\Gamma, +\Gamma', -\Delta, -\Delta' \Rightarrow -\psi} \text{ (ex)} \quad \rightsquigarrow \quad \frac{\Gamma, \varphi \Rightarrow \Delta \quad \Gamma' \Rightarrow \varphi, \Delta'}{\Gamma, \Gamma', \psi \Rightarrow \Delta, \Delta'} \\
\\
\frac{+\Gamma, -\Delta, -\varphi \Rightarrow +\psi \quad +\Gamma', -\Delta', +\varphi \Rightarrow +\psi}{+\Gamma, +\Gamma', -\Delta, -\Delta' \Rightarrow +\psi} \text{ (bem)} \quad \rightsquigarrow \quad \frac{\Gamma \Rightarrow \varphi, \psi, \Delta, \quad \Gamma', \varphi \Rightarrow \psi, \Delta'}{\Gamma, \Gamma' \Rightarrow \psi, \Delta, \Delta'} \\
\\
\frac{+\Gamma, -\Delta, -\varphi \Rightarrow -\psi \quad +\Gamma', -\Delta', +\varphi \Rightarrow -\psi}{+\Gamma, +\Gamma', -\Delta, -\Delta' \Rightarrow -\psi} \text{ (bem)} \quad \rightsquigarrow \quad \frac{\Gamma, \psi \Rightarrow \varphi, \Delta, \quad \Gamma', \varphi, \psi \Rightarrow \Delta'}{\Gamma, \Gamma' \Rightarrow \psi, \Delta, \Delta'}
\end{array}$$

More interestingly,  $(+\forall \text{ I})$  and  $(+\forall \text{ E})$  correspond to Boričić's multiple-conclusion natural deduction rules  $(\forall \text{ I})^+$  and  $(\forall \text{ E})^+$  from Section 6.2.3:

$$\begin{array}{c}
\frac{+\Gamma, -\Delta \Rightarrow +\varphi(y/x)}{+\Gamma, -\Delta \Rightarrow +\forall x\varphi} \text{ (+}\forall \text{ I)} \quad \rightsquigarrow \quad \frac{\Gamma \Rightarrow \varphi(y/x), \Delta}{\Gamma \Rightarrow \forall x\varphi, \Delta} \text{ (}\forall \text{ I)}^+ \\
\\
\frac{+\Gamma, -\Delta \Rightarrow +\forall x\varphi}{+\Gamma, -\Delta \Rightarrow +\varphi(y/x)} \text{ (+}\forall \text{ E)} \quad \rightsquigarrow \quad \frac{\Gamma \Rightarrow \forall x\varphi, \Delta}{\Gamma \Rightarrow \varphi(y/x), \Delta} \text{ (}\forall \text{ E)}^+
\end{array}$$

Now, Theorem 6.12 shows that the non-normal interpretation  $V^{\mathcal{B}}$  is consistent with Boričić's rules  $(\forall \text{ I})^+$  and  $(\forall \text{ E})^+$ . But since these rules have the same global models as  $(+\forall \text{ I})$  and  $(+\forall \text{ E})$ , the non-normal interpretation  $V^{\mathcal{B}}$  is also consistent<sub>b</sub> with  $(+\forall \text{ I})$  and  $(+\forall \text{ E})$ . Officially, then:

**Theorem 6.18.**  $V^{\mathcal{B}}$  is consistent<sub>b</sub> with  $(+\forall \text{ I})$  and  $(+\forall \text{ E})$ .

It is also easy to see that Theorem 6.18 implies that  $V^{\mathcal{B}}$  satisfies<sub>b</sub> all sequents provable from  $\mathbf{R}_{\forall}$ , so attempting to read the usual semantics off the bilateral consequence relation that our rules give rise to, rather than from the rules themselves, would fail as well. All in all, then, the situation can be put as follows: when it comes to Carnap's Problem, bilateral approaches have the same effect as unilateral approaches that make use of multiple-conclusions. This means that in a propositional setting the normal semantics is pinned down by the bilateral consequence relation, as well as by the bilateral rules for the connectives. In first-order settings, however, bilateral rules of inference can't secure the normal interpretation of the quantifiers.

## 6.4 McGee: Open-Endedness

In this section we will examine Vann McGee’s (2000) approach to Carnap’s Problem. His take is conceptually quite different to those we have encountered until now, so it will be useful to start off with an overview of his position. According to McGee the meaning of logical vocabulary is determined by the rules of inference governing its use. Crucially, he thinks those rules are *open-ended*, meaning that they remain valid “however our language may be enriched by the addition of new sentences” (McGee 2000, p. 66), or “in every mathematically possible extension of our language” (2000, p. 70). Here is the idea at play:

We accept *reductio ad absurdum* because its validity follows from the meaning of classical ‘ $\neg$ ’. We do not accept it because we have surveyed the forms of expression found in English and found that its expressive power is circumscribed in such a way as to validate the rule. When we introduce a new predicate into the language [...] we do not have to inquire whether *reductio ad absurdum* remains valid for inferences involving the new predicate. (2000, p. 66)

McGee also claims that, for any class of models, there is a ‘mathematically possible extension of our language’ which contains a sentence that is true exactly in those models. He then argues that any non-normal interpretation of a language  $\mathcal{L}$  invalidates some classically valid inference when extended to a richer language  $\mathcal{L}'$ . Non-normal interpretations, therefore, are ruled out on the grounds that they violate open-endedness.

Our first goal in this section is to make this more precise. We will begin by looking at McGee’s rules. He adopts a proof-system from (Mates 1972) in which only sentences (closed formulas) occur in derivations. Let  $\Gamma$  range over (finite) sets of sentences,  $\varphi$  and  $\psi$  range over sentences, and  $\theta$  range over formulas with one variable  $x$  free. Then the calculus in question —call it  $\mathcal{C}$ — is:

$$\begin{array}{c} \frac{}{\Gamma, \varphi \Rightarrow \varphi} \text{ (Id)} \qquad \frac{\Gamma, \varphi \Rightarrow \psi \quad \Gamma' \Rightarrow \varphi}{\Gamma, \Gamma' \Rightarrow \psi} \text{ (Cut)} \\ \\ \frac{}{\varphi, \psi \Rightarrow \varphi \wedge \psi} \text{ (\wedge I)} \qquad \frac{}{\varphi \wedge \psi \Rightarrow \varphi} \text{ (\wedge E}_1\text{)} \qquad \frac{}{\varphi \wedge \psi \Rightarrow \psi} \text{ (\wedge E}_2\text{)} \\ \\ \frac{\Gamma, \varphi \Rightarrow \perp}{\Gamma \Rightarrow \neg\varphi} \text{ (\neg I)} \qquad \frac{}{\neg\neg\varphi \Rightarrow \varphi} \text{ (\neg E}_1\text{)} \qquad \frac{}{\varphi, \neg\varphi \Rightarrow \psi} \text{ (\neg E}_2\text{)} \\ \\ \frac{\Gamma \Rightarrow \theta(c/x)}{\Gamma \Rightarrow \forall x\theta} \text{ (\forall I)} \qquad \frac{\Gamma \Rightarrow \forall x\theta}{\Gamma \Rightarrow \theta(t/x)} \text{ (\forall E)} \end{array}$$

where  $\perp := \psi \wedge \neg\psi$  for some sentence  $\psi$ , in  $(\forall I)$   $c$  is any individual constant

not occurring in  $\forall x\theta$  or in  $\Gamma$ , and in  $(\forall E)$   $t$  is any closed term.

Here we hit the first bump in the road. To begin with,  $(\forall I)$  assumes that the language we are working with has at least one individual constant. And to continue,  $\mathcal{C}$  isn't complete for classical consequence in any language with a *finite* stock of individual constants. The reason, briefly put, is that when every constant of a language occurs in a finite set of sentences  $\Gamma$ ,  $\mathcal{C}$  can't derive all classically valid sequents of the form  $\Gamma \Rightarrow \varphi$ . For example, in a language with a single individual constant  $c$  and two atomic predicates  $P, Q$ , the calculus cannot derive  $\forall x(Qc \wedge Px) \Rightarrow \forall xPx$ .<sup>18</sup>

Is this a problem? On the one hand, it is somewhat tempting to argue that natural languages can have infinitely many 'closed terms' but not infinitely many 'individual constants'.<sup>19</sup> On the other hand, McGee could reply that we follow the rules of  $\mathcal{C}$  'open-endedly', that is, no matter how our language is expanded. Therefore, if we need a fresh individual constant to draw an inference according to  $(\forall I)$ , we can always add it to our language. This may leave some potentially awkward questions about whether we also follow rules of inference no matter how our language is reduced, rather than expanded, and what this means in a scenario where we eliminate all individual constants, but I won't try to press the point here. Let's say that, at the very least if we grant open-endedness,  $\mathcal{C}$  is a sensible description of how classical logicians argue.

Next we have to say more about interpretations. McGee keeps his discussion quite informal, and I will (have to) follow suit, but here is the general idea. In order to consider interpretations in all 'mathematically possible languages', McGee says, we have to adopt a certain abstract point of view:

To make sense of "mathematically possible languages" we have to conceive of a language abstractly, as perhaps an ordered pair consisting of a collection of expression types, appropriately arrayed in grammatical categories, and a function assigning to each expression a semantic value. (McGee 2000, p. 62, my italics)

Officially, then, interpretations are assignments of semantic values to sub-sentential expressions, and these interpretations determine the truth-conditions of whole sentences. McGee doesn't say anything else about interpretations (for instance, the range of possible semantic values for each type of expression, or how truth-conditions are determined by sub-sentential semantic values). As a general background example, though, we can think of the way taking the semantic value

---

<sup>18</sup>This is easy to see 'informally', by trying out a few derivations and noting why they fail. A strict proof is routine but tedious: tweak the rules for  $\wedge$  and  $\neg$  in obvious ways to get a cut-free calculus  $\mathcal{C}'$  that derives the same sequents as  $\mathcal{C}$ , and then show by an upwards proof-search that  $\mathcal{C}'$  can't derive the sequent  $\forall x(Qc \wedge Px) \Rightarrow \forall xPx$ .

<sup>19</sup>The labels 'closed term' and 'individual constant' should be taken with a (large) pinch of salt when applied to natural languages. The idea, however, is that there are infinitely many English expressions that would ordinarily be formalised as closed terms (think of iterating the phrase 'the father of' in front of the name 'Mary'), but only finitely many syntactically simple expressions (e.g. proper names) that would be formalised as individual constants.

of ‘ $\wedge$ ’ to be the usual truth-function determines the truth-conditions of sentences of the form  $\varphi \wedge \psi$ .

Now, how should we define normal interpretations? McGee’s idea is that an interpretation is normal as long as it determines certain truth conditions.<sup>20</sup> Given some interpretation  $\mathcal{I}$  for a language  $\mathcal{L}$ , let’s write ‘ $\mathcal{M} \models \varphi$ ’ to mean that  $\varphi$  is true in the  $\mathcal{L}$ -model  $\mathcal{M}$ . Then  $\mathcal{I}$  is  $\wedge$ -normal or  $\neg$ -normal, respectively, if for every  $\mathcal{L}$ -model:

$$(\wedge)_m \mathcal{M} \models \varphi \wedge \psi \text{ iff } \mathcal{M} \models \varphi \text{ and } \mathcal{M} \models \psi.$$

$$(\neg)_m \mathcal{M} \models \neg\varphi \text{ iff } \mathcal{M} \not\models \varphi.$$

As for the universal quantifier, let  $c$  be an individual constant of  $\mathcal{L}$  and say that  $\mathcal{L}$ -models  $\mathcal{M}$  and  $\mathcal{M}'$  are *c*-variants when they differ only in the value they assign to  $c$ . Then  $\mathcal{I}$  is  $\forall$ -normal when:

$$(\forall)_m \text{ If } c \text{ does not appear in } \theta, \text{ then } \mathcal{M} \models \forall x\theta(x) \text{ iff } \mathcal{M}' \models \theta[c/x] \text{ in every } c\text{-variant } \mathcal{M}' \text{ of } \mathcal{M}.$$

Here we hit another snag. To start with, not all languages have individual constants. And to continue, if a language has a finite stock of individual constants, some sentences of the form  $\forall x\theta(x)$  will contain all of them, but  $(\forall)_m$  can’t be applied to such sentences. Simply put:  $(\forall)_m$  defines  $\forall$ -normal interpretations only for a restricted class of languages.

This time the fact that any language can be extended is irrelevant; the point is that we want a definition of  $\forall$ -normal interpretations for arbitrary languages, not for some extensions of any given language. This seems like a good reason to reject McGee’s definition of normality, but let’s put the issue aside for the rest of this section. From now on, all our languages have infinitely many individual constants.

The final piece of the puzzle is explaining how rules constrain interpretations. The first part of McGee’s answer is simple enough: acceptable interpretations should make the rules of  $\mathcal{C}$  validity-preserving. Given some interpretation  $\mathcal{I}$ , a sequent  $\Gamma \Rightarrow \psi$  is valid when, for all models  $\mathcal{M}$  of our language,  $\mathcal{M} \models \Gamma$  implies  $\mathcal{M} \models \psi$ , and a rule  $R$  is validity-preserving when, for every application of  $R$ , if the premiss-sequents are valid under  $\mathcal{I}$ , so is the conclusion-sequent. In the case of rules with no premiss-sequents, like (Id) or  $(\wedge I)$ , validity-preservation boils down to the validity of certain sequents. For this reason, and to avoid repetition, I will sometimes use the expressions ‘rule  $R$  is valid’ and ‘rule  $R$  is validity-preserving’ interchangeably.<sup>21</sup>

<sup>20</sup>The fact that interpretations are assignments of semantic values to sub-sentential expressions while normality is defined at the sentential level could be problematic. As we will see in the next section, some interpretations are ‘sub-sententially non-normal’ (i.e. they assign sub-sentential expressions an unusual semantic value) and ‘sententially normal’ (i.e. they determine the same semantic value for sentences as a ‘sub-sententially normal’ interpretation would). For more details see 6.5.1

<sup>21</sup>McGee often says that he takes the rules of inference to be “truth preserving” (e.g. McGee 2000, pp. 66, 67, 70), but it is clear from context, and obvious from his proofs, that he means what I am calling ‘validity-preserving’ (or ‘validity’ for short).

The second part of McGee’s answer is where open-endedness comes into play. The idea, recall, is that the rules of  $\mathcal{C}$  should come out valid ‘no matter how our language is extended’. McGee massages this into a more concrete requirement as follows:

The semantic role of a sentence is determined, uniquely up to logical equivalence, by specifying the models in which the sentence is true. *So the effect of introducing a new sentence is to identify a new class of models*, and the effect of open-endedness is to broaden the class of models with respect to which [the rules of  $\mathcal{C}$  are required to be valid]. (McGee 2000, p. 66, my italics)

There are certain complications to unpack here. When McGee says that the effect of adding new expressions to a language  $\mathcal{L}$  is “to identify a new class of models”, what he means is a new class of  $\mathcal{L}$ -models —this much is clear from his proofs, as we will see below. This is a problem. Suppose we enrich a language  $\mathcal{L}$  with a new predicate  $P$ , and call the resulting language  $\mathcal{L}'$ . Then  $\mathcal{L}'$ -models are distinct from  $\mathcal{L}$ -models, and strictly speaking no sentence of  $\mathcal{L}'$  identifies a class of  $\mathcal{L}$ -models.

A possible explanation is that McGee also assumes that “changing the logical operators does not change the class of models” (McGee 2000, p. 65). Perhaps, then, his talk of mathematically possible extensions should be understood as talk of mathematically possible extensions by means of new logical operators. This, however, raises problems of its own. The most obvious is that McGee offers no clear reason to assume that adding new logical operators to a language  $\mathcal{L}$  never requires us to change the notion of an  $\mathcal{L}$ -model. He states in a footnote that “most, but not all, of the connectives and quantifiers we encounter in practice meet this condition”. He then considers “one [quantifier] that does not”, Keisler’s probability quantifier, and concludes that “it is not altogether clear that [it] ought to count as a logical operator”, since it is not invariant under permutations (see McGee 2000, fn. 24). Unfortunately, it is difficult to see why we should derive a general lesson from this specific example, and in any case, cashing out logicity in terms of invariance under permutations is not the only option.<sup>22</sup>

Let’s put this aside for a minute and keep following McGee, though. The effect of open-endedness, he says, is to “broaden the class of models with respect to which the rules of  $\mathcal{C}$  are valid”. But how much does it broaden this class?

There are surely limits —though no one knows much about them— on what classes of models it would be possible to define in some psychologically and practically feasible extension of English. Surely, however, there are no limits on which classes of models it is possible, in principle, to define in some mathematically possible language. (McGee 2000, p. 67)

---

<sup>22</sup>See e.g. (Gómez-Torrente 2002), (Brandom 1994), and (Warmbröd 1999) for further references.

Once again we face some issues here. The first is that McGee confidently asserts that any class of models is definable in some mathematically possible language, but gives no reason to share his certainty. The second is that, given the earlier quote, what seems relevant is whether every class of models of a given language  $\mathcal{L}$  is definable in an extension of  $\mathcal{L}$  by means of logical operators only. This, however, doesn't follow even if we grant that every class of models is definable in some language. And finally, the point at issue is whether all interpretations of logical vocabulary can be pinned down by inferential practice, so it isn't clear that we can appeal, as McGee does, to 'psychologically and practically' unfeasible extensions of our language.<sup>23</sup>

Let's take stock. McGee's argues that, by open-endedness, interpretations should make the rules of  $\mathcal{C}$  come out valid in every mathematically possible extension of our language. He also holds that every class of models is definable in some mathematically possible language, and he concludes that the rules of  $\mathcal{C}$  must come out valid in every class of models for our language. There are three problems with this argument. The first is that it is a non-sequitur. The conclusion only follows if we assume that every class of models of a language  $\mathcal{L}$  is definable in one of its 'mathematically possible extension' by means of logical operators, and also assume that adding new logical operators to  $\mathcal{L}$  doesn't change the class of  $\mathcal{L}$ -models. The second problem is that we have been given no reason to make those assumptions. And the third problem is that, even if we grant McGee's assumptions, it doesn't go without saying that we can appeal to all extensions of a given language by means of logical operators, rather than to extensions that are 'psychologically and practically feasible', as he puts it.

Be that as it may, we now have enough background to go through McGee's proof that the rules for  $\mathcal{C}$  (and open-endedness) pin down normal interpretations. Fix some language  $\mathcal{L}$  and consider any  $\mathcal{L}$ -interpretation  $\mathcal{I}$  that makes the rules of  $\mathcal{C}$  come out valid with respect to any class of  $\mathcal{L}$ -models. Our goal is to show that  $\mathcal{I}$  is normal. The case of conjunction is almost trivial. By assumption  $\mathcal{I}$  makes the following rules valid:

$$\frac{}{\varphi, \psi \Rightarrow \varphi \wedge \psi} (\wedge I) \quad \frac{}{\varphi \wedge \psi \Rightarrow \varphi} (\wedge E_1) \quad \frac{}{\varphi \wedge \psi \Rightarrow \psi} (\wedge E_2)$$

In other words, for every  $\mathcal{L}$ -model  $\mathcal{M}$ , if the antecedents of any of these sequents are true in  $\mathcal{M}$ , so is the conclusion. From this it immediately follows that for any model  $\mathcal{M}$ :

$$(\wedge)_m \mathcal{M} \models \varphi \wedge \psi \text{ iff } \mathcal{M} \models \varphi \text{ and } \mathcal{M} \models \psi.$$

Things start to get interesting with negation. McGee's rules for it, recall, are:

---

<sup>23</sup>For more on this last point see (Piccolo, forthcoming).

$$\frac{\Gamma, \varphi \Rightarrow \perp}{\Gamma \Rightarrow \neg\varphi} (\neg I) \quad \frac{}{\neg\neg\varphi \Rightarrow \varphi} (\neg E_1) \quad \frac{}{\varphi, \neg\varphi \Rightarrow \psi} (\neg E_2)$$

where  $\perp := \chi \wedge \neg\chi$  for some sentence  $\chi$ .

At this point McGee introduces the extra assumption that there is no model in which all sentences are true (McGee 2000, p. 70). Let's make this official:

**(Non-Triv):** There is no  $\mathcal{L}$ -model  $\mathcal{M}$  such that  $\mathcal{M} \models \varphi$  for every sentence  $\varphi$  of  $\mathcal{L}$ .

**(Non-Triv)** is McGee's analogue of Garson's restriction to non-trivial interpretations (see Section 6.2.1), and much like Garson, McGee offers no specific reason to adopt **(Non-Triv)**. In any case, from **(Non-Triv)** and the assumption that  $(\neg E_2)$  is valid we immediately get that:

**(a)** There is no  $\mathcal{L}$ -model  $\mathcal{M}$  and  $\mathcal{L}$ -sentence  $\varphi$  such that  $\mathcal{M} \models \varphi$  and  $\mathcal{M} \models \neg\varphi$ .

Given **(a)**, all that remains in order to show that  $\mathcal{I}$  is  $\neg$ -normal is:

**(c)** There is no  $\mathcal{L}$ -model  $\mathcal{M}$  and  $\mathcal{L}$ -sentence  $\varphi$  such that  $\mathcal{M} \not\models \varphi$  and  $\mathcal{M} \not\models \neg\varphi$ .

As the labels suggest, we need a quick intermediate lemma to prove **(c)**, namely:

**(b)** There is no  $\mathcal{L}$ -model  $\mathcal{M}$  such that  $\mathcal{M} \models \perp$ .

To see that **(b)** holds, just note that all the rules of  $\mathcal{C}$  are valid (or validity-preserving) by assumption, so every sequent provable in  $\mathcal{C}$  is valid. Since we can derive the sequent  $\perp \Rightarrow \psi$  for arbitrary  $\psi$ :

$$\frac{\frac{\frac{}{\chi \wedge \neg\chi \Rightarrow \chi} (\wedge E_1) \quad \frac{}{\chi, \neg\chi \Rightarrow \psi} (\neg E_2)}{\chi \wedge \neg\chi, \neg\chi \Rightarrow \psi} (\text{Cut}) \quad \frac{}{\chi \wedge \neg\chi \Rightarrow \neg\chi} (\wedge E_2)}{\chi \wedge \neg\chi \Rightarrow \psi} (\text{Cut})$$

lemma **(b)** follows by **(Non-Triv)** once again.

We can now tackle McGee's proof of **(c)** —see (McGee 2000, p. 71). First, he says, consider an extension  $\mathcal{L}^+$  of  $\mathcal{L}$  which contains a sentence  $\delta$  that is true in all and only the models where  $\mathcal{M} \not\models \varphi$  and  $\mathcal{M} \not\models \neg\varphi$ . Since there is no model such that  $\mathcal{M} \models \delta$  and  $\mathcal{M} \models \varphi$ , the sequent  $\delta, \varphi \Rightarrow \perp$  is valid. Since  $(\neg I)$  preserves validity, the sequent  $\delta \Rightarrow \neg\varphi$  is valid as well, and by a similar reasoning the sequent  $\delta \Rightarrow \neg\neg\varphi$  is valid too. Given that the rules preserve validity, the end-sequent of the following derivation is also valid:

$$\frac{\frac{\delta \Rightarrow \neg\varphi \quad \frac{\quad}{\neg\varphi, \neg\neg\varphi \Rightarrow \perp} (\wedge E_2)}{\delta, \neg\neg\varphi \Rightarrow \perp} (\text{Cut})}{\delta \Rightarrow \perp} \quad \frac{\delta \Rightarrow \neg\neg\varphi}{\delta \Rightarrow \perp} (\text{Cut})$$

Therefore, it follows by **(b)** that there is no  $\mathcal{M}$  such that  $\mathcal{M} \models \delta$ , and by definition of  $\delta$ :

**(c)** There is no  $\mathcal{L}$ -model  $\mathcal{M}$  and  $\mathcal{L}$ -sentence  $\varphi$  such that  $\mathcal{M} \not\models \varphi$  and  $\mathcal{M} \not\models \neg\varphi$ .

There is a further unstated assumption in this proof. McGee’s argument, in a nutshell, is that if we extend the language  $\mathcal{L}$  to  $\mathcal{L}^+$ , then there is no model such that  $\mathcal{M} \not\models \varphi$  and  $\mathcal{M} \not\models \neg\varphi$ . This tells us something about the original language  $\mathcal{L}$  only if we assume that the truth-value of  $\varphi$  in every model remains unchanged as we go from  $\mathcal{L}$  to  $\mathcal{L}^+$  and vice versa; otherwise, it could be precisely expanding  $\mathcal{L}$  into  $\mathcal{L}^+$  that makes it the case that there is no model such that  $\mathcal{M} \not\models \varphi$  and  $\mathcal{M} \not\models \neg\varphi$ . Note, however, that it is possible to define interpretations for which extending the language in certain ways can change the truth-value of sentences (e.g. the substitutional interpretation of quantifiers in Bonevac 1985, p. 230), so it is unclear whether this assumption is legitimate.

Finally, let’s move on to the universal quantifier. The rules governing  $\forall$ , recall, are:

$$\frac{\Gamma \Rightarrow \theta(c/x)}{\Gamma \Rightarrow \forall x\theta} (\forall I) \qquad \frac{\Gamma \Rightarrow \forall x\theta}{\Gamma \Rightarrow \theta(t/x)} (\forall E)$$

where  $\theta(x)$  is a formula with one variable  $x$  free, and  $c$  is any constant not occurring in  $\forall x\theta$  or in  $\Gamma$ . Once again, McGee introduces an extra assumption to deal with this case:

To get the truth conditions for quantified sentences, we need to think about how to utilize the proviso “the individual constant  $c$  does not appear in  $\forall x\theta$  or in  $\Gamma$ ” in rule  $(\forall I)$ . Let’s say that two models  $\mathcal{M}$  and  $\mathcal{M}'$  are  $c$ -variants if they are just alike, except perhaps in the value they assign to  $c$ . If  $c$  does not appear in the sentence  $\psi$ , then whether  $\psi$  is true in  $\mathcal{M}$  *should not* depend on what value  $\mathcal{M}$  assigns to  $c$ , so that, if  $\mathcal{M}$  and  $\mathcal{M}'$  are  $c$ -variants and  $\psi$  is true in one of them, it is true in both. *Thus we assume that*, if the constant  $c$  does not occur in  $\psi$ , then the class of models in which  $\psi$  is true is closed under  $c$ -variants, and, conversely, that if a class of models is closed under  $c$ -variants, then there is, in some mathematically permissible language, a sentence not containing  $c$  that is true in all and only the members of the class. (McGee 2000, p. 71, with added italics and changes in notation)

Now, the assumption being introduced is a certain closure condition on all interpretations, normal and non-normal. Let's isolate it:

**(Clo)**

- (i) If the constant  $c$  does not occur in  $\psi$ , then the class of models in which  $\psi$  is true is closed under  $c$ -variants, and
- (ii) if a class of models is closed under  $c$ -variants, then there is, in some mathematically permissible language, a sentence not containing  $c$  that is true in all and only the members of the class.

The obvious question is: why should we impose this condition? As far as I can tell, McGee offers no motivation. All that we are told is that “if  $c$  does not appear in the sentence  $\psi$ , then whether  $\psi$  is true in  $\mathcal{M}$  *should not* depend on what value  $\mathcal{M}$  assigns to  $c$ ”, but the question is precisely why it shouldn't (and in what sense of ‘should’). On the face of it, McGee is simply ruling out some non-normal interpretations by fiat.

In any case, once **(Clo)** is in place  $\forall$ -normality follows easily. Recall that an interpretation is  $\forall$ -normal if, for every model  $\mathcal{M}$ :

$(\forall)_m$  If  $c$  does not appear in  $\theta$ , then  $\mathcal{M} \models \forall x\theta(x)$  iff  $\mathcal{M}' \models \theta[c/x]$  in every  $c$ -variant  $\mathcal{M}'$  of  $\mathcal{M}$ .

To show the left-to-right direction, suppose  $c$  doesn't occur in  $\forall x\theta(x)$  and  $\mathcal{M} \models \forall x\theta(x)$ . Then by **(Clo)(i)** we have that  $\mathcal{M}' \models \forall x\theta(x)$  in any  $c$ -variant  $\mathcal{M}'$  of  $\mathcal{M}$ . But since the sequent  $\forall x\theta \Rightarrow \theta(c/x)$  is valid (since  $(\forall E)$  preserves validity), it follows that  $\mathcal{M}' \models \theta(c/x)$  in any  $c$ -variant  $\mathcal{M}'$  of  $\mathcal{M}$ .

For the right-to-left direction, suppose  $\mathcal{M}' \models \theta(c/x)$  in any  $c$ -variant  $\mathcal{M}'$  of  $\mathcal{M}$ . By **(Clo)(ii)** there is a sentence  $\psi$  not containing  $c$  that is true in all and only the  $c$ -variants of  $\mathcal{M}$ . This means that the sequent  $\psi \Rightarrow \theta(c/x)$  is valid, and since rule  $(\forall I)$  preserves validity, so is the sequent  $\psi \Rightarrow \forall x\theta(x)$ . Finally, since  $\mathcal{M} \models \psi$ , we have that  $\mathcal{M} \models \forall x\theta(x)$ .

All in all the discussion in this section can be summarised as follows. McGee's general idea is that open-endedness rules out non-normal interpretations of logical vocabulary. On closer inspection, however, one needs more than open-endedness to do this: McGee's results rely on assumptions about the ‘mathematically possible extensions’ of our language and the effect of adding new logical operators, and on relatively ad hoc restrictions like **(Clo)** which haven't been clearly justified. In addition, McGee's definition of  $\forall$ -normal interpretations doesn't apply to arbitrary first-order languages.

## 6.5 Bonnay and Westerståhl: Semantic Constraints

In this section we will look at Bonnay and Westerståhl's (2016) approach to Carnap's Problem. Their overall strategy is easy to explain. According to

Bonnay and Westerståhl some interpretations can be ruled out from the get go; the “space of possible interpretations” as they put it “is a priori restricted by universal semantic principles” (Bonnay and Westerståhl 2016, p. 721). They also claim that, given these universal semantic constraints, we should restrict attention to interpretations that are (a) compositional, (b) non-trivial, and (c) in the case of the quantifiers, invariant under permutations. But with these restrictions in place, they say, Carnap’s Problem is avoided. We will begin by briefly looking at how this strategy plays out for propositional languages, so as to set the stage, and then move on to the first-order case.

An interpretation for a propositional language  $\mathcal{L}$  is an arbitrary valuation, i.e. an arbitrary function from  $\mathcal{L}$ -sentences to  $\{1,0\}$ . Valuations are  $\diamond$ -normal when they respect the truth-table for  $\diamond$ , and *normal* when they are  $\diamond$ -normal for all connectives  $\diamond$  of  $\mathcal{L}$ .

Bonnay and Westerståhl read interpretations off of consequence relations, as in Section 6.1: a valuation  $v$  is *consistent with* a consequence relation  $\vdash$  when it makes all valid arguments of  $\vdash$  truth-preserving: if  $\Gamma \vdash A$  and  $v(\Gamma) = 1$ , then  $v(A) = 1$ .

As we know from Section 6.1, though, it is easy to find non-normal interpretations that are consistent with the classical consequence relation  $\vdash_{\mathcal{L}}$  of any (truth-functionally complete) language  $\mathcal{L}$ . Our main examples were the trivial valuation  $v_T$  and the valuation  $v_{cl}$  such that  $v_{cl}(A) = 1$  iff  $A$  is a classical tautology. This is where the idea of imposing restrictions on the type of valuation we take into account comes into action. Bonnay and Westerståhl begin by stating that “as speakers, we know that our language [...] is going to have some true and some false sentences” (Bonnay and Westerståhl 2016, p. 733). This is, again, the requirement of non-triviality familiar from Garson and McGee’s accounts.

The second restriction Bonnay and Westerståhl impose is compositionality. The usual, fast and loose formulation of the principle of compositionality says that the meaning of a complex expression depends on the meanings of its parts and the syntactic rules used to put them together. This formulation can be made more precise in a number of non-equivalent ways.<sup>24</sup> Bonnay and Westerståhl go for the following one: an assignment  $\mu$  of semantic values to the expressions of some language is compositional if:

**(PC):** For every n-ary syntactic rule  $\mathcal{O}$  there is a semantic composition function  $F_{\mathcal{O}}$  such that for any well-formed expression  $\mathcal{O}(e_1, \dots, e_n)$  we have  $\mu(\mathcal{O}(e_1, \dots, e_n)) = F_{\mathcal{O}}(\mu(e_1), \dots, \mu(e_n))$ .

In the case of propositional logic, where  $\mu$  is a valuation  $v$ , **(PC)** just says that  $v$  must treat each connective as a truth-function.

These two requirements, compositionality and non-triviality, are enough to rule out all non-normal interpretations of propositional languages. In fact, compositionality alone gets rid of all valuations that are non-normal for  $\wedge$ ,  $\vee$  or  $\rightarrow$ . The requirement of non-triviality is only needed to dispose of the valuation  $v_T$ ,

<sup>24</sup>See (Pagin and Westerståhl 2010a) for discussion and further references.

which is compositional and normal for all our connectives except  $\neg$ . Let's make this into a theorem, for future reference:

**Theorem 6.19.** *Let  $\mathcal{L}$  be a propositional language and  $v$  a valuation for  $\mathcal{L}$ . If  $v$  is compositional, non-trivial, and consistent with  $\vdash_{\mathcal{L}}$ , then  $v$  is normal.*

*Proof.* See Bonnay and Westerståhl 2016. □

There is an obvious question that we haven't asked yet: why should we require non-triviality and compositionality? In the case of non-triviality, we are simply told that "it is a very weak requirement, hardly in need of motivation" (Bonnay and Westerståhl 2016, p. 725). It is unclear why we should accept this, though. Of course, taking every sentence of a language to be true is a very contrived interpretation, but the question was never whether non-normal interpretations are far-fetched, but whether they are possible. In this context, we can't just flatly assert that "as speakers, we know that our language is going to have some true and some false sentences". The point is precisely whether we do, and more importantly, how.

As for compositionality, we get the usual argument concerning learnability: natural languages have infinitely many sentences, so speakers can't learn their meanings one by one. Compositionality, therefore, seems like "our currently best explanation of learnability" (Bonnay and Westerståhl 2016, p. 725). As it is often pointed out, however, learnability constraints don't establish the need for compositionality in the strong sense demanded by Bonnay and Westerståhl.<sup>25</sup> We will come back to compositionality and its consequences several times throughout this section. First, though, we need to explain how Bonnay and Westerståhl go about Carnap's Problem in first-order settings.

As we know well, the usual Tarskian clauses for the satisfaction of quantified formulas in a model under an assignment are:

$$(iv) \mathcal{M}, \sigma \models \forall x \varphi \text{ iff } \mathcal{M}, \sigma[a/x] \models \varphi \text{ for all } a \in D.$$

$$(v) \mathcal{M}, \sigma \models \exists x \varphi \text{ iff } \mathcal{M}, \sigma[a/x] \models \varphi \text{ for some } a \in D.$$

When  $\forall$  and  $\exists$  are treated as generalised quantifiers, clauses (iv) and (v) are slightly modified. Quantifiers, in these contexts, denote sets of subsets of the domain:  $\forall$  denotes the set  $\{D\}$  containing only the domain  $D$  of the structure, and  $\exists$  denotes the set of all non-empty subsets of  $D$  (let's call it  $E_D$  for short). The corresponding clauses for satisfaction are:

$$(iv)' \mathcal{M}, \sigma \models \forall x \varphi \text{ iff } \{a \in D \mid \mathcal{M}, \sigma[a/x] \models \varphi\} \in \{D\}.$$

$$(v)' \mathcal{M}, \sigma \models \exists x \varphi \text{ iff } \{a \in D \mid \mathcal{M}, \sigma[a/x] \models \varphi\} \in E_D.$$

---

<sup>25</sup>See e.g. (Pagin and Westerståhl 2010b) or (Szabó 2012) for details.

It should be clear that clauses (iv)-(v) are equivalent to (iv)'-(v)', in the sense that they give rise to the same satisfaction relation. Bonnay and Westerståhl prefer the latter formulation, and we will follow suit. They introduce the general notion of a first-order interpretation as follows:

Since we assume compositionality, our interpretations amount to giving syntactically adequate semantic values to the logical and non-logical vocabulary. Since we furthermore assume non-triviality, we need not worry about the interpretation of connectives, which has to be standard by [Theorem 6.19]. Hence, our interpretations can be taken to be pairs of the form  $(\mathcal{M}, Q)$  where  $\mathcal{M}$  is a standard  $\mathcal{L}$ -structure [...] interpreting the non-logical vocabulary of  $\mathcal{M}$ , and  $Q$  is a set of subsets of [the domain of the structure], interpreting  $\forall$ . (Bonnay and Westerståhl 2016, p. 729)

An interpretation for a language  $\mathcal{L}$ , then, is a *weak model*  $\mathcal{M}, Q$ , where  $\mathcal{M} = (D, I)$  is an  $\mathcal{L}$ -structure and  $Q$  is a set of subsets of  $D$ . Given a weak model  $\mathcal{M}, Q$ , satisfaction is defined in the obvious way:<sup>26</sup>

- (1)  $\mathcal{M}, Q, \sigma \models P(x_1, \dots, x_n)$  iff  $(\sigma(x_1), \dots, \sigma(x_n)) \in I(P)$ .
- (2)  $\mathcal{M}, Q, \sigma \models \varphi \wedge \psi$  iff  $\mathcal{M}, Q, \sigma \models \varphi$  and  $\mathcal{M}, Q, \sigma \models \psi$ .
- (3)  $\mathcal{M}, Q, \sigma \models \neg\varphi$  iff  $\mathcal{M}, Q, \sigma \not\models \varphi$ .
- (4)'  $\mathcal{M}, Q, \sigma \models \forall x\varphi$  iff  $\{a \in D \mid \mathcal{M}, Q, \sigma[a/x] \models \varphi\} \in Q$ .

Clauses (2) and (3) are just the usual clauses for  $\wedge$  and  $\neg$ . This is to be expected, since Bonnay and Westerståhl claim that compositionality and non-triviality fix the normal interpretation of connectives. Clause (4)' is almost identical to (iv)', the standard clause for  $\forall$ . The only difference is that in (4)' the denotation of  $\forall$  can be any set of subsets of  $D$ , and not just the usual one,  $\{D\}$ . In other words, the only possible source of non-normality in a weak model comes from deviant interpretations of  $\forall$ . A weak model  $\mathcal{M}, Q$ , then, is normal if  $Q = \{D\}$ , and non-normal otherwise.

Adapting the notion of consistency to weak models is straightforward. Bonnay and Westerståhl take the classical consequence relation  $\vdash_{\mathcal{L}}$  of a first-order language  $\mathcal{L}$  to be a relation between sentences (closed formulas) only; this means that a weak model  $\mathcal{M}, Q$  is consistent with the classical consequence relation  $\vdash_{\mathcal{L}}$  of  $\mathcal{L}$  if, whenever  $\Gamma \vdash_{\mathcal{L}} \varphi$  and  $\mathcal{M}, Q \models \Gamma$ , we have that  $\mathcal{M}, Q \models \varphi$  (no reference to variable assignments is needed, since we are dealing with sentences).

Finally, Bonnay and Westerståhl also require that the interpretation of quantifiers be invariant under permutations. They define this notion in the usual way. A permutation  $\pi$  of the domain  $D$  of a structure is a bijection of  $D$  onto itself.

<sup>26</sup>Of course, if  $\mathcal{L}$  contains individual constants and function symbols, clause (1) is modified in the usual way to account for terms.

If  $\pi$  is a permutation of  $D$  and  $X \subseteq D$ , let  $\pi(X) = \{\pi(a) \in D \mid a \in X\}$ . Then  $Q \subseteq \mathcal{P}(D)$  is invariant under permutations if  $Q = \{\pi(X) \in \mathcal{P}(D) \mid X \in Q\}$  for all permutations  $\pi$ . It is easy to check that  $\{D\}$ , the normal interpretation of  $\forall$ , is indeed invariant under permutations.

According to Bonnay and Westerståhl, all weak models that are (non-trivial, compositional, invariant under permutations and) consistent with the classical consequence relation of their language must be normal. In the rest of this section I will focus on two problems with this claim. The first is that Bonnay and Westerståhl's restrictions don't seem to rule out non-normal interpretations for all languages. The second is that their definition of a first-order interpretation either violates compositionality or else begs the question against non-normal interpretations of connectives. These problems are explained in 6.5.1 and 6.5.2, respectively.

### 6.5.1 Carnap's Problem and Definability

The main result of Bonnay and Westerståhl's paper is a characterisation of the interpretations of the universal quantifier  $\forall$  that are consistent with the classical consequence relation of a language. This characterisation uses the notion of a principal filter closed under the interpretation of terms. Recall that a principal filter  $F$  on a set  $S$  is a non-empty set of subsets of  $S$  such that for some  $G \subseteq S$ , we have that  $F = \{X \subseteq S \mid G \subseteq X\}$ . Given a weak model  $\mathcal{M}, Q$  for a language  $\mathcal{L}$ , possibly containing individual constants and function symbols, we say that  $Q$  is a principal filter closed under the interpretation of terms in  $\mathcal{M}$  if:

- (i)  $Q$  is a principal filter on  $D$  generated by some  $G \subseteq D$ .
- (ii) For any  $n$ -ary function symbol  $f$  of  $\mathcal{L}$ , if  $a_1, \dots, a_n \in G$ , then  $I(f)(a_1, \dots, a_n) \in G$ .
- (iii) For any individual constant  $c$  of  $\mathcal{L}$ ,  $I(c) \in G$ .

Bonnay and Westerståhl's characterisation is as follows:

**(BW1)** Let  $\mathcal{L}$  be a language with  $\forall$  primitive, let  $\mathcal{M}, Q$  be a weak model for it. Then  $\mathcal{M}, Q$  is consistent with  $\vdash_{\mathcal{L}}$  iff  $Q$  is a principal filter closed under the interpretation of terms in  $\mathcal{M}$ .

They combine this characterisation with an ancillary result:

**(BW2)** A principal filter  $Q$  on  $D$  is invariant under permutations iff  $Q = \{D\}$ .

In theory, (BW1) and (BW2) should take care of Carnap's Problem. According to (BW1) consistency with the classical consequence relation forces us to interpret  $\forall$  as a principal filter closed under the interpretation of terms. (BW2)

then shows that the only such filter that is also invariant under permutations is  $\{D\}$ , the normal interpretation of  $\forall$ . Therefore, non-normal weak models are ruled out (or so it seems).

The problem with Bonnay and Westerståhl’s strategy is easy to explain: they only prove (BW1) for languages with predicate *variables* (that is, for ‘second-order languages without second-order quantifiers’). While Bonnay and Westerståhl claim that the addition of predicate variables is a “simplifying assumption” without which “similar results would hold” (Bonnay and Westerståhl 2016, p. 729), it is difficult to see what they mean. Without predicate variables, some weak models falsify (BW1) and are both non-normal and invariant under permutations. In other words, without predicate variables the normal interpretation of quantifiers is not secured. To see this we need an extra bit of terminology: given a weak model  $\mathcal{M}, Q$  for a language  $\mathcal{L}$ , the *extension* of a formula  $\varphi$  relative to  $\sigma$  and  $x$  is  $\|\varphi\|_{\sigma, x}^{\mathcal{M}, Q} := \{a \in D \mid (\mathcal{M}, Q), \sigma[a/x] \models \varphi\}$ , and a subset  $X$  of  $D$  is *definable* if  $X = \|\varphi\|_{\sigma, x}^{\mathcal{M}, Q}$  for some  $\varphi, \sigma$  and  $x$ . The following lemma gives a counterexample to (BW1) that is simultaneously non-normal and invariant under permutations:<sup>27</sup>

**Lemma 6.20.** *There is a first-order language  $\mathcal{L}_1$  with a non-normal weak model  $\mathcal{N}, Q$  consistent with the classical consequence relation  $\vdash_{\mathcal{L}_1}$ .*

*Proof.* Take the language  $\mathcal{L}_1$  with  $\wedge, \neg$  and  $\forall$  as logical constants and a unary predicate symbol  $P$  as non-logical vocabulary. Consider the  $\mathcal{L}_1$ -structure  $\mathcal{N} = (D, I)$  where  $D$  is any set such that  $|D| > 1$  and  $I(P) = D$ . Next we need a (non-normal) interpretation for  $\forall$  to make  $\mathcal{N}$  into a weak model. Say we interpret  $\forall$  with  $E_D$ , the set of all non-empty subsets of  $D$  —in other words, we will (mis)interpret  $\forall$  as  $\exists$ .

We now have a weak model  $(\mathcal{N}, E_D)$ . It is clear that  $(\mathcal{N}, E_D)$  is non-normal, and it is easy to check that  $E_D$  is invariant under permutations. The only thing left to show is that  $(\mathcal{N}, E_D)$  is consistent with the classical consequence relation  $\vdash_{\mathcal{L}_1}$  of  $\mathcal{L}_1$ . A routine induction on formulas shows that for any  $\varphi, \sigma$  and  $x$ :

$$\begin{aligned} \text{(d}_1\text{)} \quad & \|\varphi\|_{\sigma, x}^{\mathcal{N}, \{D\}} = D \text{ or } \|\varphi\|_{\sigma, x}^{\mathcal{N}, \{D\}} = \emptyset. \\ \text{(d}_2\text{)} \quad & \|\varphi\|_{\sigma, x}^{\mathcal{N}, E_D} = \|\varphi\|_{\sigma, x}^{\mathcal{N}, \{D\}}. \end{aligned}$$

A second induction on formulas then shows that, for any  $\varphi$  and  $\sigma$ , we have that  $\mathcal{N}, \{D\}, \sigma \models \varphi$  iff  $\mathcal{N}, E_D, \sigma \models \varphi$ . All steps of this second induction are trivial except the quantifier one, which goes as follows:

---

<sup>27</sup>Incidentally, Lemma 6.20 is also an example of those ‘sub-sententially non-normal’ but ‘sententially-normal’ interpretations mentioned in fn. 19 of Section 6.4.

$$\begin{aligned}
& \mathcal{N}, \{D\}, \sigma \models \forall x\varphi \text{ iff by (4)'} \\
& \|\varphi\|_{\sigma, x}^{\mathcal{N}, \{D\}} = D \text{ iff by (d}_1\text{), (d}_2\text{)} \\
& |\varphi|_{\sigma, x}^{\mathcal{N}, E_D} = D \text{ iff by (d}_1\text{), (d}_2\text{)} \\
& |\varphi|_{\sigma, x}^{\mathcal{N}, E_D} \in E_D \text{ iff by (4)'} \\
& \mathcal{N}, E_D, \sigma \models \forall x\varphi.
\end{aligned}$$

Finally,  $\mathcal{N}, \{D\}$  is a normal interpretation, and hence consistent with  $\vdash_{\mathcal{L}_1}$ . But by the above the normal weak model  $\mathcal{N}, \{D\}$  is equivalent to the non-normal weak model  $\mathcal{N}, E_D$ , so  $\mathcal{N}, E_D$  is also consistent with  $\vdash_{\mathcal{L}_1}$ .  $\square$

Lemma 6.20 is formulated for the very simple language  $\mathcal{L}_1$ , but it can be adapted to richer ones. If a language has function symbols and individual constants, the function  $I$  can give them any value whatsoever, and if it has further predicate letters  $R$  of arity  $n$  we just need to set  $I(R) = D^n$ . It is also important to note that the fact that we (mis)interpreted  $\forall$  as  $\exists$  is largely irrelevant. For instance, let  $E_n$  be the set of subsets of  $D$  containing at least  $n$  elements. Then Lemma 6.20 also applies if the denotation of  $\forall$  is  $E_n$  for any  $n > 1$ . And finally, Lemma 6.20 can be adapted to languages with  $\exists$ , rather than  $\forall$ , as a primitive symbol. Let a weak model  $\mathcal{M}, Q$  for a language where  $\exists$  is primitive be  $\exists$ -normal if  $Q = E_D$ . Then:

**Corollary 6.21.** *There is a language  $\mathcal{L}_2$  with  $\exists$  as a primitive symbol and a non  $\exists$ -normal weak model  $\mathcal{N}, Q$  consistent with  $\vdash_{\mathcal{L}_2}$ .*

*Proof.* Let  $\mathcal{L}_2$  be exactly as  $\mathcal{L}_1$  but with  $\exists$  primitive instead of  $\forall$ . It follows from Lemma 1 that  $\mathcal{N}, \{D\}$  is a non  $\exists$ -normal weak model for  $\mathcal{L}_2$  consistent with  $\vdash_{\mathcal{L}_2}$ .  $\square$

We can explain what's going on from a more general perspective by adapting a result from (Antonelli 2013). Consider once again the normal interpretation of  $\exists$ :

$$\mathcal{M}, E_D, \sigma \models \exists x\varphi \text{ iff } \{a \in D \mid \mathcal{M}, \sigma[a/x] \models \varphi\} \in E_D$$

There is a certain sense in which this interpretation overshoots the mark. We don't really need  $\exists$  to denote *all* non-empty subsets of the domain in order to get the usual satisfaction relation. It is enough that it denotes all the *definable* non-empty subsets of the domain, those that are the extension of some formula. Let's make this observation more precise and more general. Given a weak model  $\mathcal{M}, Q$ , let  $\text{Def}(\mathcal{M}, Q)$  be the set of definable subsets of  $D$ . Then:<sup>28</sup>

<sup>28</sup>Lemma 6.22 is a modification of the Hull Theorem from (Antonelli 2013). The modification is needed to (a) adapt Antonelli's idea to the present setting and (b) avoid a problem with Antonelli's definition of the hull of a model, pointed out in (Antonelli 2017).

**Lemma 6.22.** *If  $\mathcal{M}, Q$  and  $\mathcal{M}, Q'$  are weak models such that  $Q' = Q \cap \text{Def}(\mathcal{M}, Q)$ , then  $\mathcal{M}, Q, \sigma \models \varphi$  iff  $\mathcal{M}, Q', \sigma \models \varphi$  for any  $\varphi$  and  $\sigma$ .*

*Proof.* The proof is by induction on formulas. The weak models agree on the interpretation of atomic formulas and connectives, so all steps are trivial except the quantifier one. Let  $\bar{Q}$  be whatever quantifier  $Q$  and  $Q'$  are meant to interpret. Then:

$$\begin{aligned} \mathcal{M}, Q, \sigma \models \bar{Q}x\psi &\text{ iff} \\ \|\psi\|_{\sigma,x}^{\mathcal{M},Q} \in Q \cap \text{Def}(\mathcal{M}, Q) &\text{ iff (by induction hypothesis)} \\ \|\psi\|_{\sigma,x}^{\mathcal{M},Q'} \in Q \cap \text{Def}(\mathcal{M}, Q) &\text{ iff (by assumption)} \\ \|\psi\|_{\sigma,x}^{\mathcal{M},Q'} \in Q' &\text{ iff} \\ \mathcal{M}, Q', \sigma \models \bar{Q}x\psi. & \end{aligned}$$

□

Corollary 6.21 can now be seen as a special case of Lemma 6.22. The only definable sets of  $\mathcal{N}, E_D$  are  $D$  and  $\emptyset$ . Thus  $\{D\} = E_D \cap \text{Def}(\mathcal{N}, E_D)$ , and so the non  $\exists$ -normal model  $\mathcal{N}, \{D\}$  is equivalent to the  $\exists$ -normal model  $\mathcal{N}, E_D$ .

Something similar happens with the normal interpretation of  $\forall$ :

$$\mathcal{M}, \{D\}, \sigma \models \forall x\varphi \text{ iff } \{a \in D \mid \mathcal{M}, \sigma[a/x] \models \varphi\} \in \{D\}$$

In this case we don't need  $\forall$  to denote the set containing *only*  $D$ . We can, for instance, clutter its denotation with undefinable subsets and get an equivalent weak model:

**Lemma 6.23.** *Let  $\mathcal{M}, Q$  and  $\mathcal{M}, Q'$  be weak models such that  $Q' = Q \cup B$  for some  $B \subseteq \mathcal{P}(D)$  such that  $B \cap \text{Def}(\mathcal{M}, Q) = \emptyset$ . Then  $\mathcal{M}, Q, \sigma \models \varphi$  iff  $\mathcal{M}, Q', \sigma \models \varphi$  for any  $\varphi$  and  $\sigma$ .*

*Proof.* Again by induction on formulas. All steps are trivial except the quantifier one. For the left-to-right direction note that

$$\begin{aligned} \mathcal{M}, Q, \sigma \models \bar{Q}x\psi &\Rightarrow \|\psi\|_{\sigma,x}^{\mathcal{M},Q} \in Q \Rightarrow \text{(by induction hypothesis)} \\ \|\psi\|_{\sigma,x}^{\mathcal{M},Q'} \in Q &\Rightarrow \|\psi\|_{\sigma,x}^{\mathcal{M},Q} \in Q \cup B \Rightarrow \mathcal{M}, Q', \sigma \models \bar{Q}x\psi. \end{aligned}$$

For the right-to left direction, note that

$$\begin{aligned} \mathcal{M}, Q', \sigma \models \bar{Q}x\psi &\Rightarrow \|\psi\|_{\sigma,x}^{\mathcal{M},Q'} \in Q \cup B \Rightarrow \text{(by induction hypothesis)} \\ \|\psi\|_{\sigma,x}^{\mathcal{M},Q} \in Q \cup B &\Rightarrow \|\psi\|_{\sigma,x}^{\mathcal{M},Q} \in Q \Rightarrow \mathcal{M}, Q, \sigma \models \bar{Q}x\psi. \end{aligned}$$

□

Lemma 6.20 can then be seen as a special case of Lemma 6.23. The only definable sets of  $\mathcal{N}, \{D\}$  are  $D$  and  $\emptyset$ . Thus  $E_D = \{D\} \cup B$  for a  $B \subseteq \mathcal{P}(D)$  such that  $B \cap \text{Def}(\mathcal{N}, \{D\}) = \emptyset$ , and hence the non  $\forall$ -normal weak model  $\mathcal{N}, E_D$  is equivalent to the  $\forall$ -normal weak model  $\mathcal{N}, \{D\}$ .

Let’s summarise the discussion. Bonnay and Westerståhl’s use of predicate variables is a core piece of the machinery that gets them their results. The non-normal weak models above exploit the fact that, given a first-order language and a structure for it, there are often subsets of the domain that can’t be defined by any formula.<sup>29</sup> Adding predicate variables to a first-order language boosts its expressive power, and therefore rules out a host of non-normal interpretations. Bonnay and Westerståhl’s proof of (BW1) —for second-order languages— hinges on this, and cannot be adapted to first-order ones.<sup>30</sup> Note, incidentally, that the results in this subsection show that there are non-normal interpretations of the quantifiers *even if we assume that the standard interpretation of the connectives is fixed*; indeed, the ‘normality’ of the connectives is built into the way Bonnay and Westerståhl define weak models. As we will see in the next subsection, however, there is a problem with this.

## 6.5.2 Carnap’s Problem and Compositionality

We are now going to focus on a problem with the way Bonnay and Westerståhl define interpretations. The key fact to keep in mind is that the usual, two-valued semantics for first-order languages is *not* compositional. This calls for some explanation. As we know from Section 6.2.3 the usual clauses defining satisfaction are not an assignment of semantic values; they are the definition of a ternary relation  $\models$  between structures, variable assignments, and formulas. The notion of compositionality, as Bonnay and Westerståhl define it, only applies to assignments of semantic values, so it doesn’t really make sense to say that a satisfaction relation is, or isn’t, compositional. Still, as we also saw in 6.2.3 there is an obvious way in which a satisfaction relation  $\models$  determines a class of valuations: just set  $v_\sigma^{\mathcal{M}}(\varphi) = 1$  if  $\mathcal{M}, \sigma \models \varphi$ , else  $v_\sigma^{\mathcal{M}}(\varphi) = 0$ , for any formula  $\varphi$ .

Each satisfaction relation, then, defines a two valued semantics for first-order languages. But there is a catch: this semantics is not always compositional. Take a language  $\mathcal{L}$ , an  $\mathcal{L}$ -structure  $\mathcal{M}$ , and the standard satisfaction relation

<sup>29</sup>Note that undefinability is a pervasive phenomenon that applies to languages with more expressive power than I have considered above. For instance, a simple cardinality argument shows that given any countable  $\mathcal{L}$  and any  $\mathcal{L}$ -structure  $\mathcal{M} = (D, I)$  with a countably infinite  $D$ , some subsets of  $D$  are not definable by any  $\mathcal{L}$ -formula.

<sup>30</sup>Perhaps one could argue that the addition of predicate variables doesn’t make a language second-order, and that what matters is whether we can quantify over predicate variables (I owe this point to Dag Westerståhl). Ultimately, I think this point doesn’t matter much. If we count languages with predicate variables as first-order, then the problem with Bonnay and Westerståhl’s approach is that it only solves Carnap’s Problem for a restricted (and rather unusual) class of first-order languages.

$\models$  between them. In any valuation determined by  $\models$ , the semantic value of a formula  $\varphi \wedge \psi$  is a function of the semantic value of  $\varphi$  and the semantic value of  $\psi$ :  $v_\sigma^{\mathcal{M}}(\varphi \wedge \psi) = 1$  iff  $v_\sigma^{\mathcal{M}}(\varphi) = v_\sigma^{\mathcal{M}}(\psi) = 1$ , for any  $\sigma$ . Something similar happens with negation:  $v_\sigma^{\mathcal{M}}(\neg\varphi) = 1$  iff  $v_\sigma^{\mathcal{M}}(\varphi) = 0$  (again, for arbitrary  $\sigma$ ). But the same does not hold for quantifiers: the truth value of  $\forall x\varphi$  does not depend only on the truth value of  $\varphi$ . Put more formally, the value of  $v_\sigma^{\mathcal{M}}(\forall x\varphi)$  is not a function of the value of  $v_\sigma^{\mathcal{M}}(\varphi)$ , but of the values of  $v_{\sigma'}^{\mathcal{M}}(\varphi)$  for all appropriate  $\sigma'$ .<sup>31</sup> Indeed, it is easy to find structures  $\mathcal{M}$  such that for some  $\varphi, \psi$  and  $\sigma$ , we have that  $v_\sigma^{\mathcal{M}}(\varphi) = v_\sigma^{\mathcal{M}}(\psi) = 1$  but  $v_\sigma^{\mathcal{M}}(\forall x\varphi) = 0$  and  $v_\sigma^{\mathcal{M}}(\forall x\psi) = 1$ . In those cases no function on semantic values (i.e. no truth-function) corresponds to the syntactic operation generating  $\forall x\varphi$  from  $\varphi$  —and similarly for  $\exists$ .<sup>32</sup>

The fact that the valuations  $v_\sigma^{\mathcal{M}}$  are not compositional is not a problem in and of itself. They assign truth-values to formulas in a natural, systematic, and finitely specifiably way, following the recursive clauses for satisfaction. Nevertheless, this failure of compositionality *must* be a problem for Bonnay and Westerståhl. For them compositionality is a universal constraint that all interpretations, normal or not, must abide by. Unfortunately, they seem to have a two-valued semantics in mind when defining interpretations of first-order languages. Recall what they say:

Since we assume compositionality, our interpretations amount to giving syntactically adequate semantic values to the logical and non-logical vocabulary. Since we furthermore assume non-triviality, we need not worry about the interpretation of connectives, which has to be standard, by [Theorem 6.19]. Hence, our interpretations can be taken to be pairs of the form  $(\mathcal{M}, Q)$  where... (Bonnay and Westerståhl 2016, p. 729)

Theorem 6.19 says that *if our only semantic values are 0 and 1*, compositionality and non-triviality rule out all non-normal interpretations of the connectives. But then something must be wrong with the passage above. If the only semantic values are 0 and 1, then normal interpretations are not (always) compositional. If, on the other hand, there are semantic values besides 1 and 0, then the appeal to Theorem 6.19 is moot, and we can't assume that compositionality and non-triviality pin down the standard interpretation of connectives. This is the second problem with Bonnay and Westerståhl's approach: their definition of a first-order interpretation either violates compositionality or begs the question against non-normal interpretations of the connectives.

<sup>31</sup>The appropriate  $\sigma'$  are of course those of the form  $\sigma[a/x]$  for some element  $a$  of the domain.

<sup>32</sup>Note that if we use partial valuations that assign truth-values only to sentences the resulting semantics is not compositional either. Set, for all sentences  $\varphi$ ,  $v^{\mathcal{M}}(\varphi) = 1$  if  $\mathcal{M} \models \varphi$ , else  $v^{\mathcal{M}}(\varphi) = 0$ . Compositionality, on Bonnay and Westerståhl's formulation, requires that for each syntactic operation  $\mathcal{O}$  there should be a corresponding operation on semantic values. The syntactic components of sentences, however, are not in general sentences themselves. For instance, the syntactic components of  $\forall xP(x)$  are all subsentential, so no operation on semantic values corresponds to the syntactic operation that generates  $\forall xP(x)$ .

It is possible to give compositional semantics for first-order languages, which leads to a natural question. Suppose we reformulate Bonnay and Westerståhl’s notion of an interpretation so that (a) normal interpretations are always compositional and (b) we don’t beg the question against non-normal interpretations of the connectives. Do compositionality, non-triviality and invariance under permutations pin down the normal interpretation of logical vocabulary?

The answer is ‘no’. The reason, roughly put, is that compositional semantics for first-order languages require a large number of semantic values. This, in its turn, makes Carnap’s Problem more difficult to solve, since more semantic values entail more possible interpretations, and therefore more unintended ones to rule out. As we will see, this means that in the new setting Bonnay and Westerståhl’s restrictions are not enough to fix the standard interpretation of connectives, let alone quantifiers.

The most common strategy to set up a compositional semantics for first-order languages is to take the semantic value  $\llbracket \varphi \rrbracket^{\mathcal{M}}$  of a formula  $\varphi$  in a structure  $\mathcal{M}$  to be the set of variable assignments  $\sigma$  such that  $\mathcal{M}, \sigma \models \varphi$ .<sup>33</sup> Let’s bring back some notation from Section 6.2.3: given a structure  $\mathcal{M} = (D, I)$  for a language  $\mathcal{L}$ ,  $A^{\mathcal{M}}$  is the set of all variable assignments from  $\mathcal{L}$  to  $D$ . Then first-order formulas can be interpreted as follows:<sup>34</sup>

- (1)  $\llbracket P(x_1, \dots, x_n) \rrbracket^{\mathcal{M}} = \{ \sigma \in A^{\mathcal{M}} \mid (\sigma(x_1), \dots, \sigma(x_n)) \in I(P) \}$
- (2)  $\llbracket \varphi \wedge \psi \rrbracket^{\mathcal{M}} = \llbracket \varphi \rrbracket^{\mathcal{M}} \cap \llbracket \psi \rrbracket^{\mathcal{M}}$
- (3)  $\llbracket \neg \varphi \rrbracket^{\mathcal{M}} = A^{\mathcal{M}} - \llbracket \varphi \rrbracket^{\mathcal{M}}$
- (4)  $\llbracket \exists x \varphi \rrbracket^{\mathcal{M}} = \{ \sigma \in A^{\mathcal{M}} \mid \sigma \sim_x \sigma' \text{ for some } \sigma' \in \llbracket \varphi \rrbracket^{\mathcal{M}} \}$
- (5)  $\llbracket \forall x \varphi \rrbracket^{\mathcal{M}} = \{ \sigma \in A^{\mathcal{M}} \mid \sigma' \in \llbracket \varphi \rrbracket^{\mathcal{M}} \text{ for all } \sigma' \text{ st. } \sigma \sim_x \sigma' \}$

The general idea is clear: the semantic values of formulas are sets of assignments, and conjunction and negation are interpreted as operations on semantic values (intersection and complementation, respectively). As pointed out in (Wehmeier 2018), however, the resulting semantics is not fully compositional either. The problem lies again with the interpretation of quantifiers—that is, with clauses (4) and (5). Compositionality, remember, demands that for each syntactic operation  $\mathcal{O}$  there should be a corresponding operation on semantic values. The usual way of setting up the syntax of first-order languages has *one* syntactic operation that takes the quantifier  $\forall$ , any variable  $x$ , and any formula  $\varphi$ , and returns a formula  $\forall x \varphi$ —and similarly for  $\exists$ . Clause (5), on the other hand, covertly defines one semantic operation *for each variable  $x$* ; the right-hand side of (5) refers to  $x$  qua syntactic object—and so does (4).

<sup>33</sup>See (Kreisel and Krivine 1967), (Monk 1976), (McGee 1996) or (Janssen and Partee 2011).

<sup>34</sup>Once again, if a language contains individual constants and function symbols clause (1) is modified in the obvious way to account for terms. The notation ‘ $\sigma \sim_x \sigma'$ ’ means that the assignment  $\sigma$  differs from the assignment  $\sigma'$  at most in the value it assigns to  $x$ .

Different (and somewhat strained) solutions to this problem have been put forward. One option is giving semantic values to variables. We could, for instance, take the value  $\llbracket x \rrbracket^{\mathcal{M}}$  of  $x$  in  $\mathcal{M}$  to be the variable  $x$  itself, and reformulate the interpretations of  $\exists$  and  $\forall$  accordingly, as functions that take a set of assignments and a variable as arguments, and map them to a set of assignments.<sup>35</sup> On this account the normal interpretations of  $\wedge$  and  $\neg$  would still be intersection and complementation, but the normal interpretations of quantifiers would be the following functions from  $\text{Vars} \times \mathcal{P}(A^{\mathcal{M}})$  to  $\mathcal{P}(A^{\mathcal{M}})$ :

$$f_{\exists}(x, Y) = \{\sigma \in A^{\mathcal{M}} \mid \sigma \sim_x \sigma' \text{ for some } \sigma' \in Y\}$$

$$f_{\forall}(x, Y) = \{\sigma \in A^{\mathcal{M}} \mid \sigma' \in Y \text{ for all } \sigma' \text{ st. } \sigma \sim_x \sigma'\}$$

Another option is to slightly change the formulation of the syntax, so that for each variable  $x$  we have a syntactic operation  $\mathcal{O}_x$  that takes a string ‘ $\forall x$ ’ and a formula  $\varphi$  and returns a formula  $\forall x \varphi$  (similarly for  $\exists$ ); on this account the string ‘ $\forall x$ ’ is then interpreted as a single unit. The normal interpretations of  $\wedge$  and  $\neg$  are intersection and complementation, as before, but the normal interpretations of quantifier-variable strings are functions from  $\mathcal{P}(A^{\mathcal{M}})$  to  $\mathcal{P}(A^{\mathcal{M}})$ :

$$f_{\exists x}(Y) = \{\sigma \in A^{\mathcal{M}} \mid \sigma \sim_x \sigma' \text{ for some } \sigma' \in Y\}$$

$$f_{\forall x}(Y) = \{\sigma \in A^{\mathcal{M}} \mid \sigma' \in Y \text{ for all } \sigma' \text{ st. } \sigma \sim_x \sigma'\}$$

All of what I will say applies, with obvious modifications, to either option. The second one saves some time and some brackets, so I will take it as the official normal interpretation. Again, nothing hinges on this.

Now we have to adapt the rest of the notions involved in formulating Carnap’s Problem. It will be useful to keep the simpler propositional case in mind while we do it, just to make sure we don’t go astray. An interpretation for a propositional language is a valuation  $v$  from formulas to  $\{1, 0\}$ . In the first-order case interpretations depend on structures  $\mathcal{M} = (D, I)$ . Given such an  $\mathcal{M}$ , an interpretation is a function from formulas to  $\mathcal{P}(A^{\mathcal{M}})$ .

A compositional valuation associates a truth-function to each connective. Thus, compositional valuations can be defined as tuples  $v = (v_o, t_{\wedge}, t_{\vee}, t_{\neg}, t_{\rightarrow})$ , where  $v_o$  is a valuation for atoms and the  $t_{\diamond}$  are truth-functions. In the first-order case compositional interpretations can also be defined as tuples  $\mathcal{T} = (f_o, F_{\wedge}, F_{\neg}, F_{\vee}, F_{\exists})$ . Here  $f_o$  is a function from atoms to  $\mathcal{P}(A^{\mathcal{M}})$ , and the  $F_{\diamond}$  are operations on  $\mathcal{P}(A^{\mathcal{M}})$ .<sup>36</sup>

In a normal, compositional valuation the  $t_{\diamond}$  are the intended truth-functions. Similarly, in a normal, compositional interpretation  $\mathcal{T} = (f_o, F_{\wedge}, F_{\neg}, F_{\vee}, F_{\exists})$  the

<sup>35</sup>This is a simplification of (T. E. Zimmermann and Sternefeld 2013, Ch. 10), which takes a similar but more elaborate route.

<sup>36</sup>Strictly speaking, given our official normal interpretation  $F_{\vee}$  and  $F_{\exists}$  are *sets* of operations (one for each variable).

$F_\circ$  are the intended operations. There is an additional wrinkle to take care of, though. Given the usual definition of satisfaction, the semantic value of an atomic formula  $P(x_1, \dots, x_n)$  is closed under assignments that agree on the free variables  $x_1, \dots, x_n$ . Formally, this means that if  $\sigma \in \llbracket P(x_1, \dots, x_n) \rrbracket^{\mathcal{M}}$  and  $\sigma, \sigma'$  agree on the value of  $x_1, \dots, x_n$ , then  $\sigma' \in \llbracket P(x_1, \dots, x_n) \rrbracket^{\mathcal{M}}$ . Our interpretations are tuples  $\mathcal{T} = (f_o, F_\wedge, F_\neg, F_\forall, F_\exists)$  where  $f_o$  is an *arbitrary* function from atomic formulas to  $\mathcal{P}(A^{\mathcal{M}})$ , and arbitrary functions need not respect that closure condition. In order for an interpretation to be normal, then, we must also demand that  $f_o$  obeys clause (1) above. And finally, a valuation is non-trivial if it does not assign the value 1 to all formulas. Similarly,  $(f_o, F_\wedge, F_\neg, F_\forall, F_\exists)$  is non-trivial if it does not assign the value  $A^{\mathcal{M}}$  to all formulas.

This takes care of the normal interpretations, as well as the non-normal, compositional and non-trivial ones. Now we need to address invariance under permutations. Bonnay and Westerståhl's notion of invariance works well if we take quantifiers to denote sets of subsets of  $D$ . In order to achieve compositionality, however, we have had to interpret quantifiers as operations on sets of assignments, so the definition of invariance has to be adapted. Luckily there is a well-known, off-the-shelf way to do this due to (McGee 1996). If  $\pi$  is a permutation of  $D$  and  $X$  is a set of variable assignments, let  $\pi^*(X) = \{\pi \circ \sigma \in A^{\mathcal{M}} \mid \sigma \in X\}$ . Then an  $n$ -ary operation  $\mathcal{O}$  on  $A^{\mathcal{M}}$  is invariant under permutations if  $\pi^*(\mathcal{O}(X_1, \dots, X_n)) = \mathcal{O}(\pi^*(X_1), \dots, \pi^*(X_n))$  for all permutations  $\pi$ .

The last missing piece of the puzzle is consistency with a consequence relation  $\vdash$ . This is also easy to address. We will say that  $\mathcal{T}$  is consistent with  $\vdash$  if, whenever  $\Gamma \vdash \varphi$ , we have that  $\bigcap_{\gamma \in \Gamma} \llbracket \gamma \rrbracket^{\mathcal{T}} \subseteq \llbracket \varphi \rrbracket^{\mathcal{T}}$ , where  $\llbracket \psi \rrbracket^{\mathcal{T}}$  is the semantic value of  $\psi$  on the interpretation  $\mathcal{T}$ .<sup>37</sup> Following Bonnay and Westerståhl  $\vdash_{\mathcal{L}}$  is a relation between sentences only (although this doesn't really matter).

We have, at last, formulated all the notions related to Carnap's Problem in a way that is compatible with Bonnay and Westerståhl's assumptions. The following lemma shows that, in this setting, compositionality, non-triviality, and invariance under permutations don't fix the standard meaning of logical vocabulary.

**Lemma 6.24.** *There is a first-order language  $\mathcal{L}_3$  with a compositional, non-normal, non-trivial, invariant-under-permutations interpretation  $\mathcal{T}$  that is consistent with  $\vdash_{\mathcal{L}}$ .*

*Proof.* Take the language  $\mathcal{L}_3$  with  $\wedge, \neg$  and  $\forall$  as logical constants and a binary predicate symbol  $R$  as non-logical vocabulary. Consider the  $\mathcal{L}_3$ -structure  $\mathcal{M} = (D, I)$  where  $D$  is any set such that  $|D| > 1$  and  $I(R) = D^2$ . We will now construct an interpretation  $\mathcal{T} = (f_o, F_\wedge, F_\neg, F_\forall, F_\exists)$  based on  $\mathcal{M}$ . Let  $C \subset A^{\mathcal{M}}$  be the set of variable assignments such that  $\sigma(x) = \sigma(y)$  for all variables  $x, y$ . Note that  $C$  is indeed a proper subset of  $A^{\mathcal{M}}$ , since  $|D| > 1$ . Then  $\mathcal{T}$  is defined as follows:

<sup>37</sup>At the risk of stating the obvious: this definition of consistency just says that, given a valid argument, any assignment that makes the premises true must make the conclusion true as well.

$$f_0(R(x_1, x_2)) = \{\sigma \in A^{\mathcal{M}} \mid (\sigma(x_1), \sigma(x_2)) \in I(R)\}$$

$$F_\wedge(X, Y) = X \cap Y.$$

$$F_-(A^{\mathcal{M}}) = C, \text{ and if } X \neq A^{\mathcal{M}} \text{ then } F_-(X) = A^{\mathcal{M}}.$$

$$F_{\exists y}(X) = F_{\forall y}(X) = X \text{ for all variables } y.$$

$\mathcal{T}$  is clearly compositional, so it remains to show that (a)  $\mathcal{T}$  is non-normal, (b)  $\mathcal{T}$  is non-trivial, (c) the  $F_\diamond$  are invariant under permutations and (d)  $\mathcal{T}$  is consistent with  $\vdash_{\mathcal{L}_3}$ .

The fact that (a)  $\mathcal{T}$  is non-normal is obvious;  $F_-, F_\forall$  and  $F_\exists$  are not the intended operations. To see that (b)  $\mathcal{T}$  is non-trivial just note that  $\llbracket \neg \exists x_1 \exists x_2 R(x_1, x_2) \rrbracket^{\mathcal{T}} = C \neq A^{\mathcal{M}}$ . Next we need to prove that (c) the  $F_\diamond$  are invariant under permutations. This is obvious for  $F_\wedge$ , which is just intersection, and also for the operations in  $F_\exists$  and  $F_\forall$ , which are just the identity function. To show the same for  $F_-$  first note that, for any permutation  $\pi$  of the domain,  $\pi^*(A^{\mathcal{M}}) = A^{\mathcal{M}}$  and  $\pi^*(C) = C$ . In other words,  $A^{\mathcal{M}}$  and  $C$ , seen as 0-ary operations, are invariant under permutations. But then:

$$F_-(\pi^*(A^{\mathcal{M}})) = F_-(A^{\mathcal{M}}) = C = \pi^*(C) = \pi^*(F_-(A^{\mathcal{M}}))$$

$$F_-(\pi^*(X)) = A^{\mathcal{M}} = \pi^*(A^{\mathcal{M}}) = \pi^*(F_-(X)) \text{ for } X \neq A^{\mathcal{M}}$$

where the second identity rests on the fact that if  $X \neq A^{\mathcal{M}}$ , then  $\pi^*(X) \neq A^{\mathcal{M}}$  for any permutation  $\pi$ .

Finally, we need to prove that (d)  $\mathcal{T}$  is consistent with  $\vdash_{\mathcal{L}_3}$ . A trivial induction on formulas shows that for any  $\varphi$ :

$$(d_1) \llbracket \varphi \rrbracket^{\mathcal{T}} = A^{\mathcal{M}} \text{ or } \llbracket \varphi \rrbracket^{\mathcal{T}} = C.$$

$$(d_2) \llbracket \varphi \rrbracket^{\mathcal{T}} = A^{\mathcal{M}} \text{ iff } \llbracket \varphi \rrbracket^{\mathcal{M}} = A^{\mathcal{M}}.$$

$$(d_3) \llbracket \varphi \rrbracket^{\mathcal{T}} = C \text{ iff } \llbracket \varphi \rrbracket^{\mathcal{M}} = \emptyset.$$

Now we can prove the result by contraposition. Let  $\Gamma \cup \{\varphi\}$  be a set of sentences, and suppose that  $\bigcap_{\gamma \in \Gamma} \llbracket \gamma \rrbracket^{\mathcal{T}} \not\subseteq \llbracket \varphi \rrbracket^{\mathcal{T}}$ . By (d<sub>1</sub>) we must have that  $\llbracket \gamma \rrbracket^{\mathcal{T}} = A^{\mathcal{M}}$  for all  $\gamma \in \Gamma$  and  $\llbracket \varphi \rrbracket^{\mathcal{T}} = C$ . By (d<sub>2</sub>) and (d<sub>3</sub>) this means that  $\llbracket \gamma \rrbracket^{\mathcal{M}} = A^{\mathcal{M}}$  for all  $\gamma \in \Gamma$  and  $\llbracket \varphi \rrbracket^{\mathcal{M}} = \emptyset$ . Therefore,  $\Gamma \not\vdash_{\mathcal{L}_3} \varphi$ . Note, incidentally, that if we had taken  $\vdash_{\mathcal{L}_3}$  to be a relation between all formulas, open and closed, the proof would have still gone through. □

As before, Lemma 6.24 is stated for the very limited language  $\mathcal{L}_3$ , but can be adapted to richer ones: if a language has constants and function symbols, the

function  $I$  can give them any value whatsoever, and if it has further predicate letters  $P$  of arity  $n$ , we just need to set  $I(P) = D^n$ .

There is a second, more important sense in which Lemma 6.24 can be generalised. It turns out that if we take the semantic values of formulas to be sets of assignments, there are non-normal interpretations *even if every subset of  $D$  (and every subset of  $D^n$ , for any  $n$ ) is definable*.<sup>38</sup> The reason is simple: given a sufficiently large domain, certain sets of assignments can never be the semantic value of any formula on any normal interpretation. This allows us to construct non-normal interpretations that differ from normal ones only in the way they behave with respect to these ‘extra’ semantic values, and this is possible regardless of the interpretation of non-logical vocabulary. Let’s spell out the details.

The semantic value  $\llbracket \varphi \rrbracket^{\mathcal{M}}$  of an open formula  $\varphi$  on the normal interpretation based on a structure  $\mathcal{M}$  is the set of variable assignments  $\sigma$  such that  $\mathcal{M}, \sigma \models \varphi$ , where  $\models$  is the usual satisfaction relation. Clearly, whether  $\mathcal{M}, \sigma \models \varphi$  or not hinges on the value  $\sigma$  assigns to the (finitely many) free variables  $\vec{x}$  of  $\varphi$ . We need some terminology to describe this. Given a structure  $\mathcal{M}$  for  $\mathcal{L}$  and a (non-empty) finite sequence of variables  $x_1, \dots, x_n = \vec{x}$ , we will say that a set of assignments  $Y \subseteq A^{\mathcal{M}}$  *depends on  $\vec{x}$*  if there is a  $\sigma$  and a  $\sigma'$  that differ at most in the values they assign to some variables in  $\vec{x}$ , but such that  $\sigma \in Y$  and  $\sigma' \notin Y$ . A set  $Y \subseteq A^{\mathcal{M}}$  is *dependent* if it depends on some finite  $\vec{x}$ , and *independent* otherwise. Now, it is easy to check that:

**Observation 6.25.** The semantic value  $\llbracket \varphi \rrbracket^{\mathcal{M}}$  of an arbitrary formula  $\varphi$  in a normal interpretation based on  $\mathcal{M}$  is always  $A^{\mathcal{M}}$ ,  $\emptyset$ , or a dependent subset of  $A^{\mathcal{M}}$ .

Crucially, Observation 6.25 holds regardless of whether every subset of the domain  $D$  is definable. In addition:

**Observation 6.26.** Given a structure  $\mathcal{M}$  for  $\mathcal{L}$ , some independent subsets  $Y$  of  $A^{\mathcal{M}}$  are invariant under permutations.

For instance, given any structure  $\mathcal{M}$ , the set  $C_\infty$  of assignments that give the same value to (at least) countably infinitely many variables is independent and invariant under permutations. By combining Observations 6.25 and 6.26 we get that:

**Lemma 6.27.** *Let  $\mathcal{L}$  be any first-order language. Then there is a compositional, non-normal, non-trivial, invariant-under-permutations interpretation  $\mathcal{T}$  consistent with  $\vdash_{\mathcal{L}}$ .*

*Proof.* Take any  $\mathcal{L}$ -structure  $\mathcal{M} = (D, I)$  with  $|D| \geq \omega$ . By Observations 6.25 and 6.26 some independent subsets of  $A^{\mathcal{M}}$  are invariant under permutations.

<sup>38</sup>Adapting the notion of definability to this setting is trivial: the extension of a formula  $\varphi$  (relative to  $\sigma$  and  $x$ , on an interpretation  $\mathcal{I}$ ) is  $\llbracket \varphi \rrbracket_{\sigma, x}^{\mathcal{I}} := \{a \in D \mid \sigma[a/x] \in \llbracket \varphi \rrbracket^{\mathcal{I}}\}$ , and a subset  $X$  of  $D$  is definable if  $X = \llbracket \varphi \rrbracket_{\sigma, x}^{\mathcal{I}}$  for some  $\varphi, \sigma$  and  $x$ . This is extended to cover subsets of  $D^n$  in the obvious way.

Since  $|D| \geq \omega$ , we have that  $C_\infty \neq A^\mathcal{M}$ . Thus, some independent subsets of  $A^\mathcal{M}$  are invariant under permutations *and* not the semantic value of any formula under the normal interpretation based on  $\mathcal{M}$ . Next we will define a non-normal  $\mathcal{T} = (f_0, F_\wedge, F_\neg, F_\forall, F_\exists)$  based on  $\mathcal{M}$ . Let  $f_0$  interpret atoms normally, and let the  $F_\diamond$  behave like the normal operations except when applied to subsets of  $A^\mathcal{M}$  that are independent, invariant under permutations, and not the semantic value of a formula under the normal interpretation; for those sets (pairs of sets, in the case of  $F_\wedge$ ), the  $F_\diamond$  are the constant function to  $C_\infty$ . Then  $\mathcal{T} = (f_0, F_\wedge, F_\neg, F_\forall, F_\exists)$  is compositional, non-normal and invariant under permutations. Moreover, since the  $F_\diamond$  behave non-normally only for sets of assignments that are not the semantic value of a formula, and otherwise behave like the normal operations,  $\mathcal{T}$  is non-trivial and consistent with  $\vdash_{\mathcal{L}}$ . Note that, as mentioned above, this construction works regardless of how the function  $I$  interprets non-logical vocabulary □

It is time to take stock now. Bonnay and Westerståhl's strategy to solve Carnap's Problem is imposing restrictions on the interpretations we take into account, based on what they call semantic universals. There are three problems with this approach. The first, and more philosophical, is that it isn't clear that Bonnay and Westerståhl have adequately motivated their semantic constraints. The other two problems are more technical. First, Bonnay and Westerståhl's results rely on the use of predicate variables, and secondly, their definition of a first-order interpretation either violates compositionality or else begs the question against non-normal interpretations of connectives.

## 6.6 Another Approach to Carnap's Problem

So far I have been focussing on the shortcomings of different approaches to Carnap's Problem. In this section I will change tack, and propose an alternative solution. The approach I will explore is a slight modification of Bonnay and Westerståhl's, and the overall idea behind it is easy to explain: using second-order variables won't do, but if we look at our inferential practice, something similar to it can be justified.

The way forward is to focus on the notion of consistency with a consequence relation. Up until now we have identified consistency and truth-preservation: an interpretation was consistent with the classical consequence relation of a language if it made all classically valid arguments truth-preserving. A closer look at the way we draw inferences, though, suggests that this is too weak a requirement.

Consider the inference from '7 is a counterexample to Goldbach's conjecture and 8 is a counterexample to Goldbach's conjecture' to '7 is a counterexample to Goldbach's conjecture'. We accept this inference despite not knowing the extension of the predicate 'is a counterexample to Goldbach's conjecture' (and regardless of what it happens to be). More generally, we accept all inferences

of the form  $P(c_1) \wedge P(c_2) \vdash P(c_1)$  despite not knowing the extension of every predicate  $P$  and every individual constant  $c_1$  and  $c_2$  (and regardless of what they happen to be). If we want to account for the inferential role of  $\wedge$ , then, it isn't enough to give it a semantic value that makes our inferences truth-preserving *given that*  $P$ ,  $c_1$  and  $c_2$  are interpreted a certain way; we typically don't know the interpretation of non-logical vocabulary. To account for the inferential role of  $\wedge$ , it seems, we must give it a semantic value that makes valid arguments truth-preserving *regardless* of how of non-logical vocabulary is interpreted.

This is not an isolated example, but the norm. Take any inference of the form  $\forall xP(x) \vdash P(c)$ . We also accept it without knowing the extension of  $P$  or the reference of  $c$ , whatever they happen to be. To account for the role of  $\forall$  in inferences, then, we must give it a semantic value that makes arguments of this form truth-preserving no matter how  $P$  and  $c$  are interpreted.<sup>39</sup>

Let me make the point more general: so far we have read semantics off of consequence relations by looking at interpretations that make valid arguments truth-preserving. But to adequately represent the role of logical vocabulary, to do justice to the way we actually use it, is to give it semantic values that make valid arguments truth-preserving regardless of the interpretation of non-logical vocabulary. The notion of consistency, therefore, should be strengthened accordingly.

The technical set-up this requires is straightforward. Taking sets of assignments as the semantic values of formulas, as we saw, makes Carnap's Problem artificially difficult; it saddles us with more semantic values than formulas. Here we will go for a simpler and more natural option: the semantic values of formulas will be truth-values. The rest of modifications are obvious. In previous sections we built interpretations on top of structures  $\mathcal{M} = (D, I)$ . We are going to require truth-preservation across reinterpretations of non-logical vocabulary, so the function  $I$  has to go. Everything else is largely as before: given a domain  $D$ , an interpretation is a tuple  $\mathcal{T} = (F_\wedge, F_\neg, Q)$ , where  $F_\wedge$  and  $F_\neg$  are truth-functions interpreting  $\wedge$  and  $\neg$ , respectively, and  $Q$  is a set of subsets of  $D$ , interpreting  $\forall$ . When we add a function  $I$  for the non-logical vocabulary, formulas are evaluated in the obvious way:

$$\mathcal{T}_\sigma(Px_1, \dots x_n) = 1 \text{ iff } (\sigma(x_1), \dots, \sigma(x_n)) \in I(P).$$

$$\mathcal{T}_\sigma(\varphi \wedge \psi) = F_\wedge[\mathcal{T}_\sigma(\varphi), \mathcal{T}_\sigma(\psi)].$$

$$\mathcal{T}_\sigma(\neg\varphi) = F_\neg[\mathcal{T}_\sigma(\varphi)].$$

$$\mathcal{T}_\sigma(\forall x\varphi) = 1 \text{ iff } \{a \in D \mid \mathcal{T}_{\sigma[a/x]}(\varphi) = 1\} \in Q.$$

---

<sup>39</sup>A different way to argue for the same point is to note that the extension of non-logical predicates changes over time, but we continue to endorse valid arguments regardless of how it changes. To paraphrase McGee, if I am told that some red object has been destroyed, I don't pause to consider whether, given the new extension of 'red', the inference from 'Everything is red and sticky' to 'Everything is red' remains acceptable. To account for this fact, then, we must give  $\wedge$  and  $\forall$  semantic values that make the argument truth-preserving regardless of how non logical vocabulary is interpreted.

Finally, an interpretation  $\mathcal{T}$  is consistent<sup>+</sup> with  $\vdash_{\mathcal{L}}$  if it is truth-preserving given *any* function  $I$  for non-logical vocabulary. In other words, it is consistent<sup>+</sup> with  $\vdash_{\mathcal{L}}$  if, given any  $I$ , we have that  $\Gamma \vdash_{\mathcal{L}} \varphi$  and  $\mathcal{T}_{\sigma}(\Gamma) = 1$  imply  $\mathcal{T}_{\sigma}(\varphi) = 1$  (where  $\Gamma$  and  $\varphi$  can be formulas, not just sentences).<sup>40</sup>

If we look at things this way, it is easy to show that the role of classical logical vocabulary in inferences does rule out non-normal interpretations:

**Lemma 6.28.** *Let  $\mathcal{T} = (F_{\wedge}, F_{\neg}, Q)$  be an interpretation (with underlying domain  $D$ ) for a language  $\mathcal{L}$ . If  $\mathcal{T}$  is consistent<sup>+</sup> with  $\vdash_{\mathcal{L}}$ , then  $\mathcal{T}$  is normal.*

*Proof.* We will work by contraposition:

**Conjunction:** Suppose e.g. that  $F_{\wedge}(1, 1) = 0$ . Take any  $\varphi, \psi$  and  $\sigma$  such that  $\mathcal{T}_{\sigma}(\varphi) = \mathcal{T}_{\sigma}(\psi) = 1$  (there must be some such  $\varphi, \psi$  and  $\sigma$ , because  $\vdash_{\mathcal{L}}$  contains tautologies and  $\mathcal{T}$  is consistent<sup>+</sup> with it, so tautologies must have value 1 under any  $\sigma$ ). Since  $F_{\wedge}(1, 1) = 0$ , we have that  $\mathcal{T}$  invalidates  $\varphi, \psi \vdash_{\mathcal{L}} \varphi \wedge \psi$ . Remaining cases are similar.

**Negation:** Suppose e.g. that  $F_{\neg}(0) = 0$ . Set  $I(P) = \emptyset$  ( $P$  can be a predicate symbol of any arity, but for simplicity we will take it to be unary). Let  $\sigma$  be an arbitrary assignment. Since  $I(P) = \emptyset$  we must have  $\mathcal{T}_{\sigma}(Px) = 0$ . Then  $\mathcal{T}_{\sigma}(Px \wedge \neg Px) = 0$ , since one conjunct is false. But  $F_{\neg}(0) = 0$ , so  $\mathcal{T}_{\sigma}(\neg(Px \wedge \neg Px)) = 0$ . Thus,  $\mathcal{T}$  invalidates the tautology  $\vdash_{\mathcal{L}} \neg(Px \wedge \neg Px)$ . Remaining cases are similar.

**Universal Quantifier:** We have to show that  $\{D\} = Q$ . First, suppose for a contradiction that  $D \not\subseteq Q$ . Since  $\vdash_{\mathcal{L}} \neg(Px \wedge \neg Px)$ , we must have  $\mathcal{T}_{\sigma'}(\neg(Px \wedge \neg Px)) = 1$  for all  $\sigma'$ . Let's use the abbreviation  $\varphi := \neg(Px \wedge \neg Px)$  for readability. Now, let  $\sigma$  be arbitrary. Then  $\{a \in D \mid \mathcal{T}_{\sigma[a/x]}(\varphi) = 1\} = D \not\subseteq Q$ , so  $\mathcal{T}_{\sigma}(\forall x \neg(Px \wedge \neg Px)) = 0$ . Therefore,  $\mathcal{T}$  invalidates  $\vdash_{\mathcal{L}} \forall x \neg(Px \wedge \neg Px)$ , contradicting consistency<sup>+</sup>. Next, suppose that there is some  $C \subset D$  in  $Q$ . Let  $I(P) = C$ , and let  $\sigma'$  be an assignment such that  $\sigma'(x) \notin C$ . Then  $\{a \in D \mid \mathcal{T}_{\sigma'[a/x]}(Px) = 1\} = C \in Q$ , so  $\mathcal{T}_{\sigma'}(\forall x Px) = 1$ . But  $\mathcal{T}_{\sigma'}(Px) = 0$ , so  $\mathcal{T}$  invalidates  $\forall x Px \vdash_{\mathcal{L}} Px$ , contradicting consistency<sup>+</sup>.

□

A couple of comments are in order. First, it should be clear that quantifying over interpretations in the definition of consistency<sup>+</sup> has more or less the same effect as second-order variables in Bonnay and Westerståhl's account. The current option, however, has a clear motivation rooted in inferential practice.

<sup>40</sup>Bonnay and Westerståhl use a similar approach, *mutatis mutandis*, when they consider Carnap's Problem in the context of possible worlds semantics for propositional logic. They say quantifying over interpretations in the definition of consistency is only done "for simplicity" (Bonnay and Westerståhl 2016, p. 731), but it can be shown that it is a necessary condition for the strategy to work.

Moreover, it is a weaker assumption than the alternative. Predicate variables make every subset of  $D$  and  $D^n$  (for any  $n$ ) definable, whereas it is not in general the case that, given a language  $\mathcal{L}$  and a subset of  $D$  or  $D^n$ , we can always find a function  $I$  under which that subset is definable.

Secondly, the way I have defined interpretations presupposes what type of semantic value corresponds to each syntactic category: connectives are interpreted by truth-functions, and quantifiers by sets of subsets of the domain. This is something Bonnay and Westerståhl also take for granted:

One must choose the semantic values of expressions belonging to a given syntactic category [...] Our hypothetical language learner already knows, or guesses, what kind of language is to be learnt: what the syntactic categories are, and what kinds of things expressions of these categories stand for. (Bonnay and Westerståhl 2016, p. 726)

I will say more about this assumption in the next section. For now, it is enough to note that it doesn't make interpretations compositional. This is a feature, not a bug. As we have seen, compositionality (in the strict sense preferred by Bonnay and Westerståhl) is too strong a requirement, and part of what gets them into trouble. Recall also that, as we discussed in Section 6.5, Bonnay and Westerståhl motivate the requirement of compositionality through the usual learnability argument, but learnability arguments don't establish the need for compositionality in the strong sense they intend. In other words, nothing is lost with the switch to non-compositional interpretations.

The current approach also allows us to read classical semantics off of any standard calculus for classical logic. I will take  $\forall$  as an example, but the cases for  $\wedge$  and  $\neg$  are similar. Consider these typical natural deduction rules (in sequent notation) for  $\forall$ :

$$\frac{\Gamma \Rightarrow \varphi(y/x)}{\Gamma \Rightarrow \forall x\varphi} (\forall \text{ I}) \qquad \frac{\Gamma \Rightarrow \forall x\varphi}{\Gamma \Rightarrow \varphi(y/x)} (\forall \text{ E})$$

where  $y$  must be free for  $x$  in  $\varphi$ ,  $\varphi(y/x)$  is the result of replacing all free occurrences of  $x$  in  $\varphi$  for  $y$ , and  $y$  must not occur free in  $\Gamma$  or  $\forall x\varphi$ .

How should we adapt the notion of consistency<sup>+</sup> to rules? It is clear that we often draw inferences according to  $(\forall \text{ I})$  and  $(\forall \text{ E})$ —and take them to be valid—without knowing the interpretation of the non-logical vocabulary involved. Therefore, to adequately represent the inferential role of  $\forall$ , we must give it a semantic value that makes inferences sanctioned by  $(\forall \text{ I})$  and  $(\forall \text{ E})$  valid regardless of the interpretation of non-logical vocabulary. Let's make this more formal. Given an interpretation  $\mathcal{T}$  and some  $I$  for the non-logical vocabulary, we will say that  $\Delta \Rightarrow \psi$  is valid in  $\mathcal{T}$  if  $\mathcal{T}_\sigma(\Delta) = 1$  implies  $\mathcal{T}_\sigma(\psi) = 1$  (for any  $\sigma$ ). Moreover, an interpretation  $\mathcal{T} = (F_\wedge, F_\neg, Q)$  is consistent<sup>+</sup> with a rule  $\mathbf{R}$  if, given any  $I$ , whenever the premiss-sequents of  $\mathbf{R}$  are valid in  $\mathcal{T}$ , so is its conclusion-sequent. Then it is easy to show that:

**Lemma 6.29.** *Let  $\mathcal{T} = (F_\wedge, F_\neg, Q)$  be an interpretation (with underlying domain  $D$ ) for a language  $\mathcal{L}$ . If  $\mathcal{T}$  is consistent<sup>+</sup> with  $(\forall I)$  and  $(\forall E)$  then  $Q = \{D\}$ .*

*Proof.* We will assume for simplicity that  $\mathcal{L}$  has a unary predicate  $P$ , and work by contraposition. Suppose  $D \notin Q$ . Let  $I(P) = D$ . Then we have that  $\mathcal{T}_\sigma(Px) = 1$  and  $\mathcal{T}_\sigma(\forall xPx) = 0$  for any assignment  $\sigma$ . In other words,  $\Rightarrow Px$  is valid, and  $\Rightarrow \forall xPx$  is not. But we can infer  $\Rightarrow \forall xPx$  from  $\Rightarrow Px$  by  $(\forall I)$ , so  $\mathcal{T}$  is not consistent<sup>+</sup> with  $(\forall I)$ . Similarly, suppose there is some  $C \subset D$  in  $Q$ . Let  $I(P) = C$ , and let  $\sigma'$  be an assignment such that  $\sigma'(x) \notin C$ . Then we have that  $\mathcal{T}_\sigma(\forall xPx) = 1$  for any  $\sigma$ , but  $\mathcal{T}_{\sigma'}(Px) = 0$ . Therefore  $\Rightarrow \forall xPx$  is valid but  $\Rightarrow Px$  is not. Since we can infer the latter from the former using  $(\forall E)$ ,  $\mathcal{T}$  is not consistent<sup>+</sup> with  $(\forall E)$ .  $\square$

A similar argument works for any other standard set of natural deduction (or sequent) rules for  $\forall$ , and the corresponding result for the connectives is easy to prove.

Let's close the section with a brief 'check-list' comparison with the approaches we have seen so far. Garson's solution to Carnap's Problem, in the propositional case, relies either on non-triviality or on multiple conclusion rules. Both options are problematic, the first because we have been given no reason to assume non-triviality, and the second because, as we saw in Chapter 3, it isn't clear that multiple conclusions are an adequate representation of our inferential practice. In contrast, the current approach works with single conclusion rules and consequence relations, and doesn't assume that all interpretations are non-trivial.

The comparison with Garson's approach to the first-order case is more complicated. We rejected his 'substitutional' and 'objectual' definitions of normality:

$$(s\forall) v(\forall x\varphi) = 1 \text{ iff } v(\varphi[y/x]) = 1 \text{ for all variables } y \text{ of the language.}$$

$$(o\forall) v(\forall x\varphi) = 1 \text{ iff } v(\varphi[d/x]) = 1 \text{ for all } d \in D.$$

on the grounds that they are neither recognisably normal nor applicable to all languages. Neither of these problems arises for the current approach. I also proposed an alternative valuational set-up in Section 6.2.3. Comparison with this last approach will have to wait until the next section, when I say a bit more about interpretations in general.

We also saw that, from a technical point of view, the use of bilateral rules or consequence relations is equivalent to the use of multiple conclusions. This means that bilateral approaches avoid Carnap's Problem in propositional settings, but the negative results about the valuational semantics of 6.2.3 carry over to bilateral approaches. In contrast, the current approach secures the normal interpretation of quantifiers.

Now let's move on to McGee's proposal. In a certain sense open-endedness plays the same role as quantification over interpretations of non-logical vocabulary: it bypasses the expressive limitations of first-order languages, and ensures that each non-normal interpretation invalidates a classically valid inference. At the same time, open-endedness is stronger and more problematic than consistency<sup>+</sup>. This is, in the first place, because it isn't clear what the range of mathematically possible extensions of a language consists in, and the extent to which we need to modify the notion of a model as we enrich first-order languages with arbitrary sentences.<sup>41</sup> But secondly, it is unclear that acceptance of the rules of classical logic actually *is* open-ended. For instance, it is fairly common to hold that classical logic has to be abandoned when first-order languages are extended in certain ways (say, with vague predicates, or with truth-predicates).<sup>42</sup> In contrast, here we only assume that an adequate interpretation must make valid arguments truth-preserving under arbitrary 'well-behaved' interpretations of the non-logical vocabulary *that we already have* (where 'well-behaved' just means that the interpretations give standard, unproblematic semantic values to non-logical vocabulary, and don't smuggle in vagueness, self-reference and the like).

Incidentally, Murzi and Topey (Murzi and Topey 2021) have recently put forward a solution similar to McGee's. Like McGee, they attempt to read classical semantics from rules of inference, and take those rules to be open-ended. They cash out open-endedness more modestly than McGee, though: for Murzi and Topey rules must remain valid across all extensions of a language obtained by adding new predicate letters and individual constants. This is much less problematic than McGee's talk of "all mathematically possible extensions of a language", and only slightly stronger than our assumption that inference rules must remain valid across arbitrary interpretations of the *current* non-logical vocabulary. Murzi and Topey also assume which type of semantic value can be given to quantifiers.<sup>43</sup> And lastly, the inference rules from which they read the semantics are slightly unusual: their (2021) uses higher-order rules in the style of (Schroeder-Heister 1984), but their approach also works for ('lower-order') natural deduction or sequent rules that allow empty succedents.<sup>44</sup> In contrast, the current approach reads the usual semantics from any standard calculus for classical logic.

Finally, the comparison with Bonnay and Westerståhl's proposal is straightforward: their approach assumes compositionality, non-triviality, invariance under permutations, and doesn't work for (all) first-order languages, since it is essentially tied to the use of predicate variables. The current approach, on the other hand, starts from a similar notion of interpretation and only uses

<sup>41</sup>Not to mention McGee's additional assumptions about the definability of all classes of models, as well as ad hoc add-ons like **(C1o)**.

<sup>42</sup>See e.g. (Beall 2003) for some references.

<sup>43</sup>Their language sometimes suggests they take the denotation of  $\forall$  to be a *subset* of the domain of quantification, rather than a set of subsets (see Murzi and Topey 2021, p. 3407). If this is so, it is unclear how the approach could be extended to other quantifiers like  $\exists$ .

<sup>44</sup>I owe this last point to Julien Murzi.

consistency<sup>+</sup>, a substitute for predicate variables motivated by our inferential practice.

## 6.7 Carnap's Problem and Interpretations

Let's take a step back and put what we have done so far in perspective. We began with a relatively simple question: whether the inferential role of (classical) logical vocabulary rules out its non-normal interpretations. This question is sensible only insofar as we can explain three things:

- (i) What we mean by 'interpretations'.
- (ii) What we mean by 'inferential roles'.
- (iii) What we mean by inferential roles 'ruling out' interpretations.

We have seen the two standard ways of spelling out what inferential roles amount to: in terms of consequence relations and in terms of inference rules. The approach I have put forward works with both. We have also seen that 'ruling out' is usually cashed out in terms of truth (or validity) preservation. I have argued that it is better explained in terms of truth (or validity) preservation across all reinterpretations of non-logical vocabulary, or consistency<sup>+</sup>. This takes care of (ii) and (iii); now it is time to say more about (i).

By and large, the literature on Carnap's Problem is split into two camps. According to the first interpretations are assignments of truth-conditions to formulas (or sentences). According to the second, interpretations are assignments of semantic values to subsentential expressions.

Garson's work and the Garson-like semantics of sections 6.2.3 and 6.3 are clear examples of the first type of approach. In both of them interpretations assign semantic values only to formulas, and non-logical vocabulary is treated syncategorematically.

McGee is somewhere in between the sentential and sub-sentential approaches. Officially his interpretations are assignments of semantic values to sub-sentential expressions. At the same time, normality is defined in terms of the truth-conditions of full sentences, so the specific semantic value of logical expressions is somewhat irrelevant.

The fully sub-sentential approach is also common. The usual set-up (at the propositional level) has interpretations assign truth-functions to connectives.<sup>45</sup> Bonnay and Westerståhl extend this to the first-order case by taking quantifiers to be interpreted by sets of subsets of the domain, and I have followed suit. Seeing this, it is natural to ask: is there any reason to prefer sentential or subsentential definitions of interpretations?

The idea that sub-sentential expressions have meanings is perhaps the orthodoxy—to the extent that there is an orthodoxy—both in philosophy and formal semantics. If we take this idea on board, it is reasonable to demand

---

<sup>45</sup>See e.g. (Peacocke 2004), (Hodes 2004), (Hacking 1979) or, in a different setting, (Button and Walsh 2018).

that interpretations assign semantic values to individual expressions, and to define  $\diamond$ -normal interpretations in terms of the specific semantic value of the expression ' $\diamond$ '. There are, however, well-known challenges to this orthodoxy. On Davidson's program, for example, interpreting a language is ultimately a matter of assigning truth-conditions to its sentences, and "words, meanings of words, reference, and satisfaction are posits we need to implement a theory of truth" (Davidson 2001a, p. 222). As Davidson puts it:

It makes no sense, on this approach, to complain that a theory comes up with the right truth conditions time after time, but has the logical form (or deep structure) wrong. We should take the same view of reference. (Davidson 2001a, p. 223)

From this perspective it is misguided to define normal interpretations in terms of the semantic value of sub-sentential expressions, rather than in terms of the truth-conditions of whole sentences. I will not attempt to settle this general debate here; it goes well beyond Carnap's Problem and the scope of this thesis, and worse still, I don't know which side of the debate is right. What I will try to do instead is signpost some false leads, and say a few words in favour of the way I have defined interpretations.

First, it may seem that the sentential approach to valuations is more congenial to inferentialism. Inferentialists take meaning to consist in, or be determined by, the role linguistic expressions play in inferences, and it is primarily sentences that are inferentially articulated: we don't draw inferences between individual words. Be that as it may, it would be rushed to conclude that sub-sentential interpretations are incompatible with inferentialism. It is one thing to say that inferentialist semantics must begin by ascribing contents to sentences, and a different one to say that it can disregard the meaning of individual words. Indeed, many prominent inferentialists put a lot of emphasis on talk about the meaning of specific logical operators (recall the notion of separability from Chapter 2, and its importance for the discussion of Harmony in Chapter 3).

Secondly, it could be tempting to argue that sentential interpretations are preferable because they 'assume less' about the languages we are interested in. Sub-sentential interpretations take for granted some division of the expressions of a language into different syntactic categories, and the type of semantic value that expressions of each category can be given. To some extent sentential interpretations do the same: they assume, at least, a division between the syntactic category 'formula' (or 'sentence') and all others, and the semantic value that formulas or sentences can have.<sup>46</sup> Of course, some degree of stipulation is inevitable, and any way to define interpretations will rule out some options from the get go. The worry, however, is that in building more into the definition of interpretations, sub-sentential approaches make Carnap's Problem easier.

This objection has more bite, except for the fact that it is not clear what 'easier' means in this context. Stipulating the kind of semantic value that cor-

---

<sup>46</sup>For instance, when defining interpretations Garson and McGee assume that there are exactly two truth-values, ruling out gaps, gluts, and supervaluations.

responds to different syntactic categories does rule out some assignments of truth-conditions to complete sentences. For instance, no assignment of truth-functions to the propositional connectives can result in the valuation  $v_{cl}$  such that  $v_{cl}(A) = 1$  iff  $A$  is a classical tautology (this is what sits at the core of Bonnay and Westerståhl’s Theorem 6.19). At the same time, on the sub-sentential approach interpretations can be non-normal even when they assign all formulas the same truth-value as a normal interpretation would; this was one of the upshots of Lemma 6.20. In other words, the sub-sentential approach rules out some sentential interpretations from the start, but also introduces more a demanding notion of normality.

Now for some good news. Deciding, once and for all, whether the sentential or the sub-sentential approach is to be preferred is a tall order, but we can at least evaluate individual proposals on their own merits. I rejected Garson’s substitutional reading of the quantifiers:

$$(s\forall) v(\forall x\varphi) = 1 \text{ iff } v(\varphi[y/x]) = 1 \text{ for all variables } y \text{ of the language.}$$

on the grounds that it doesn’t match the standard truth-conditions of quantified sentences, and his objectual definition of normal interpretations:

$$(o\forall) v(\forall x\varphi) = 1 \text{ iff } v(\varphi[d/x]) = 1 \text{ for all } d \in D.$$

on the grounds that it is not applicable to arbitrary languages. This was also the reason I took issue with McGee’s definition of normality:

$$(\forall)_m \text{ If } c \text{ does not appear in } \theta, \text{ then } \mathcal{M} \models \forall x\theta(x) \text{ iff } \mathcal{M}' \models \theta[c/x] \text{ in every } c\text{-variant } \mathcal{M}' \text{ of } \mathcal{M}.$$

As we saw in Section 6.4,  $(\forall)_m$  is only applicable to languages with infinitely many individual constants. In that respect our definition of normal interpretations does better, since it is recognisably normal and applicable to arbitrary languages. There are other ways to define interpretations which fare equally well on both accounts, like the first-order valuational semantics I considered in Sections 6.2.3 and 6.3. But then again, it is not surprising that there is more than one reasonable thing to mean by ‘interpretation’.

The interpretations I have used are a natural, straightforward generalisation of standard first-order semantics: logical expressions can take any semantic value of the type they usually receive. Some could complain that this is still not general enough, that I am building ‘too much’ into the notion of an interpretation. The complaint presupposes some theory of how much a definition of interpretations can take for granted before it is ‘too much’, and I am not sure that such a theory is available. But in any case, readers who think my definition of interpretations is too narrow can relativise the results of Section 6.6. What Section 6.6 shows is that, *if* connectives and quantifiers are interpreted with the usual semantic types, then they must be given their normal semantic values. Or put more suggestively: that if connectives and quantifiers are interpreted *as* connectives and quantifiers, they must be interpreted normally. This

is not the last word on Carnap’s Problem. It couldn’t be: there are as many ‘Carnap’s Problems’ as there are ways to define interpretations. But it does, I hope, help to clarify the relation between what logical constants mean and how they are used in arguments.

## 6.8 Coda: Harmony Again

We are almost done, but it is never too late to reminisce. The first problem we examined, back in Chapter 3, was the Problem of Harmony. This problem arises from the combination of two claims:

**(AV):** Some inferences are valid in virtue of the meaning of expressions that occur in them.

**(AR):** Any set of rules can define the meaning of an expression that it governs.

As Prior pointed out, it seems to follow from **(AV)** and **(AR)** that any sentence can be validly inferred from any other. By **(AR)** we can define a connective ‘tonk’ with the following rules:

$$\frac{A}{A \text{ tonk } B} \text{ (tonk I)} \qquad \frac{A \text{ tonk } B}{B} \text{ (tonk E)}$$

Since (tonk I) and (tonk E) supposedly define the meaning of tonk, by **(AV)** all inferences carried out according to them are valid in virtue of the meaning of ‘contoktion’. But of course, consecutive applications of (tonk I) and (tonk E) allow us to derive any two sentences from each other.

In Chapter 3 we focussed on solutions to the Problem of Harmony that lay down syntactic restrictions limiting which rules we can take as definitions. Now that we know more about moderate inferentialism, we can revisit the idea from a different point of view. Moderate inferentialists think the meaning of logical expressions is determined by inference rules. But, just like strict inferentialists claim that some (sets of) rules fail to define a connective, moderate inferentialists can claim that some (sets of) rules fail to determine a meaning. This is the line Peacocke takes in his (2004):

In showing the legitimacy of introducing a constant as conforming to certain laws, the fundamental task for a realist must be to demonstrate that *there is* a semantical value for it which makes those laws necessarily truth-preserving. I will refer to the requirement that there exists such a value as ‘the Semantic Constraint’. (Peacocke 2004, p. 167)<sup>47</sup>

---

<sup>47</sup>A realist, for Peacocke, is “a theorist who allows that a sentence or content can be true though unverifiable by us.” (Peacocke 2004, p. 153).

According to Peacocke, the problem with the rules for tonk is that they fail to pick out a semantic value. By this he means something quite specific:

For the realist, the fundamental objection to Prior’s runabout inference ticket tonk is semantical. [...] The semantical objection to tonk is that *there is no binary function on truth values* which validates both its introduction and its elimination rules. (Peacocke 2004, p. 167, my italics)

The underlying assumption is that, since tonk is a binary sentential connective like conjunction or disjunction, it should receive the same type of semantic value as them. Once again, though, we face the problem of deciding whether Peacocke is assuming ‘too much’ about semantic values. In this specific case, however, the point is less problematic than in the previous section: something like Peacocke’s argument goes through regardless of what we take as the range of possible semantic values of tonk. Suppose only that sentences have exactly one of two semantic values, True and False. Peacocke’s Semantic Constraint requires that the meaning of tonk —if there is one— should make (tonk I) and (tonk E) necessarily truth-preserving. It follows that such a meaning should make the tonk rules truth-preserving regardless of the interpretation of non-logical vocabulary. But now consider two sentences  $P(c)$  and  $Q(c)$ , where  $P, Q$ , and  $c$  interpreted so that  $P(c)$  has semantic value True and  $Q(c)$  has semantic value False. The tonk rules allow us to infer  $Q(c)$  from  $P(c)$ , and therefore (tonk I) and (tonk E) don’t preserve truth. Whatever semantic values are, generally speaking, no semantic value can do the job the tonk rules would require it to.

This second argument still depends on two substantial assumptions. First, it assumes that sentences are either true or false (and never both), and secondly, it assumes that rules pick out (subsentential) interpretations which make them necessarily truth-preserving. Cook (2005) and Fjellstad (2015) have pointed out that, without these assumptions, it is possible to find a meaning (of sorts) for tonk. Suppose that sentences can be true (T), false (F), neither true nor false (N), or both true and false (B), as in the logic FDE. Next define an ad hoc notion of logical consequence —call it  $\text{consequence}_t$ — which tracks preservation of either truth or non-falsity (where truth and falsity aren’t mutually exclusive). A bit more formally:

$\Gamma \models_t \varphi$  iff either **(a)** or **(b)** hold:

**(a)** whenever all  $\gamma \in \Gamma$  are T or B, so is  $\varphi$ .

**(b)** whenever  $\varphi$  is F or B, so is some  $\gamma \in \Gamma$ .

If we demand that rules be sound for  $\text{consequence}_t$ , it is possible to interpret tonk with a function on the values T,F,N, and B (see Cook 2005 for details). Fjellstad describes a similarly baroque environment in which sentences receive different interpretations depending on whether they occur as premises or conclusions, and proves that his set-up allows for an interpretation of tonk (see Fjellstad 2015).

What should we make of these results? Cook and Fjellstad show that, in the context of a tailor-made model theory, it is possible to assign tonk a denotation. At the same time, one could wonder whether this denotation is worth calling a *semantic* value. This takes us to something we briefly touched upon in Chapter 2, when introducing the idea of moderate inferentialism:

It is possible to associate many sorts of things with sentences and other linguistic expressions. What makes the association a *semantic* one is precisely the possibility of appealing to it to explain the proprieties that govern the use of those expressions. Calling what one associates with expressions ‘contents’, ‘propositions’, ‘sets of possible worlds’, ‘truth conditions’, ‘extensions’, or ‘referents’ is at best issuing a promissory note that hints at how what are put forward as their semantic correlates ought to be taken to be relevant to determining how those expressions are correctly used. [...] Semantics answers to pragmatics, attributions of content to explanations of use. (Brandom 1994, pp. 187-8)

Following Brandom, we can distinguish between a formal semantics —roughly, a systematic assignment of something-or-others to linguistic expressions— and a semantics proper, that is, a systematic assignment of something-or-others to linguistic expressions by appeal to which we can explain proprieties of linguistic use. This is a *very* vague distinction, but even vague distinctions can be helpful when dealing with extreme cases like tonk. Both Cook and Fjellstad are careful to note that their respective set-ups aren’t intended for the analysis of our inferential practice, or more generally, for the analysis of ‘proprieties of use’ in Brandom’s sense.<sup>48</sup> In this sense, their results are compatible with the (fuzzy) idea that no ‘proper semantics’ validates the rules for tonk.

Let’s take stock and regroup. We have seen that if we assume that (a) all sentences are true or false (and never both), and (b) rules pick out interpretations that make them necessarily truth-preserving, then the rules for tonk don’t pick out a semantic value. We have also seen that, if we let go of assumptions (a) and (b), it is possible to give tonk a denotation, but it isn’t clear whether we should call this denotation a semantic value. Those with a strong classical bent will share Peacocke’s assumptions, and are bound to be content with this situation. Most non-classical logicians, however, will reject (a), (b), or both, and so for them Peacocke’s solution isn’t available (indeed, it isn’t a solution at all). This takes us to the moral Cook himself extracts from his result:

As a more general lesson, the existence of Tonk-Logic suggests that discussion of the legitimacy of logical connectives should focus not only on which ones are and are not acceptable *simpliciter*, but should also take into account which connectives are acceptable in which context. [...] A mismatch between our account of the general acceptability of a particular connective (such as tonk) and its acceptability

---

<sup>48</sup>See (Cook 2005, p. 221) and (Fjellstad 2015, p. 16), respectively.

within a particular formal logic (such as Tonk-logic) should then lead us to rethink both the acceptability of the connective itself and the legitimacy of the particular conception of consequence in question. What the present example illustrates, however, is that such mismatches always present us with a choice. (Cook 2005, pp. 223-4)

Moderate inferentialists may want to rephrase Cook's point in terms of meaning and meaning-determination, but the idea seems important nonetheless. Whether a set of rules picks out a semantic value depends on what we mean by 'picking out' and by 'semantic value'. In cases like tonk the choice between rejecting an expression as meaningless, on the one hand, or analysing linguistic use in terms of a contrived semantics, on the other, is likely to be obvious. To that extent, classical logicians and (most) non-classical logicians will agree that tonk is meaningless. In other cases, though, the situation is bound to be less clear. The fact that a certain set of rules is validated by something we are comfortable calling a semantics may lead us to accept the rules in question, and conversely, the fact that a given semantics can't accommodate expressions we would hesitate to call meaningless might lead us to search for a different semantics.<sup>49</sup>

The picture of harmony —if it can still be called that— which emerges is unusual, but perhaps worth more attention than it has received. What counts as a defective set of rules can't be decided from the outset. Whether a rule picks out a semantic value depends on the background language, and can only be decided after we weigh in the suitability of a complete semantic framework. In some fringe cases the answer will be reasonably clear, but generally speaking no one-size-fits-all criterion is likely to be useful. This is, perhaps, less satisfying that the sharp(er) criteria we reviewed in Chapter 3. If Cook is on the right track, though, this outcome is inevitable; after all, our notion of what fails to determine a meaning can't be any less vague than our notion of what counts as meaning in the first place. And we will leave it at that, so as to end, as we started, with a platitude.

---

<sup>49</sup>Perhaps some good examples of this interplay are the change in logician's perception of modal logics before and after Kripke's work, and the strict-tolerant semantics Ripley has advocated for in order to extend classical logic with a well-behaved truth-predicate (Ripley 2012). Note that in strict-tolerant semantics logical consequence is not transitive, and the rejection of transitivity leads Ripley to doubt that there is, all things considered, something wrong with tonk (Ripley 2015).

## Chapter 7

# Concluding Remarks

And when you've finished writing it  
I think you'll find a great joy in it,  
or so...

---

Glenn Gould, *So You Want to  
Write a Fugue*

Inferentialism is one way to flesh out the slogan that use determines meaning. It is especially promising when applied to logical vocabulary, but even then, developing an inferentialist theory of meaning involves taking stances on some difficult debates. Here I have focused on the Problem of Harmony, the Problem of Meaning-Variance, and Carnap's Problem. In each case my goal has been to help clear some roadblock, or suggest new avenues for inferentialists to explore.

In the case of proof-theoretic harmony, I have argued that the qualms inferentialists have traditionally had about classical logic should be dismissed. This is for two reasons. The first is that close inspection of the usual criteria of harmony shows that certain formalisations of classical logic are harmonious according to many standards, and that there are serious problems with the stricter harmony criteria that rule out classical logic. The second reason is that it is possible to give bilateral calculi for classical logic that satisfy all standard harmony criteria, even when these are adapted (strengthened, really) so as to suit the bilateralist framework.

In the case of meaning-variance the situation is more complicated. I have taken issue with Quine's arguments for the Meaning-Variance thesis, the Use and the Translation arguments. At the same time, I have tried to show that current minimalist proposals are untenable. This still leaves open the question of whether partisans of different logics use different logical vocabulary. But I have also suggested that there may not be an answer to this question (at least not in all cases), and that, regardless, the meaning-variance debate has less momentous consequences than it is usually made out to have.

Finally, I have suggested one way to 'solve' Carnap's Problem, that is, a

way of formulating moderate inferentialism that doesn't run into complications about the categoricity of logical vocabulary, and I have argued that this approach avoids some of the issues that affect related proposals.

Along the way some of the connections between these topics have come to the surface. Whether structural rules partly determine the meaning of connectives plays a role in the harmony debate (for instance, in conservativeness-based criteria), as well as in the meaning-variance debate. Similarly, a choice concerning which rules determine the same meaning must inform any stance on harmony: presumably rules that determine the same meaning must fall or stand together, as far as harmony goes. Bilateralism also cuts across the problems we have looked at. It can be seen as a way to approach the harmony debate, but also goes some way towards solving Carnap's Problem. And adopting a moderate inferentialist point of view, in its turn, allows us to see the harmony debate in a new light.

Many other points of contact between our three problems have been left unexplored, and there is further work to be done regarding each of them. The normalisation results of Chapter 4 should be extended to first-order calculi. A deeper investigation into the consequences of the approach to meaning variance suggested in Chapter 5 is called for. And Chapter 6 suggests that the debate surrounding Carnap's Problem would benefit from a clearer account of what can go into the notion of an interpretation, and the role model-theoretic semantics plays in moderate inferentialism. But while the work is not finished, I hope, at least, to have gone some way into clarifying the prospects of inferentialist approaches to logic.

# Bibliography

- Adams, Fred and Ken Aizawa (2021). “Causal Theories of Mental Content”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2021. Metaphysics Research Lab, Stanford University.
- Antonelli, Aldo (2013). “On the General Interpretation of First-Order Quantifiers”. In: *Review of Symbolic Logic* 6.4, pp. 637–658.
- (2017). “Completeness and Decidability of General First-Order Logic”. In: *Journal of Philosophical Logic* 46.3, pp. 233–257.
- Avron, Arnon (1988). “The Semantics and Proof Theory of Linear Logic”. In: *Theoretical Computer Science* 57.2, pp. 161–184.
- (1991). “Simple consequence relations”. In: *Information and Computation* 92.1, pp. 105–139. ISSN: 0890-5401. DOI: [https://doi.org/10.1016/0890-5401\(91\)90023-U](https://doi.org/10.1016/0890-5401(91)90023-U). URL: <https://www.sciencedirect.com/science/article/pii/089054019190023U>.
- Barwise, Jon and John Perry (1985). “Shifting Situations and Shaken Attitudes”. In: *Linguistics and Philosophy* 8.1, pp. 105–161. DOI: 10.1007/bf00653379.
- Beall, J. C., ed. (2003). *Liars and Heaps: New Essays on Paradox*. Oxford, England: Oxford University Press UK.
- Belnap, Nuel (1962). “Tonk, Plonk and Plink”. In: *Analysis* 22.6, pp. 130–134.
- Belnap, Nuel and Gerald J. Massey (1990). “Semantic Holism”. In: *Studia Logica* 49.1, pp. 67–82.
- Berger, Alan (1990). “A Central Problem for a Speech-Dispositional Account of Logic and Language”. In: *Perspectives on Quine*. Ed. by Robert B. Barrett and Roger F. Gibson. Cambridge, Mass., USA: Blackwell.
- Birkhoff, Garrett and John Von Neumann (1936). “The Logic of Quantum Mechanics”. In: *Annals of Mathematics* 37.4, pp. 823–843.
- Boghossian, Paul (1994). “Inferential Role Semantics and the Analytic/Synthetic Distinction”. In: *Philosophical Studies* 73.2-3, pp. 109–122. DOI: 10.1007/bf01207660.
- (1996). “Analyticity Reconsidered”. In: *Noûs* 30.3, pp. 360–391. DOI: 10.2307/2216275.
- (2000). “Knowledge of Logic”. In: *New Essays on the A Priori*. Ed. by Paul Boghossian and Christopher Peacocke.

- Boghossian, Paul (2003). “Blind Reasoning”. In: *Supplement to the Proceedings of the Aristotelian Society* 77.1, pp. 225–248. DOI: 10.1111/1467-8349.00110.
- Bonevac, Daniel (1985). “Quantity and Quantification”. In: *Noûs* 19.2, pp. 229–247.
- Bonnay, Denis (2014). “Logical Constants, or How to Use Invariance in Order to Complete the Explication of Logical Consequence”. In: *Philosophy Compass* 9.1, pp. 54–65. DOI: 10.1111/phc3.12095.
- Bonnay, Denis and Dag Westerståhl (2016). “Compositionality Solves Carnap’s Problem”. In: *Erkenntnis* 81.4, pp. 721–739. DOI: 10.1007/s10670-015-9764-8.
- Boričić, Branislav R. (1985). “On Sequence-Conclusion Natural Deduction Systems”. In: *Journal of Philosophical Logic* 14.4, pp. 359–377. DOI: 10.1007/bf00649481.
- Brandom, Robert (1983). “Asserting”. In: *Noûs* 17.4, pp. 637–650. DOI: 10.2307/2215086.
- (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, Mass.: Harvard University Press.
- (2000). *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, Mass.: Harvard University Press.
- (2007). “Inferentialism and Some of its Challenges”. In: *Philosophy and Phenomenological Research* 74.3, pp. 651–676. DOI: 10.1111/j.1933-1592.2007.00044.x.
- Brîncuş, Constantin C. (2024). “Categorical Quantification”. In: *The Bulletin of Symbolic Logic*, pp. 1–27. DOI: 10.1017/bsl.2024.3.
- Bryant, Amanda (2017). “Resolving Quine’s Conflict: A Neo-Quinean View of the Rational Revisability of Logic”. In: *Australasian Journal of Logic* 14.1. DOI: 10.26686/ajl.v14i1.4026.
- Burgess, John P. (2005). “No Requirement of Relevance”. In: *Oxford Handbook of Philosophy of Mathematics and Logic*. Ed. by Stewart Shapiro. Oxford University Press, pp. 727–750.
- Button, Tim and Sean P. Walsh (2018). *Philosophy and Model Theory*. Ed. by Sean Walsh and Wilfrid Hodges. Oxford, UK: Oxford University Press.
- Carnap, Rudolf (1937). *The Logical Syntax of Language*. Ed. by Amethe Smeaton. London, K. Paul, Trench, Trubner & co.
- (1943). *Formalization of Logic*. Cambridge, Mass., Harvard university press.
- Ceragioli, Leonardo (2022). “New Problems for Tennant’s Definition of Harmony”. In: *Theoria* 88.4, pp. 829–849. DOI: 10.1111/theo.12400.
- Cook, Roy T. (2005). “What’s Wrong with Tonk”. In: *Journal of Philosophical Logic* 34.2, pp. 217–226. DOI: 10.1007/s10992-004-7805-x.
- Davidson, Donald (1967). “Truth and Meaning”. In: *Synthese* 17.1, pp. 304–323. DOI: 10.1007/bf00485035.
- (1973). “Radical Interpretation”. In: *Dialectica* 27.1, pp. 314–328. DOI: 10.1111/j.1746-8361.1973.tb00623.x.
- (1974). “Belief and the Basis of Meaning”. In: *Synthese* 27.July-August, pp. 309–323. DOI: 10.1007/bf00484597.

- Davidson, Donald (1979). “The Inscrutability of Reference”. In: *Southwestern Journal of Philosophy* 10.2, pp. 7–19. DOI: 10.5840/swjphil197910229.
- (1984a). *Inquiries Into Truth and Interpretation*. Oxford: Oxford University Press.
- (1984b). “The Method of Truth in Metaphysics”. In: *Inquiries Into Truth and Interpretation*. Oxford University Press, pp. 199–214.
- (1984c). “Thought and Talk”. In: *Inquiries Into Truth and Interpretation*. Oxford University Press, pp. 155–170.
- (Sept. 2001a). “Reality Without Reference”. In: *Inquiries into Truth and Interpretation*. Oxford University Press.
- (2001b). “Three Varieties of Knowledge”. In: *Subjective, Intersubjective, Objective*. Ed. by Donald Davidson. Clarendon Press.
- (2004). “A Unified Theory of Thought, Meaning, and Action”. In: *Problems of Rationality*. Ed. by Donald Davidson. Oxford University Press.
- Dicher, Bogdan (2016). “A Proof-Theoretic Defence of Meaning-Invariant Logical Pluralism”. In: *Mind* 125.499, pp. 727–757. DOI: 10.1093/mind/fzv214.
- (2020). “Hopeful Monsters: A Note on Multiple Conclusions”. In: *Erkenntnis* 85.1, pp. 77–98. DOI: 10.1007/s10670-018-0019-3.
- Dicher, Bogdan and Francesco Paoli (2018). “The Original Sin of Proof-Theoretic Semantics”. In: *Synthese* 198.1, pp. 615–640. DOI: 10.1007/s11229-018-02048-x.
- Došen, Kosta (1989). “Logical Constants as Punctuation Marks”. In: *Notre Dame Journal of Formal Logic* 30.3, pp. 362–381. DOI: 10.1305/ndjfl/1093635154.
- Došen, Kosta and Peter Schroeder-Heister (1985). “Conservativeness and uniqueness”. In: *Theoria* 51.3, pp. 159–173. DOI: <https://doi.org/10.1111/j.1755-2567.1985.tb00093.x>.
- Dretske, Fred (2000). *Perception, Knowledge and Belief: Selected Essays*. New York: Cambridge University Press.
- Duhem, Pierre Maurice Marie (1954). *The Aim and Structure of Physical Theory*. Princeton, Princeton University Press.
- Dummett, Michael (1973). *Frege: Philosophy of Language*. London: Duckworth.
- (1975). “The Philosophical Basis of Intuitionistic Logic”. In: *Truth and Other Enigmas*. Ed. by Michael Dummett. Harvard UP, pp. 215–247.
- (1978). “The Justification of Deduction”. In: *Truth and other enigmas*. Ed. by Michael Dummett. Harvard University Press.
- (1991). *The Logical Basis of Metaphysics*. Cambridge: Harvard University Press.
- (2006). *Thought and Reality*. New York: Oxford University Press.
- Ferreira, Fernando (2008). “The Co-Ordination Principles: A Problem for Bilateralism”. In: *Mind* 117.468, pp. 1051–1057. DOI: 10.1093/mind/fzn036.
- Feyerabend, Paul K. (1962). “Explanation, Reduction and Empiricism”. In: *Scientific Explanation, Space, and Time, (Minnesota Studies in the Philosophy of Science, Volume III)*. Ed. by H. Feigl and G. Maxwell, pp. 103–106.
- Fiore, Camillo and Lucas Rosenblatt (2023). “Recapture Results and Classical Logic”. In: *Mind* 132.527, pp. 762–788. DOI: 10.1093/mind/fzad006.

- Fjellstad, Andreas (2015). “How a Semantics for Tonk Should Be”. In: *Review of Symbolic Logic* 8.3, pp. 488–505. DOI: 10.1017/s1755020314000513.
- Fodor, Jerry A. (1990). *A Theory of Content and Other Essays*. MIT Press.
- Francez, Nissim (2014). “Bilateralism in Proof-Theoretic Semantics”. In: *Journal of Philosophical Logic* 43.2-3, pp. 239–259. DOI: 10.1007/s10992-012-9261-3.
- Frege, Gottlob (1997). *Negation*. Ed. by Michael Beaney. Cambridge: Blackwell.
- (2002). “From Frege to Gödel: a sourcebook in mathematical logic”. In: *Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought*. Ed. by Jean Van Heijenoort. Harvard UP.
- García-Carpintero, Manuel (2012). “Foundational Semantics I: Descriptive Accounts”. In: *Philosophy Compass* 7.6, pp. 397–409. DOI: 10.1111/j.1747-9991.2012.00484.x.
- Garson, James W. (2013). *What Logics Mean: From Proof Theory to Model-Theoretic Semantics*. New York: Cambridge University Press.
- Gentzen, Gerhard (1969). *The Collected Papers of Gerhard Gentzen*. Ed. by M. E. Szabo. Amsterdam, North-Holland Pub. Co.
- Glüer, Kathrin (2011). *Donald Davidson: A Short Introduction*. New York, US: Oup Usa.
- Goldfarb, Warren (2016). “On Dummett’s “Proof-Theoretic Justifications of Logical Laws””. In: *Advances in Proof-Theoretic Semantics*. Ed. by Thomas Piecha and Peter Schroeder-Heister. Cham: Springer International Publishing, pp. 195–210. DOI: 10.1007/978-3-319-22686-6\_13.
- Gómez-Torrente, Mario (1989). “The Problem of Logical Constants and the Semantic Tradition: From Invariantist Views to a Pragmatic Account”. In: *The Semantic Conception of Logic: Essays on Consequence, Invariance, and Meaning*. Ed. by J. Woods and G. Sagi. Oxford University Press, pp. 35–55.
- (2002). “The Problem of Logical Constants”. In: *Bulletin of Symbolic Logic* 8.1, pp. 1–37. DOI: 10.2178/bs1/1182353851.
- Gordeev, Lev (1987). “On Cut Elimination in the Presence of Peirce Rule”. In: *Archive for Mathematical Logic* 26.1, pp. 147–164.
- Gustafsson, Martin (2017). “Quine on Translation and Logical Deviance”. In: *The Monist* 100.2, pp. 228–248.
- Haack, Susan (1982). “Dummett’s Justification of Deduction”. In: *Mind* 91.362, pp. 216–239. DOI: 10.1093/mind/xci.362.216.
- Hacking, Ian (1979). “What is Logic?” In: *Journal of Philosophy* 76.6, pp. 285–319. DOI: 10.2307/2025471.
- Hjortland, Ole (2014). “Verbal Disputes in Logic: Against Minimalism for Logical Connectives”. In: *Logique Et Analyse* 57.227, pp. 463–486.
- Hodes, Harold (2004). “On The Sense and Reference of A Logical Constant”. In: *The Philosophical Quarterly* 54.214, pp. 134–165. DOI: <https://doi.org/10.1111/j.0031-8094.2004.00345.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0031-8094.2004.00345.x>.
- Humberstone, Lloyd and David Makinson (2011). “Intuitionistic Logic and Elementary Rules”. In: *Mind* 120, pp. 1035–1051.

- Incurvati, L. and J. J. Schlöder (2017). “Weak Rejection”. In: *Australasian Journal of Philosophy* 95.4, pp. 741–760. DOI: 10.1080/00048402.2016.1277771.
- (2019). “Weak Assertion”. In: *Philosophical Quarterly* 69.277, pp. 741–770. DOI: 10.1093/pq/pqz016.
- (2023). *Reasoning with Attitude: Foundations and Applications of Inferential Expressivism*. Oxford University Press. ISBN: 9780197621004. URL: <https://books.google.es/books?id=QJBo0AEACAAJ>.
- Incurvati, L. and P. Smith (2010). “Rejection and Valuations”. In: *Analysis* 70.1, pp. 3–10.
- Janssen, Theo M.V. and Barbara H. Partee (2011). “Compositionality”. In: *Handbook of Logic and Language*. Ed. by Johan van Benthem and Alice ter Meulen. Amsterdam: North-Holland, pp. 417–473. ISBN: 978-0-444-81714-3. DOI: <https://doi.org/10.1016/B978-044481714-3/50011-4>.
- Kneale, William (1956). “The Province of Logic”. In: *Mind* 66.262, p. 258. DOI: 10.1093/mind/lxvi.262.258.
- Kreisel, Georg and J. L. Krivine (1967). *Éléments de Logique Mathématique Théorie des Modèles*. Dunod.
- Kuhn, Thomas S. (1962). *The Structure of Scientific Revolutions*. Ed. by Ian Hacking. Chicago, IL: University of Chicago Press.
- Kürbis, Nils (2021). “Normalisation for Bilateral Classical Logic with Some Philosophical Remarks”. In: *Journal of Applied Logics* 2.8, pp. 531–556.
- Lewis, David K. (1975). “Languages and Language”. In: *Minnesota Studies in the Philosophy of Science*. Ed. by Keith Gunderson. University of Minnesota Press, pp. 3–35.
- Martin, Ben and Ole Hjortland (2024). “Anti-Exceptionalism about Logic (Part I): From Naturalism to Anti-Exceptionalism”. In: *Philosophy Compass* 19.8, e13014.
- Mates, B. (1972). *Elementary Logic*. Oxford University Press. ISBN: 9780195014914.
- McGee, Vann (1996). “Logical Operations”. In: *Journal of Philosophical Logic* 25.6, pp. 567–580. DOI: 10.1007/bf00265253.
- (2000). “Everything”. In: *Between Logic and Intuition: Essays in Honor of Charles Parsons*. Ed. by Gila Sher and Richard Editors Tieszen. Cambridge University Press, pp. 54–78. DOI: 10.1017/CB09780511570681.005.
- Milne, Peter (2002). “Harmony, Purity, Simplicity and a ?Seemingly Magical Fact?” In: *The Monist* 85.4, pp. 498–534. DOI: 10.5840/monist200285427.
- Monk, J. Donald (1976). *Mathematical Logic*. New York: Springer Verlag.
- Morton, Adam (1973). “Denying the Doctrine and Changing the Subject”. In: *Journal of Philosophy* 70.15, pp. 503–510.
- Murzi, Julien and Brett Topey (2021). “Categoricity by Convention”. In: *Philosophical Studies* 178.10, pp. 3391–3420.
- Negri, Sara (2002). “A Normalizing System of Natural Deduction for Intuitionistic Linear Logic”. In: *Archive for Mathematical Logic* 41.8, pp. 789–810. DOI: 10.1007/s001530100136.
- Negri, Sara and Jan von Plato (2001). *Structural Proof Theory*. Ed. by Jan Von Plato. New York: Cambridge University Press.

- Oliveira, Hermógenes (2019). “On Dummett’s Pragmatist Justification Procedure”. In: *Erkenntnis* 86.2, pp. 429–455. DOI: 10.1007/s10670-019-00112-7.
- Olkhovikov, Grigory K. and Peter Schroeder-Heister (2014). “On Flattening Elimination Rules”. In: *The Review of Symbolic Logic* 7, pp. 60–72.
- Pagin, Peter and Dag Westerståhl (2010a). “Compositionality I: Definitions and Variants”. In: *Philosophy Compass* 5.3, pp. 250–264. DOI: 10.1111/j.1747-9991.2009.00228.x.
- (2010b). “Compositionality II: Arguments and Problems”. In: *Philosophy Compass* 5.3, pp. 265–282. DOI: 10.1111/j.1747-9991.2009.00229.x.
- Paoli, Francesco (2003). “Quine and Slater on Paraconsistency and Deviance”. In: *Journal of Philosophical Logic* 32.5, pp. 531–548. DOI: 10.1023/a:1025648722188.
- (2014). “Semantic Minimalism for Logical Constants”. In: *Logique et Analyse* 57.227, pp. 439–461. ISSN: 00245836, 22955836. URL: <http://www.jstor.org/stable/44085296> (visited on 09/26/2023).
- Paseau, A. C. and Owen Griffiths (2021). “Is English Consequence Compact?” In: *Thought: A Journal of Philosophy* 10.3, pp. 188–198. DOI: 10.1002/tht3.492.
- Peacocke, Christopher (2004). “Understanding Logical Constants: A Realist’s Account”. In: *Studies in the philosophy of logic and knowledge*. Ed. by T. J. Smiley and Thomas Baldwin. Published for the British Academy by Oxford University Press, p. 163.
- Peregrin, Jaroslav (2014). *Inferentialism: Why Rules Matter*. London and New York: Palgrave-Macmillan.
- Popper, K. R. (1947). “Xii.—Logic Without Assumptions”. In: *Proceedings of the Aristotelian Society* 47.1, pp. 251–292. DOI: 10.1093/aristotelian/47.1.251.
- Prawitz, Dag (1965). *Natural Deduction: A Proof-Theoretical Study*. Mineola, N.Y.: Dover Publications.
- (1973). “Towards A Foundation of A General Proof Theory”. In: *Proceedings of the Fourth International Congress for Logic, Methodology and Philosophy of Science, Bucharest, 1971*. Ed. by Patrick Suppes et al. Vol. 74. Studies in Logic and the Foundations of Mathematics. Elsevier, pp. 225–250. DOI: [https://doi.org/10.1016/S0049-237X\(09\)70361-1](https://doi.org/10.1016/S0049-237X(09)70361-1).
- (1974). “On the Idea of a General Proof Theory”. In: *Synthese* 27.1-2, pp. 63–77. DOI: 10.1007/bf00660889.
- (1987). “Some Remarks on Verificationistic Theories of Meaning”. In: *Synthese* 73.3, pp. 471–477. DOI: 10.1007/bf00484988.
- (2017). “Pragmatist and Verificationist Theories of Meaning”. In: *The Philosophy of Michael Dummett*. Ed. by Hahn L.E. Auxier R.E. Open Court, pp. 455–481.
- Price, Huw (1990). “Why ?Not??” In: *Mind* 99.394, pp. 221–238. DOI: 10.1093/mind/xcix.394.221.

- Priest, Graham (2003). “On Alternative Geometries, Arithmetics, and Logics; a Tribute to Łukasiewicz”. In: *Studia Logica* 74.3, pp. 441–468. DOI: 10.1023/a:1025123418085.
- (2014). “Revising logic”. In: *The Metaphysics of Logic*. Ed. by Penelope Rush. Cambridge University Press, pp. 211–223.
- Prior, A. N. (1960). “The Runabout Inference-Ticket”. In: *Analysis* 21.2, pp. 38–39. DOI: 10.1093/analys/21.2.38.
- Putnam, Hilary (1957). “Three-Valued Logic”. In: *Philosophical Studies* 8.5, pp. 73–80. DOI: 10.1007/bf02304905.
- (1969). “Is Logic Empirical?” In: *Boston Studies in the Philosophy of Science: Proceedings of the Boston Colloquium for the Philosophy of Science 1966/1968*. Ed. by Robert S. Cohen and Marx W. Wartofsky. Dordrecht: Springer Netherlands, pp. 216–241.
- Quine, W. V. (1954). “Carnap and Logical Truth”. In: *Synthese* 12.4, pp. 350–74. DOI: 10.1007/bf00485423.
- (1960). *Word and Object*. Cambridge, MA, USA: MIT Press.
- (1986). *Philosophy of Logic (2Nd Edition)*. Cambridge, MA: Harvard University Press.
- Read, Stephen (2000). “Harmony and Autonomy in Classical Logic”. In: *Journal of Philosophical Logic* 29.2, pp. 123–154. DOI: 10.1023/a:1004787622057.
- (2010). “General-Elimination Harmony and the Meaning of the Logical Constants”. In: *Journal of Philosophical Logic* 39.5, pp. 557–576. DOI: 10.1007/s10992-010-9133-7.
- (2015a). “General-Elimination Harmony and Higher-Level Rules”. In: *Dag Prawitz on Proofs and Meaning*. Ed. by Heinrich Wansing. Springer International Publishing, pp. 293–312. DOI: 10.1007/978-3-319-11041-7\_13.
- (2015b). “Proof-Theoretic Validity”. In: *Foundations of Logical Consequence*. Ed. by Colin R. Caret and Ole T. Hjortland. Oxford University Press, pp. 136–158.
- Restall, Greg (2002). “Carnap’s Tolerance, Meaning, and Logical Pluralism”. In: *Journal of Philosophy* 99.8, pp. 426–443.
- (2005). “Multiple Conclusions”. In: *Logic, Methodology, and Philosophy of Science*. Ed. by Petr Hájek, Luis Valdés-Villanueva, and Dag Westerståhl. College Publications.
- (2014). “Pluralism and Proofs”. In: *Erkenntnis* 79.S2, pp. 279–291. DOI: 10.1007/s10670-013-9477-9.
- Ripley, Ellie (2011). “Negation, denial, and rejection”. In: *Philosophy Compass* 6.9, pp. 622–629.
- (2012). “Conservatively Extending Classical Logic with Transparent Truth”. In: *Review of Symbolic Logic* 5.2, pp. 354–378. DOI: 10.1017/s1755020312000056.
- (2015). “Anything Goes”. In: *Topoi* 34.1, pp. 25–36. DOI: 10.1007/s11245-014-9261-8.
- (2017). “Bilateralism, coherence, warrant”. In: *Act-Based Conceptions of Propositional Content*. Ed. by Friederike Moltmann and Mark Textor. Oxford University Press, pp. 307–324.

- Ripley, Ellie (2020). “Denial”. In: *The Oxford Handbook of Negation*. Ed. by Viviane Déprez and M. Teresa Espinal. Oxford University Press, pp. 47–57.
- (2021). “One Step is Enough”. In: *Journal of Philosophical Logic* 51.6, pp. 1–27. DOI: 10.1007/s10992-021-09615-7.
- Rumfitt, Ian (2000). “Yes and No”. In: *Mind* 109.436, pp. 781–823. DOI: 10.1093/mind/109.436.781.
- (2002). “Unilateralism Disarmed: A Reply to Dummett and Gibbard”. In: *Mind* 111.442, pp. 305–322. DOI: 10.1093/mind/111.442.305.
- (2008). “Co-Ordination Principles: A Reply”. In: *Mind* 117.468, pp. 1059–1063. DOI: 10.1093/mind/fzn032.
- (2017). “Against Harmony”. In: *Blackwell Companion to the Philosophy of Language*. Ed. by B. Hale and Crispin Wright. Blackwell.
- Schroeder-Heister, Peter (1984). “A Natural Extension of Natural Deduction”. In: *The Journal of Symbolic Logic* 49.4, pp. 1284–1300. ISSN: 00224812. URL: <http://www.jstor.org/stable/2274279> (visited on 11/25/2023).
- (2023). “Proof-Theoretic Semantics”. In: *Stanford Encyclopedia of Philosophy*.
- Sellars, Wilfrid (1953). “Inference and Meaning”. In: *Mind* 62.247, pp. 313–338. DOI: 10.1093/mind/lxii.247.313.
- (1954). “Some Reflections on Language Games”. In: *Philosophy of Science* 21.3, pp. 204–228. DOI: 10.1086/287344.
- Shapiro, Stewart (2015). “The Meaning of Logical Terms”. In: *Foundations of Logical Consequence*. Ed. by Colin R. Caret and Ole T. Hjortland. Oxford University Press.
- (2019). “Translating Logical Terms”. In: *Topoi* 38.2, pp. 291–303. DOI: 10.1007/s11245-016-9432-x.
- Shoesmith, D. J. and Timothy Smiley (1978). *Multiple Conclusion Logic*. Cambridge, England / New York London Melbourne: Cambridge University Press.
- Simonelli, Ryan (2023). “How to Be a Hyper-Inferentialist”. In: *Synthese* 202.163, pp. 1–24. DOI: 10.1007/s11229-023-04382-1.
- Smiley, Timothy (1996). “Rejection”. In: *Analysis* 56.1, pp. 1–9. DOI: 10.1111/j.0003-2638.1996.00001.x.
- Stafford, Will and Victor Nascimento (2023). “Following All the Rules: Intuitionistic Completeness for Generalized Proof-Theoretic Validity”. In: *Analysis* 83.3, pp. 507–516. DOI: 10.1093/analys/anac100.
- Stalnaker, Robert (1997). “Reference and Necessity”. In: *A Companion to the Philosophy of Language*. Ed. by Bob Hale, Crispin Wright, and Alexander Miller. Wiley-Blackwell, pp. 902–919.
- Steinberger, Florian (2009). “Not so Stable”. In: *Analysis* 69.4, pp. 655–661. DOI: 10.1093/analys/anp100.
- (2011a). “What Harmony Could and Could Not Be”. In: *Australasian Journal of Philosophy* 89.4, pp. 617–639. DOI: 10.1080/00048402.2010.528781.
- (2011b). “Why Conclusions Should Remain Single”. In: *Journal of Philosophical Logic* 40.3, pp. 333–355. DOI: 10.1007/s10992-010-9153-3.

- Steinberger, Florian (2013). “On the Equivalence Conjecture for Proof-Theoretic Harmony”. In: *Notre Dame Journal of Formal Logic* 54.1, pp. 79–86. DOI: 10.1215/00294527-1731398.
- Steinberger, Florian and Julien Murzi (2017). “Inferentialism”. In: *Blackwell Companion to Philosophy of Language*. Ed. by Bob Hale and Crispin Wright, pp. 197–224.
- Szabó, Zoltán Gendler (Feb. 2012). “The case for compositionality”. In: *The Oxford Handbook of Compositionality*. Oxford University Press. ISBN: 9780199541072. DOI: 10.1093/oxfordhb/9780199541072.013.0003.
- Tahko, Tuomas E. (2021). “A Survey of Logical Realism”. In: *Synthese* 198.5, pp. 4775–4790. DOI: 10.1007/s11229-019-02369-5.
- Tennant, Neil (1978). *Natural Logic*. Edinburgh University Press.
- (1987). *Anti-Realism and Logic: Truth as Eternal*. New York: Oxford University Press.
- (1997). *The Taming of the True*. New York: Oxford University Press.
- (1999). “Negation, Absurdity and Contrariety”. In: *What is Negation?* Ed. by Dov M. Gabbay and Heinrich Wansing. Dordrecht: Springer Netherlands, pp. 199–222. ISBN: 978-94-015-9309-0. DOI: 10.1007/978-94-015-9309-0\_10.
- (May 2010). “Harmony in a sequent setting”. In: *Analysis* 70.3, pp. 462–468. ISSN: 0003-2638. DOI: 10.1093/analys/anq026.
- (2020). “Inferentialism, Logicism, Harmony, and a Counterpoint”. In: *Logic, Language, and Mathematics: Themes from the Philosophy of Crispin Wright*. Ed. by Alexander Miller. Inferentialism, Logicism, Harmony, and a Counterpoint. Oxford University Press, pp. 223–247.
- (2021). “What is a rule of Inference?” In: *The Review of Symbolic Logic* 14.2, pp. 307–346. DOI: 10.1017/S1755020320000441.
- Tranchini, Luca (2018). “Stabilizing Quantum Disjunction”. In: *Journal of Philosophical Logic* 47.6, pp. 1029–1047. DOI: 10.1007/s10992-018-9460-7.
- Urbas, Igor (1996). “Dual-Intuitionistic Logic”. In: *Notre Dame Journal of Formal Logic* 37.3, pp. 440–451. DOI: 10.1305/ndjfl/1039886520.
- Warmbröd, K. (1999). “Logical Constants”. In: *Mind* 108.431, pp. 503–538. DOI: 10.1093/mind/108.431.503.
- Warren, Jared (2015). “Talking with Tonkers”. In: *Philosophers’ Imprint* 15.
- (2020). *Shadows of Syntax: Revitalizing Logical and Mathematical Conventionalism*. New York, USA: Oxford University Press.
- Wehmeier, Kai F. (2018). “The Proper Treatment of Variables in Predicate Logic”. In: *Linguistics and Philosophy* 41.2, pp. 209–249.
- Weiner, Matthew (2005). “Must We Know What We Say?” In: *Philosophical Review* 114.2, pp. 227–251. DOI: 10.1215/00318108-114-2-227.
- Woods, Jack (2012). “Failures of Categoricity and Compositionality for Intuitionistic Disjunction”. In: *Thought: A Journal of Philosophy* 1.4, pp. 281–291.
- Zimmermann, Ernst (2002). “Peirce’s Rule in Natural Deduction”. In: *Theoretical Computer Science* 275.1, pp. 561–574. ISSN: 0304-3975.

Zimmermann, Thomas Ede and Wolfgang Sternefeld (2013). *Introduction to Semantics. An Essential Guide to the Composition of Meaning*. Berlin, Boston: De Gruyter Mouton.