UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Explanatory Constraints on Logic and Truth

Permalink

https://escholarship.org/uc/item/2740v1kj

ISBN

9798173788931

Author

Villalon, Daniel

Publication Date

2026-07-10

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Explanatory Constraints on Logic and Truth

By

Daniel Villalon

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Philosophy

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor John MacFarlane, Co-Chair
Professor Paolo Mancosu, Co-Chair
Associate Professor Shamik Dasgupta
Professor Amy Rose Deal

Summer 2026

Explanatory Constraints on Logic and Truth

Copyright 2026
by
Daniel Villalon

Abstract

Explanatory Constraints on Logic and Truth

By

Daniel Villalon

Doctor of Philosophy in Philosophy

University of California, Berkeley

Professors John MacFarlane and Paolo Mancosu, Co-Chairs

According to a popular view in the philosophy of logic called "inferentialism," the meanings of the logical constants are given by their rules of inference. Put another way, understanding a logical constant consists in being disposed to reason in accordance with certain rules. Variations on this view have been endorsed by prominent figures in the history of analytic philosophy, including Dummett and the early Carnap.

However, inferentialism faces a well-known challenge originally raised by Arthur Prior. Specifically, it does not seem as though every collection of rules can give the meaning of a logical constant, lest we be able to introduce a constant 'tonk' into our language that lets us infer any claim from any other. This challenge has prompted a longstanding debate on the constraints that inferentialists should impose upon meaning-giving rules. Discussions of these constraints have revolved almost entirely, though, around proof-theoretic concerns.

In this dissertation, I approach Prior's challenge from a new angle that emphasizes explanatory considerations. When speaking of "explanations", my focus is on metaphysical explanations rather than explanations of other kinds. To motivate my approach, I therefore defend a conception of metaphysical explanation that I subsequently apply to Prior's challenge.

In tackling Prior's challenge, I focus on the explanatory burdens that inferentialists should meet. I then argue that the best way of meeting these

burdens motivates a particular constraint on meaning-giving rules. This constraint rules out ‘tonk’ and therefore addresses Prior’s challenge in a non-arbitrary way.

Finally, I examine the proposed proof-theoretic constraint of “logical harmony” and the arguments that have been adduced in its favor. While I reject these arguments, I argue that the notion of logical harmony nevertheless plays an important role. I then critique existing accounts of the notion and develop my own.

Table of Contents

Acknowledgments ii

Introduction iii

Chapter 1: The Epistemic Role of the Logic of Ground 1
- Section I – *The Grounding Framework* 2
- Section II – *Two Accounts of Metaphysical Explanation* 4
- Section III – *Type B Examples* 11
- Section IV – *Inference to the Best Explanation* 16
- Conclusion 22

Chapter 2: Meta-Semantic Explanation and 'Tonk' 23
- Section I – *The Truth-Table Challenge* 26
- Section II – *Sense and Reference* 32
- Section III – *Solving the Truth-Table Challenge* 35
- Section IV – *Alternative Solutions* 39
- Section V – *Tonk* 45
- Conclusion 46

Chapter 3: An Analysis of Logical Harmony 47
- Section I – *The Harmony Requirement* 48
- Section II – *The Purpose of Harmony* 56
- Section III – *Accounts of Harmony* 59
- Section IV – *Dummett's Account of Validity* 66
- Section V – *Harmony Revisited* 75
- Conclusion 79

Concluding Remarks 80

References 82

Acknowledgments

I am extremely grateful to have had remarkable advisors at UC Berkeley. These include my co-chairs John and Paolo as well as Shamik.

Paolo introduced me to the topic of proof theory in his introductory course on the subject – a course that is seldom taught elsewhere. He also taught a seminar on proof-theoretic semantics in part to help me lay the foundations for this dissertation. I will forever be grateful for the dedication he has shown me. His breadth of technical, philosophical and historical knowledge is an inspiration as well as a reminder of how much I have yet to learn.

John has played an invaluable role in modeling many of the philosophical virtues to which I aspire. These include crystal clarity, argumentative rigor and dialectical thoroughness. I can only hope that his mentorship has influenced my philosophical style and the intellectual standards that I apply to my work. His course on philosophical logic first exposed me to some of the questions that take center stage in this dissertation.

Shamik has pushed the boundaries of my thinking through his blend of rigor and creativity. While philosophy may sometimes seem overly focused on technical details, his guidance has always helped me keep sight of the bigger picture. His approach to philosophy also embodies a rare methodological self-awareness that I have sought to apply to my work.

I would also like to thank every member of the Berkeley graduate student community with whom I have had helpful discussions. My experience would not have been the same without those interactions.

Finally, and above all, I would like to thank my family. Their support has been unwavering throughout this journey and without them I would never have made it this far.

Introduction

The twentieth century gave rise to multiple revolutions in the philosophy of language and logic. Among these, few were more significant than the application of modal reasoning to the study of reference and the semantics of names. Over the past couple decades, metaphysics has increasingly shifted its attention from the study of modal notions to hyper-intensional notions such as "ground". Such notions are widely thought to play a vital role in the task of devising metaphysical explanations as well as describing the dependency of certain facts on others. However, while explanatory notions such as ground have received abundant attention in recent metaphysics, there has been little attempt to apply these developments to issues in the philosophy of logic or language.

This dissertation approaches questions in the philosophy of logic from a vantage point that emphasizes the role of metaphysical explanation. This approach is uncommon given that questions in the philosophy of logic are typically approached through a purely formalistic lens. This lens emphasizes mathematical results and semantical considerations. However, it gives little attention to explanatory issues and the constraints that such issues impose on our answers to the relevant questions.

In applying this explanation-oriented methodology, I will focus primarily on questions that arise within the inferentialist program. Inferentialism consists, roughly speaking, of the view that the meanings of the logical constants are given by their rules of inference. My interest is in the constraints that inferentialists should (and should not) impose on the rules that may give the meaning of a logical constant. While a wide variety of constraints have been proposed, these constraints are typically evaluated solely on the basis of their proof-theoretic merits. By contrast, I will seek to determine which of these constraints can be motivated on an explanatory basis. I will also explore the role that the logic of ground plays in metaphysical explanation itself.

In chapter 1, I explore the epistemology of ground and its intersection with issues in the logic of ground. I begin by defending an account of metaphysical explanation on which some such explanations consist of a derivation of the explanandum from premises that invoke the notion of ground. Importantly, these derivations may employ rules that belong to the logic of ground in addition to the standard rules of

first order logic. With this framework in place, I argue that our knowledge of certain grounding claims is enabled by a form of inference to the best explanation (IBE) – where this form of IBE ranges over the kinds of metaphysical explanations just described. It follows that certain rules in the logic of ground play an important role in enabling our knowledge of various grounding claims. This in turn suggests that debates over the logic of ground have epistemic ramifications.

In chapter 2, I turn to a well-known and much discussed challenge raised by Arthur Prior – namely, the problem of 'tonk'. As Prior first observed, inferentialists face the following problem: if any collection of rules may give the meaning of a logical constant, then we can add a constant 'tonk' to our language the meaning of which is given by rules that let us infer any claim from any other. This observation has prompted a wide range of replies on the part of inferentialists. Such replies typically impose different constraints – often of a proof-theoretic sort – on the rules that may give the meaning of a logical constant. These constraints are rarely motivated, however, independently of the need to rule out 'tonk'. Consequently, it is unclear which of the proposed constraints, if any, inferentialists should adopt.

In this chapter, I confine my attention to inferentialists who endorse a realist (i.e. non-epistemic) conception of truth. I argue that such inferentialists must explain why our ordinary connectives obey their familiar truth tables – where explaining this fact amounts to providing a (metaphysical) explanation of the kind described in chapter 1. I then develop an explanation of this sort that requires both a) drawing a distinction between the sense and referent of a logical constant as well as b) adopting a non-deflationary theory of truth. The explanation also requires that we treat the meanings of our familiar connectives as given by all their classical rules. Moreover, I argue that my proposed explanation outdoes alternatives in important respects including explanatory power. Using the form of IBE defended in chapter 1, I argue that these advantages give us reason to adopt my explanation and by extension the commitments that it incurs.

Finally, I argue that these commitments motivate a particular constraint on the rules that may give the meaning of a logical constant. This constraint rules out 'tonk' and therefore provides us with an independently motivated response to Prior's original challenge.

In chapter 3, I turn my attention to the proposed proof-theoretic constraint of logical harmony. This constraint is formulated in different ways by different authors. On each formulation, though, the constraint is intended to capture the

idea that the introduction and elimination rules that give the meaning of a logical constant should exhibit a kind of balance. Moreover, on most formulations, the rules of intuitionistic logic are deemed harmonious whereas the rules of classical logic are not. This outcome is used as a basis for supporting the revision of classical logic in favor of intuitionistic logic. Such requirements therefore seem to pose an issue for the overall view defended in the previous chapter, which required that we treat the meanings of our familiar connectives as given by all their classical rules.

In this chapter, I examine various possible reasons for imposing a harmony constraint on the rules that may give a connective's meaning and argue that these reasons are unconvincing. It follows that such constraints do not provide an independently motivated response to Prior's challenge. Moreover, given that inferentialists have no good reason to impose a harmony constraint, it follows that they have no clear reason to replace classical with intuitionistic logic.

Despite these conclusions, however, I argue that harmony is not an idle notion. This is because it plays an important role in helping us answer the following question: namely, whether or not the meanings of our ordinary connectives are partly given by their elimination rules. Specifically, I argue that if we can devise a notion of harmony that lets us extract the elimination rules that govern our ordinary connectives from their introduction rules, this gives us reason to believe that the elimination rules do not contribute to giving the meanings of our ordinary connectives. To answer the question at hand, I first assess existing formulations of the notion of harmony and find them all lacking. I then devise a new formulation that builds on Dummett's proof-theoretic account of logical validity. This formulation avoids the shortcomings that other accounts face and lets us extract the elimination rules that govern our connectives from their introduction rules. This outcome supports the contention that the elimination rules do not contribute to giving the meanings of our ordinary connectives. The notion of logical harmony therefore plays a diagnostic rather than a revisionary role.

Chapter 1: The Epistemic Role of the Logic of Ground

The logic of ground is a much-studied topic, and the notion of ground is often said to obey various deductive rules. The epistemology of ground, on the other hand, has received comparatively little attention. In this chapter, I will aim to show that questions concerning the logic and epistemology of ground are more intertwined than has been recognized so far.

Specifically, I will argue that certain rules in the logic of ground play an overlooked epistemic role. In making my case, I will assume that knowledge of *non*-transparent grounding claims is possible – where a transparent grounding claim is one that is immediately obvious and knowable with little to no effort (e.g. "This is red because it is scarlet") whereas a non-transparent grounding claim is one that is not (e.g. "I am in pain because my C-fibers are firing").[1] If the reader does not share this epistemic assumption, they may read the chapter as an exploration of the consequences that come from accepting it. However, given that metaphysicians routinely endorse non-transparent grounding claims (e.g. grounding physicalism, ethical naturalism, priority monism), this assumption is far from gratuitous.[2]

My argument will focus on the structure of metaphysical explanations as well as the epistemology of non-transparent grounding claims. Roughly speaking, the chapter's overall argument will consist of the following:

1) Certain metaphysical explanations consist of derivations that employ grounding rules.
2) The best epistemology for non-transparent grounding claims is one on which we apply inference to the best explanation (henceforth "IBE") to explanations of the kind described in premise 1.
3) Therefore, grounding rules play an important epistemic role insofar as they help enable the form of IBE described in premise 2.

In section I, I will introduce the grounding framework. In section II, I will argue that some metaphysical explanations have a derivational structure that employs rules from the logic of ground. I will support this claim by showing that it is continuous with precedents from metaphysical practice – specifically, from certain debates in

[1] The distinction between transparent and non-transparent grounding claims should therefore be understood as an epistemological distinction. Transparent grounding claims are sometimes called "scrutable".

[2] See Schaffer (2010) for a defense of priority monism i.e. the view that the existence of each entity is grounded in the existence of the entire cosmos.

the metaphysics of mind and perception. I will also consider analogies between metaphysical explanation and explanation in other domains (e.g. science and mathematics). In section III, I will consider some intuitive case judgments that further support my account of metaphysical explanation. In section IV, I will argue that the best epistemology for non-transparent grounding claims is one on which we apply IBE to metaphysical explanations that have a derivational structure. Given that these derivations often employ grounding rules, it follows that grounding rules play an overlooked epistemic role.

Section I – *The Grounding Framework.*

The notion of ground is often thought to have an ancient pedigree that dates back at least as far as Aristotle's notion of real definition (see Schaffer (2009) for discussion). It is used to regiment claims that describe the way in which certain facts depend on others (see Bliss and Trogdon (2021) for a useful overview). These claims are expressed in natural language via locutions such as "because" (e.g. "there are bachelors because there are unmarried men") and "in virtue of" (e.g. "This chair is red in virtue of the fact that it is scarlet"). There are two ways of regimenting such claims, though, and by extension two competing approaches to the notion of ground. The first approach (e.g. as seen in Correia (2010)) employs a hyper-intensional operator "$\mathcal{G}$" that takes in two sentences and forms a new sentence. For example, the claim that O is red because O is scarlet may be formalized as "$\mathcal{G}$(O is red, O is scarlet)". The second approach, meanwhile, employs a relational predicate that relates worldly facts and entities. While I will adopt the second approach for the sake of convenience, nothing that I argue in this chapter will turn on this choice. Going forward, I will use the symbol "<" as the two-place grounding predicate. I will use the square brackets "[ ]" as a term forming operator that takes in a sentence (or a collection of sentences) and forms a term that stands for a worldly fact (or a collection of facts). The notation "[P]", for instance, is therefore a term that designates the worldly fact that corresponds to the statement P while the term "[P_1 … P_n]" denotes the collection of facts [P_1] … [P_n].[3]

[3] One might worry that "[p]" will lack a denotation if "p" is false, given that false sentences do not seem to single out facts. To handle this issue, I will treat "[p]" as denoting a *non*-obtaining fact when "p" is false and an obtaining fact when "p" is true. The grounding relation then relates obtaining facts only. If positing non-obtaining facts strikes the reader as ontologically costly, it is worth remembering that my arguments may all be formulated using an operator rather than a fact-based approach to ground.

The sentence "[P] < [Q]", meanwhile, is used to assert that the fact [P] grounds the fact [Q]. For example, if P is the statement "O is scarlet" and Q is the statement "O is red", then "[P] < [Q]" is used to assert that the fact that O is scarlet grounds the fact that O is red. Importantly, the grounding predicate "<" may take a term that denotes a collection of facts on the left-hand side. A sentence such as "[P_1 … P_n] < [R]" is therefore used to assert that the facts [P_1] … [P_n] collectively ground the fact [R]. Lastly, I will assume that we may intelligibly quantify into the operator "[ ]". Sentences such as "$\forall$x([Fx] < [Gx])" are therefore well formed and intelligible even though "[Fx]" is not a denoting term and does not, strictly speaking, denote a fact.[4] Such quantification will be needed later when formally regimenting general claims such as "whenever someone is in pain it is because their C-fibers are firing".

Going forward, I will follow Fine (2012a), Schaffer (2009) and Rosen (2010) in assuming that a) the notion of ground is ideologically primitive and b) is among the central vehicles of metaphysical explanation. These assumptions put me at odds with skeptics like Wilson (2014) who think that we have no distinctive notion of ground or, at the very least, think that no such notion is needed. They also put me at odds with those who think that the notion of ground can be defined in more basic terms.

It is common to distinguish between the notions of full and partial ground – where a fact F partly grounds a fact F' whenever the combination of F and other facts fully grounds F'. My focus going forward, though, will be entirely on full ground and I will simply call this notion "ground". This notion is thought to obey various deductive rules. Correia (2010), for example, develops one such formal system (see Fine (2012b) for an alternative system). While I will not provide the full system here, it includes the following schematic principles – where the meta-variables 'p' and 'q' range over sentences while the meta-variables 'Γ' and 'Δ' range over collections of sentences:[5]

---

[4] The situation here is analogous to that of quantifying into a function symbol in a sentence such as "$\forall$x(H(x) = a)". This sentence is well formed even though the function symbol "H" is a term forming operator that only forms a denoting term when it takes in another denoting term. Like "[Fx]", "H(x)" is therefore not a denoting term. Unlike "H(x)", however, I take no stand on whether or not "[Fx]" denotes anything relative to a variable assignment. For it is not my aim to provide any kind of semantics for languages that include grounding vocabulary. Doing so is a considerable challenge that all grounding theorists face and one that lies outside the scope of this paper.

[5] Given that Correia takes an operator-based approach to ground, I have reformulated his principles in terms of the fact-based approach.

a) $([\Gamma] < [p]) \rightarrow (\widehat{\Gamma} \wedge p)$ *Factivity*

b) $\neg([\Gamma, p] < [p])$ *Irreflexivity*

c) $([\Gamma, q] < [p]) \wedge ([\Delta] < [q]) \rightarrow ([\Gamma, \Delta] < [p])$ *Cut*

One obtains an instance of factivity, for example, by replacing 'p' with a sentence, 'Γ' and 'Δ' with collections of sentences and '$\widehat{\Gamma}$' with the conjunction of all the sentences in the collection that replaces 'Γ'. The following sentence is therefore an instance of factivity – where 'p' is replaced by the sentence 'O is a bachelor' and 'Γ' is replaced by the collection that includes the sentences 'O is a man' and 'O is unmarried':

d) ([O is a man, O is unmarried] < [O is a bachelor]) $\rightarrow$ ((O is a man $\wedge$ O is unmarried) $\wedge$ O is a bachelor).

Generally speaking, factivity requires that the sentences on the left and right hand sides of a true grounding claim all be true as well. Importantly, I do not mean to endorse Correia's system. Rather, I simply highlight it in order to illustrate paradigmatic rules in the logic of ground. I do not wish to endorse Correia's system here (or any other) for the simple reason that the logic of ground is a much-debated topic and it is not my aim to contribute to these debates.[6] Rather, my aim is to show that grounding rules play an important epistemic role.

Section II – *Two Accounts of Metaphysical Explanation.*

I have so far introduced the notion of ground and discussed some of its formal features. However, I have not discussed its precise relationship to metaphysical explanation. That said, this claim requires some clarification. For in stating earlier that "[P] < [Q]" is synonymous with "Q because P", one might think that I have thereby committed myself to the identification of grounding facts with explanatory facts. While there may be a worldly sense of "explanatory fact" on which this is true, my concern when speaking of "explanation" is with the human activity of giving explanations. I will call this sense of explanation the "human giving" sense. Explanations in this sense are clearly representational items insofar as they either consist of or are composed of statements rather than worldly entities.[7] Moreover, they are often thought of as answers to "why" questions. In seeking to devise an

[6] See Fine (2010) for one such contribution.

[7] For example, while causal facts may be explanatory in a worldly sense, we give causal explanations of events by describing their causes. Such descriptions are representational rather than worldly items.

account of metaphysical explanation, I am therefore concerned with the following question: namely, what sorts of representational items serve as metaphysical explanations in the human giving sense?

Two natural answers to this question suggest themselves if we adopt a grounding based approach to metaphysical explanation. These answers take the form of two general accounts of metaphysical explanation that I will call the "type A" and "type B" accounts. Importantly, I will defend the latter. On the type A account, a metaphysical explanation of a claim φ always consists of a grounding claim of the form "$[\psi_1 \ldots \psi_n] < [\varphi]$".[8] On the type B account, meanwhile, another kind of explanation is possible. Specifically, a metaphysical explanation of φ may also consist of a derivation of φ from suitable premises via standard logical resources and grounding rules. While I will not specify the precise range of premises that are suitable, I will assume that grounding claims qualify.[9] That said, other kinds of claims may be eligible too. Importantly, the type B account, unlike its type A rival, is a plural one – for it does not deny that bare grounding claims may sometimes serve as metaphysical explanations as well. If the reader wishes to restrict metaphysical explanations to those that have a derivational structure, though, nothing that I argue going forward will turn on this choice.

While the type A/B distinction is my own, I believe that this taxonomy encompasses the most natural accounts of metaphysical explanation that are available if we assume (as I have done) that the notion of ground is among the central explanatory notions of metaphysics.[10] Importantly, the taxonomy says nothing about explanatory virtues and is concerned solely with the *form* of metaphysical explanations. It is therefore capable of accommodating a wide range

[8] As we will later see, this kind of account appeals to those who wish to model metaphysical explanation entirely on the phenomenon of causal explanation.

[9] The type B account bears a loose resemblance to deductive-nomological accounts of explanation in the natural sciences (e.g. as developed by Hempel and Oppenheim (1948)). However, unlike such accounts, the type B account does not require that the premises of a metaphysically explanatory derivation must be laws in any significant sense. It is also worth distinguishing the type B account of explanation from deductive-nomological accounts of ground itself (e.g. Wilsch (2015)). For such accounts do not address the structure of explanations in the sense that concerns me. Nor do they treat the notion of ground as primitive (as I do). Rather, they seek to characterize the notion of ground in terms of the notion of metaphysical laws – where such laws are treated as something above and beyond mere regularities.

[10] If we jettison this assumption, then my taxonomy will clearly require modification.

of more specific views.[11] Suppose, however, that one thinks a metaphysical explanation may sometimes consist of something other than a grounding claim or a derivation. It is enough for the purposes of my overall argument that at least *some* metaphysical explanations consist of derivations that employ grounding rules. This view is compatible with the position that some explanations are neither grounding claims nor derivations. By extension, it is compatible with the position that neither the type A nor the type B view is fully correct. Nevertheless, I will restrict my attention to the type A and B views.

I will refer to explanations that consist of bare grounding claims as "type A" explanations and to derivational explanations as "type B" explanations. I will refer to type B explanations that employ grounding rules as "grounding-based type B" explanations. Strictly speaking, the type B view requires not just that there are type B explanations but that there are grounding-based type B explanations as well. For type B explanations are simply explanations with a derivational structure. The type B view, however, requires that some of these explanations employ grounding rules. My chief focus, going forward, will be on grounding-based type B explanations. This is because my aim is to assess the epistemic significance of debates over the logic of ground. To illustrate these kinds of explanations, let us suppose that we wish to explain why everything scarlet is red. To do this, we may start with the premise that for any x, if x is scarlet then [x is scarlet] < [x is red]. We may then derive the explanandum from this premise and the resulting derivation appears to be a grounding-based type B explanation. To carry out the derivation, we may reason as follows: suppose y is scarlet. Then it follows from the original premise that [y is scarlet] < [y is red]. From this we may conclude that y is red via the factivity rule and so we have therefore established that everything scarlet is red. I will return to this example in section III.

Why think that there are any grounding-based type B explanations? I will defend the view that there are such explanations by appealing to continuities with existing metaphysical practice. To make my case, I will focus on certain

[11] My distinction between the type A and B theories is largely independent of the debate between so-called "unionists" and "separatists" (I will return to this debate later on). The latter debate concerns the metaphysical issue of whether grounding facts are identical to or distinct from explanatory facts (see Raven (2015, section 5) for a brief overview of this debate). The distinction between the type A and B theories, on the other hand, concerns the structure and content of metaphysical explanations in the human giving sense described earlier on. This is an issue that has received little attention compared to the unionism/separatism debate.

argumentative strategies that have been employed in the metaphysics of mind and perception. Importantly, my aim is not to settle any issues in these areas. Rather, my aim is to show that metaphysical explanations sometimes consist of derivations that employ inference rules which are both metaphysically loaded and cannot be formalized in first order logic. This observation provides precedent for the claim that metaphysical explanations may employ grounding rules and that there are therefore grounding-based type B explanations.

I will begin by considering the explanatory argument for reductive physicalism defended by Bates (2009) and McLaughlin (2010). Reductive physicalism consists of the view that every mental state type is identical to some physical state type. Bates and McLaughlin defend this view on the basis that it best explains observed psycho-physical correlations. The argument takes the form of two stages. The first stage is a sequence of inferences to the best explanation, each of which is an instance of the following argumentative schema:[12]

1) $FACT_i$: Mental state-type M_i and physical state-type P_i are such that for all x, x is in M_i iff x is in P_i (i.e. $\forall x(M_i x \leftrightarrow P_i x)$).
2) One possible explanation of $FACT_i$ is H_i: M_i and P_i are one and the same state-type (i.e. $M_i = P_i$).
3) No alternative explanation of $FACT_i$ is as good as H_i.

$\therefore$ H_i is true (i.e. $M_i = P_i$).

Each instance of this template establishes a particular psycho-physical identity. The second stage of the argument for reductive physicalism is an inductive generalization from the identities established in the first stage. This generalization yields the conclusion that for every mental-state type, there is a physical-state type to which it is identical (i.e. reductive physicalism).[13]

The argument template used in part 1 raises the following question: namely, in what manner does the identity claim "$M_i = P_i$" explain the general correlation "$\forall x(M_i x \leftrightarrow P_i x)$"? Bates is explicit on this point: namely, it does so insofar as the identity claim lets us derive the correlation. Moreover, while Bates does not discuss the issue, it is obvious that the argument, as he formulates it, employs resources that exceed those of first order logic. This is because the premise "$M_i = P_i$" asserts an identity between types and moving from such claims to generalizations of the

[12] My formulation of the schema is taken directly from Bates (2009).

[13] Most of Bates' discussion is devoted to defending the explanatory argument from criticisms of it by Kim (2005). However, these criticisms are tangential to the focus of my paper and so I will not discuss them here.

form "$\forall x(M_i x \leftrightarrow P_i x)$" requires principles that govern type identity.[14] The explanatory argument for reductive physicalism therefore provides us with precedent for the view that metaphysical explanations may consist of derivations that employ inference rules which exceed those of first order logic and govern metaphysical notions. Moreover, if premises and rules concerning type identity are suitable for derivational explanations in metaphysics, then it is reasonable to conclude that premises and rules that concern the notion of ground may participate in such explanations as well. For I have assumed all along that ground is one of the central explanatory notions in metaphysics.[15] The foregoing example therefore provides precedent for the claim that there are grounding-based type B explanations.

I will now turn to the metaphysics of perceptual experience and the explanatory argument for phenomenal internalism defended by Pautz (2013) and critiqued by Epstein and Sethi (2025).[16] Phenomenal internalism (henceforth "internalism") consists of the view that the phenomenal character of a sensory experience depends wholly on the subject's neurological state rather than on the properties of the subject's external environment. Put otherwise, the character of a sensory experience is "internally dependent". Phenomenal externalism (henceforth "externalism"), by contrast, consists of the opposing view i.e. that the phenomenal character of a sensory experience depends, at least partly, on the properties of the subject's external environment.[17] To defend internalism, Pautz first highlights neurological findings which show that the structure of our experience of color (and secondary qualities in general) is a) well correlated with the structure of our neural properties but b) poorly correlated with the structure of the physical properties of our environment. For example, the degree of difference in the character of our various color experiences correlates well with the degree of difference in our neural states but correlates poorly with the degree of difference in the physical properties of the objects we perceive (e.g. spectral

[14] I take no stand though on whether or not such principles count as "logical".

[15] See Dasgupta (2014) for a defense of the claim that various metaphysical theses such as physicalism are best formulated in terms of the notion of ground.

[16] Pautz' argument is laid out more explicitly in Epstein and Sethi's critique and so I will focus primarily on their formulation of it.

[17] For example, on the internalist view, the phenomenal character of a visual sensation of red depends entirely on one's brain state. On the externalist view, meanwhile, the phenomenal character of such an experience depends partly on the fact that one is suitably related to a (presumably red) object.

reflectance profiles). The best explanation for these correlations, Pautz then argues, is that the character of our color experience is internally dependent. Lastly, Pautz appeals to an assumption that he calls the "uniformity constraint" – namely, that if the character of our color experience is internally dependent then so is the character of our shape experience. It follows that the character of our visual experience as a whole is internally dependent (i.e. phenomenal internalism).

Epstein and Sethi concede that phenomenal internalism provides the best explanation for the correlations that Pautz highlights. Moreover, they accept the uniformity constraint on a priori grounds. However, they claim that the argumentative strategy that Pautz employs leads to a stalemate between internalists and externalists. To make their case, they highlight empirical findings which show that the structure of our experience of shape is a) well correlated with the structure of the physical properties of our environment but b) poorly correlated with our neural properties. If we take these correlations as our initial explanandum rather than the correlations that Pautz highlights, we may then use his argumentative strategy to defend the externalist view instead.

The most natural interpretation of Pautz' argument is one on which a derivational account of explanation is implicit and such derivations may appeal to rules and principles that govern the notion of dependence. For suppose we claim that something like the type A account of explanation is implicit instead. The claim that internalism explains the correlations well then amounts to stating that the following grounding claim is clearly compelling: [Internalism] < [Correlations]. However, this claim asserts that a dependence fact grounds an empirical fact. It is not obvious that this is possible and it is therefore far from obvious why we should consider this to be a good type A explanation (let alone the best). Both Pautz and his critics (i.e. Epstein and Sethi), though, see it as obvious that internalism provides a good (if not the best) explanation for the correlations even if they do not accept Pautz' internalist conclusion. This mutual acceptance is puzzling on the type A account. A derivational view of metaphysical explanation therefore makes better sense of the reasoning involved in this debate.[18] Moreover, any derivation of the correlations from the internalist thesis will clearly employ rules or principles that govern the notion of dependence. If explanatory derivations may appeal to rules or principles concerning "dependence" then it is once again natural to extrapolate

[18] Even if one takes the view that dependence facts can ground empirical facts, it is enough for my point to go through that this is not at all obvious.

from this to the notion of ground.[19] The foregoing discussion therefore provides further precedent for the claim that there are grounding-based type B explanations.

The type A theorist might acknowledge the weight of my case for the type B view but claim that their view is nevertheless preferable to the type B account for the simple reason that a uniform account of metaphysical explanation is better than a plural one. But it is not obvious that uniformity considerations weigh in favor of the type A account. For they only do so if the type A account is preferable to an account on which all metaphysical explanations have a derivational structure. While I will not claim that an account of the latter kind is superior to the type A account, it enjoys some advantages and therefore seems, at least, to be its equal. First, as we have already seen, the type B account makes better sense of our explanatory judgments and also exhibits continuity with metaphysical practice. Second, an account of metaphysical explanation that has precedent in another explanatory paradigm is preferable, all else being equal, to one that is not. Where should we look for such a precedent? We might try looking to the literature on scientific explanation. But this literature is notoriously fractured over the basic form of scientific explanations themselves. Some such as Hempel/Oppenheim (1948) and Kitcher (1981) treat them all as derivations. Others such as Salmon (1984) and Lewis (1986) treat them all as causal stories. This makes the scientific explanation literature a fragile ground on which to draw. By contrast, the literature on mathematical explanation converges on the view that most, if not all, mathematical explanations have a derivational structure even if theorists disagree over specific explanatory virtues.[20] This makes the case of mathematical explanation a more stable precedent on which to draw and gives us a tentative reason to favor a view on which all metaphysical explanations consist of derivations rather than bare grounding claims. The fact that mathematical and metaphysical explanations are both directed – either often or always – at necessary truths gives us further reason to treat mathematics as the more relevant precedent.

[19] If we assume (as is common in metaphysics nowadays) that dependence claims are best captured via the notion of ground, then no extrapolation is even needed. Rather, the case at hand provides a direct example of a grounding-based type B explanation.

[20] Importantly, I am concerned solely with explanations of pure mathematical facts rather than mathematical explanations of physical facts. See Mancosu (2008, ch. 5) for a discussion of this distinction. See Steiner (1978) and Lange (2014, 2015), meanwhile, for different accounts of the explanatory virtues that pertain to mathematical explanations.

The type A theorist might argue that while type B explanations share a derivational structure with mathematical explanations, the explanatory virtues that pertain to type A explanations resemble those that pertain to causal explanations. It is worth noting, though, that any such claim requires significant defense. Absent such defense, a clear parallel between type A and causal explanations has not been established. Moreover, there are reasons to doubt that some of the key criteria that are used to assess causal explanations transfer straightforwardly to grounding claims. The criteria in question include temporal order and plausible mechanism. Temporal order does not transfer for obvious reasons. On Salmon's (1984, chs. 1, 4, 5) mechanistic view, meanwhile, causal explanations are assessed based on their level of mechanistic detail and fit with known physical laws. These criteria do not transfer clearly either since it is not obvious what would count as a mechanism in the context of grounding. Moreover, even if something analogous could be found, it would likely diverge drastically from our understanding of mechanisms in the causal domain.

The discussion above regarding mathematical and causal explanation reinforces the claim that the type A view lacks any clear advantage over a derivation only view of metaphysical explanation. I wish to reiterate, however, that I am not arguing for a derivation only view. For the type B view is a plural account. Rather, I merely seek to show that uniformity considerations give us little reason, if any, to prefer the type A account. For if we insist on a uniform account then a derivation only view seems at least as compelling as its type A competitor.[21]

Section III – *Type B Examples*.

I have so far defended the type B view on the basis of its continuity with metaphysical practice. I will now provide some intuitive examples of grounding-

[21] See Baron and Norton (2021) for an account of metaphysical explanation that applies Kitcher's (1981) unificationist account of scientific explanation to the metaphysical domain. While I, like them, believe that metaphysical explanations often have a derivational structure, I refrain from borrowing Kitcher's account wholesale as they do. For while analogies between scientific and metaphysical explanation are important, it is also true – as in the scientific case – that determining general criteria will require attention to a range of particular cases. Second, Kitcher's account requires that we be able to cleanly separate our scientific from our non-scientific beliefs. While this may be possible, it seems harder to separate our metaphysical from our non-metaphysical beliefs. Lastly, Kitcher's account is contested even in the scientific domain.

based type B explanations. These examples will serve two purposes. First, they will provide further support for the type B account. Second, they will illustrate the role that certain grounding rules such as factivity play in devising various type B explanations.[22] This illustration will then reinforce my later claim that the logic of ground plays an important epistemic role.

In providing examples of grounding-based type B explanations, I will focus on explanations of general facts – for I believe that this is where the motivation is clearest. Suppose, for example, that we wish to explain why everything scarlet is red. Now consider the following argument:

1) $\forall$x (x is scarlet $\rightarrow$ ([x is scarlet] < [x is red])). [Premise]
2) y is scarlet. [Supposition]
3) [y is scarlet] < [y is red]. [Follows from 1 and 2]
4) y is red. [Follows from 3 via Factivity]
5) y is scarlet $\rightarrow$ y is red. [Follows from 2 – 4]
6) $\forall$x (x is scarlet $\rightarrow$ x is red). [Follows from 5]

This chain of reasoning seems to provide us with an explanation for the fact at hand.

Lest the reader worry that only toy examples are possible, below is a further example from the philosophy of mind. The reasoning provides an (albeit contested) explanation for why someone is in pain whenever they are in a state of C-fiber firing:

1) $\forall$x (x is in a state of C-fiber firing $\rightarrow$ ([x is in a state of C-fiber firing] < [x is in pain])). [Premise]
2) y is in a state of C-fiber firing. [Supposition]
3) [y is in a state of C-fiber firing] < [y is in pain]. [Follows from 1 and 2]
4) y is in pain. [Follows from 3 via Factivity]
5) y is in a state of C-fiber firing $\rightarrow$ y is in pain. [Follows from 2 – 4]
6) $\forall$x (x is in a state of C-fiber firing $\rightarrow$ x is in pain). [Follows from 5]

This derivation provides us with an explanation for its conclusion regardless of whether its premise is true or not. In making this claim, I am assuming that the premises of a metaphysically explanatory argument may sometimes be false (we may think of such arguments as "incorrect explanations"). While I will not explicitly defend this assumption here, it is worth highlighting that it has precedent in other domains. For example, it is often said that we can scientifically explain the laws of

[22] While it may be possible to formulate the derivations that I will consider using a weaker grounding rule than factivity, some grounding rule will still be needed and this observation suffices for my purposes.

planetary motion by deriving them from Newton's laws even though the latter are not, strictly speaking, correct. Second, as I will discuss more later on, mathematicians sometimes defend axioms by arguing that they let us derive and thereby explain accepted results. If the premises of an explanatory derivation in mathematics must always be true, however, then this form of inference to the best explanation (IBE) would make little sense. For we presumably could not assess that a derivation is explanatory without first knowing that its premises are true. Rather, we could at most assess that it *would* be explanatory if its premises *were* true. However, any characterization of mathematical IBE in terms of such counterfactual assessments is unsatisfying. For mathematical axioms are true or false by necessity. If a derivation has a false axiom as a premise, it is therefore trivial to claim that it would be explanatory if its premises were true. This is because counterfactuals with necessarily false antecedents are trivially true on standard accounts.[23]

Some might dismiss the examples given so far in support of the type B view. But we may strengthen the case by considering the following argumentative template T – where we obtain instances of T by replacing the variables 'φ' and 'ψ' with well-formed formulas:

1) $\forall x\,(\varphi(x) \rightarrow ([\varphi(x)] < [\psi(x)]))$. [Premise]
2) $\varphi(y)$. [Supposition]
3) $[\varphi(y)] < [\psi(y)]$. [Follows from 1 and 2]
4) $\psi(y)$. [Follows from 3 via Factivity]
5) $\varphi(y) \rightarrow \psi(y)$. [Follows from 2 – 4]
6) $\forall x\,(\varphi(x) \rightarrow \psi(x))$. [Follows from 5]

Many, if not all, instances of this template seem to provide us with an explanation for their conclusion. Therefore, if the reader does not feel the force of the particular cases that I have provided, they are invited to plug in instances of the schema that they find more compelling. This will give them a basis of their own for adopting the type B view.[24] Moreover, all instances of this template employ the factivity rule. This gives us additional reason to think that the factivity rule (or something like it)

[23] The observations made here with regard to mathematical IBE also apply to the cases of metaphysical IBE considered in the previous section. Metaphysical practice therefore provides further precedent for the claim that explanatory derivations may have false premises.

[24] Importantly, I am not claiming that all type B explanations are instances of this template. Rather, I am simply claiming that instances of this template serve as a useful source of type B examples.

plays a role in devising various type B explanations (though I do not claim that the rule is employed in all such explanations).[25]

A type A theorist might object to the case judgments on which I rely by claiming that our intuitions about the given derivations are tracking some feature other than explanatory-ness. Let us call the feature that our intuitions are tracking "F". To handle this objection, I will consider some accounts of F on which F consists of something other than explanatory-ness. In each case, I will argue that the given account falls short of the explanatory interpretation of F.

A type A theorist might argue that F consists of logical validity rather than explanatory-ness. But this claim is implausible. For this account of F appears incompatible with the observation that the following deduction, while logically valid, barely exemplifies F if it does so at all:

1) $\neg\neg\forall$x (x is scarlet $\rightarrow$ x is red).
2) $\forall$x (x is scarlet $\rightarrow$ x is red).

This observation poses no problem, however, for the explanatory interpretation of F.

Alternatively, the type A theorist might claim that F consists of the fact that a derivation is an instance of the argument schema T that I provided earlier on. There are two problems, though, with this account of F. First, and most importantly, it seems that F is preserved, to some degree at least, even when we insert needless detours into my example derivations. This would not be the case if F consisted solely in the instantiation of a particular argument schema. Second, any argument – including the examples I provided – instantiates infinitely many schemata. If we

[25] One might object to this last claim by positing both a factive and non-factive notion of ground and claiming that we can define the factive notion in terms of the non-factive one as follows: $[P] <_{\text{Factive}} [Q] \equiv [P] <_{\text{Non-Factive}} [Q] \wedge (P \wedge Q)$. If we substitute the right-hand side of this definition in place of the factive grounding claims in instances of T then we may derive the conclusions from the premises without using the factivity rule. The problem with this proposal, however, is that it trivializes the derivations entirely and this outcome is incompatible with the judgment that the derivations are explanatory. The derivations are trivialized because we may derive Q from the conjunct $(P \wedge Q)$ alone via nothing more than conjunction elimination. If there are two notions of ground, I will therefore assume that neither of them is conceptually reducible to the other and that it is the factive rather than the non-factive notion that is operative in the derivations considered so far.

are to single one out as the schema to which our intuitions pertain, then some account is owed of why that schema is more salient than others.[26]

Lastly, the type A theorist might argue that F consists of the feature that a derivation has when the conjunction of its premises provides us with a compelling type A explanation for its conclusion. Put otherwise, if φ is the conjunction of the derivation's premises and ψ is the derivation's conclusion then the derivation has F if "$[\varphi] < [\psi]$" is a compelling type A explanation for ψ. However, this view also struggles with our comparative judgments. For an argument that omits any intermediate steps seems to possess F to a lesser extent than the arguments I have provided.[27] This should not be the case, though, if F consists of the feature just described. Second, in my example derivations, it is not obvious that the fact denoted by the premise is properly suited to ground the fact denoted by the conclusion. This is because it is not obvious that grounding facts are suited to ground empirical facts.[28] For whenever a fact $[\varphi]$ grounds a fact $[\psi]$, $[\varphi]$ is generally regarded as more fundamental than $[\psi]$. It is not clear, though, whether grounding facts are themselves more or less fundamental than empirical facts.[29] This gives us reason to prefer the original interpretation of F on which F consists of the explanatoriness of the derivations themselves.

I have now examined several alternative accounts of F and found them all wanting on the basis that they are either incompatible with our comparative judgments or fail to do justice to the salience of F. Of course, proponents of the type A view may devise increasingly gerrymandered accounts of F that avoid the

[26] For example, while my particular derivations are all instances of the schema T, they are also all instances of schemata that are identical to T save for the fact that they impose further restrictions on the range of the meta-linguistic variables 'φ' and 'ψ'.

[27] The type A theorist might object that an argument that omits intermediate steps is not a genuine argument. By "argument", however, I simply mean any series of inferential steps. If one also requires, in the definition of "argument", that each step conforms to a particular proof system then this effectively builds in a particular logic of ground. As I have already mentioned, though, the logic of ground is a much-contested issue and so this critique rests on a commitment that itself requires significant defense.

[28] For example, it is not clear that the grounding fact $[\forall x$ (x is scarlet $\rightarrow$ ([x is scarlet] < [x is red]))] is properly suited to ground the fact $[\forall x$ (x is scarlet $\rightarrow$ x is red)]. This is because it is not clear whether or not the former fact is more fundamental than the latter.

[29] The issue raised here is distinct from the well-known question of what grounds grounding facts. For discussions of this question see Litland (2017) and Sider (2020).

objections I have raised so far. But the explanatory account is comparatively simple and unified and so I will treat it as the best supported option going forward.

Section IV – *Inference to the Best Explanation*

I have now defended the type B view and highlighted the importance of grounding rules in devising various type B explanations. Going forward, I will examine the epistemological implications of these results. As mentioned at the beginning of this chapter, I will assume that knowledge of non-transparent grounding claims is possible. This assumption raises the following question – namely, *how* is such knowledge possible? The most viable answer, I believe, is that we use IBE to justify general grounding claims and I will defend this answer on abductive grounds.[30]

Before defending it, though, I will first elaborate on it further. Suppose we devise a type B explanation of an antecedently known claim in which a general grounding claim (e.g. "$\forall$x (x is composed of $H_2O \rightarrow$ ([x is composed of H_2O] < [x is water]))") figures as a premise. Moreover, suppose we then assess that this derivation provides us with the best explanation for its conclusion. On the account at hand, this overall process justifies believing the general grounding claim.[31] Moreover, we can then justify believing particular grounding claims (e.g. "This is water because it is composed of H_2O") by deriving them from the general claim.

In the account that I have just sketched I have assumed that IBE ranges over type B explanations. Some might wonder why I make this assumption. After all, could IBE not range over type A explanations instead? It is therefore worth noting that we cannot credibly propose IBE as an epistemology for non-transparent grounding claims unless we say something about the kind of explanations over which it ranges. I will admit that I am partial to a plural account on which IBE may range over either type A or type B explanations (though not over both simultaneously). For it seems ad hoc to restrict IBE to just one or the other. Moreover, it seems plausible that the appropriate form of a metaphysical explanation is a context sensitive matter that is determined by pragmatic factors.[32]

---

[30] Specifically, I will defend it by comparing my IBE based epistemology to alternative epistemologies and arguing that the IBE based epistemology is superior.

[31] This account of IBE in the metaphysical realm mimics the account that Lipton (2004) develops in the scientific domain.

[32] See Van Fraassen (1980, ch. 5) for a defense of this view in the scientific domain. According to Van Fraassen, what counts as an explanation in a given context is

Therefore, while type B explanations may qualify as explanations in certain contexts, they may not do so in others (and likewise for type A explanations). If this is the case, then IBE will plausibly range over different kinds of explanations in different contexts. However, if we insist on confining (metaphysical) IBE to just one category, I believe that we have (defeasible) reasons to favor type B explanations.[33] First, IBE over type B explanations has clear precedent in the abductive arguments considered in section II (i.e. those given by Bates and Pautz). Second, if we treat IBE in the metaphysical realm as ranging over type B explanations, it then bears a structural resemblance to a well-accepted form of reasoning – namely, hypothetico-deductive reasoning. This resemblance consists in the fact that both forms of reasoning involve evaluating premises in light of the conclusions that we can derive from them. All else being equal, we should prefer an account of metaphysical IBE that exhibits a close continuity with a well-accepted form of reasoning.

Additionally, if we treat metaphysical IBE as ranging over type B explanations, it then bears a close parallel to the form of IBE that is sometimes proposed as a means of assessing mathematical axioms. For in this case too, axioms are assessed in light of the consequences that can be derived from them and the explanatory virtues that those derivations exhibit.[34] Of course, I do not wish to claim that the use of IBE in mathematics is settled or uncontroversial. However, by treating IBE as ranging over type B explanations we thereby give it a close affinity with a form of reasoning that is taken seriously elsewhere.

Some might argue, however, that if we instead treat IBE as ranging over type A explanations then this will give it a greater affinity with IBE across causal explanations in science. But this affinity does not seem as strong as the affinity between mathematical IBE and IBE over type B explanations. For as I have already noted, both mathematical IBE and IBE over type B explanations involve constructing derivations. There is no comparable similarity, meanwhile, between IBE over causal explanations and IBE over grounding claims. One might object to

---

determined by pragmatic factors such as the contrast class associated with the relevant why-question. Van Fraassen, however, is more skeptical of IBE than I am.

[33] Importantly, I will not argue that IBE ranges only over grounding-based type B explanations but, rather, over type B explanations as a whole.

[34] See Maddy (1988) for a detailed discussion of the role that such considerations played in motivating the initial adoption of the ZFC framework. See Mancosu (2001) for a discussion of the important role that such reasoning plays in Russell's (1907) account of mathematical knowledge.

this claim by appealing to Schaffer's (2016) suggestion that the method of structural equation modeling (SEM) that is sometimes used to assess causal claims may be applied to grounding claims too.[35] For if we accept Schaffer's proposal, it may seem as though there is a deep affinity between IBE over causal explanations and IBE over grounding claims. But this objection mischaracterizes causal IBE. Specifically, while SEM may systematize the outputs of our causal reasoning and thereby refine it, it does not describe the way in which we actually justify our causal beliefs (at least initially). For the legitimacy of SEM is far from self-evident but is rather supported by testing its verdicts against our causal judgments in a process of reflective equilibrium. This observation highlights the important difference between SEM and causal IBE itself. By claiming that IBE over type A explanations consists of a form of SEM, the proponent of type A IBE therefore does not achieve the intended affinity with causal IBE.

Of course, by adopting Schaffer's proposal, the proponent of type A IBE may achieve a continuity between the way in which type A IBE and causal IBE are systematized. However, a similarity in systematization is not the same as a similarity in the reasoning process itself. Moreover, in order to achieve the intended similarity with causal IBE, the proponent of type A IBE must make speculative assumptions about the way in which their preferred form of IBE is correctly systematized. By contrast, the proponent of type B IBE need make no such assumptions at all in order to achieve a close affinity between type B IBE and other forms of reasoning. This gives us a reason to prefer an account of metaphysical IBE on which it ranges over type B rather than type A explanations.[36]

---

[35] The framework of structural equation modeling is as an attempt to mathematically formalize the interventionist approach to causal explanation that is found in Woodward and Hitchcock (2003). On the interventionist approach, a causal claim of the form "X causes Y" is judged to be strong to the extent that possible interventions on the variable X change the variable Y when other background variables are held constant. The relata of the causal relation are therefore variables rather than events.

[36] I wish to reiterate that I am open to a plural account on which both kinds of IBE are possible. Additionally, I do not assume that type B IBE applies only to derivations that use grounding claims as premises. For it is possible that the premises of a metaphysical explanation may consist of other claims too. However, I assume that derivations which rely on grounding claims as premises are among the strongest explanations available in metaphysics. For if we did not make this assumption, then metaphysical IBE would have little traction in establishing grounding claims. This is because there would be an unmanageably large range of candidate explanations and explanatory virtues would do

I have so far argued that if IBE gives knowledge of grounding claims then it ranges over type B rather than type A explanations. Why think that IBE gives knowledge of grounding claims at all though? To properly motivate this view as far as non-transparent grounding claims are concerned, I must say something about its main competitor – namely, rationalism. I use the term "rationalism" to describe approaches which invoke something like a faculty of rational perception that lets us immediately apprehend grounding facts. Appeals to such a faculty are often considered suspect for the simple reason that its mechanisms appear highly mysterious. Moreover, appeals to such a faculty have no clear precedent in factual domains (I say "factual" because appeals to something resembling rational perception are more widespread in normative domains).[37] We do not, for example, appeal to rational perception as a way of explaining our knowledge of causal truths. Rather, we appeal to a form of IBE. Appeals to IBE in the epistemology of ground are therefore more continuous with our epistemological theorizing in other areas.[38]

Nevertheless, it is implausible that my IBE account provides an epistemology for transparent grounding claims. For IBE is a difficult process whereas transparent grounding claims are easily knowable without empirical investigation. An a priori epistemology is therefore needed for these and those such as Siscoe (2021) argue

---

little to narrow the field. We would consequently be left without an answer to our original question – namely, how knowledge of non-transparent grounding claims is possible.

[37] Some might argue that appeals to rational perception have precedent in mathematics. However, appeals to mathematics are tenuous given the widely varying accounts of mathematical knowledge on offer and the contested role of rational perception. By contrast, IBE is much more obviously used in scientific settings. For those who are worried that my IBE based epistemology may license too many grounding claims, it is worth highlighting that there is no clear reason to think that a purely rationalist account will do any better in this regard. For on my IBE based approach, the range of non-transparent grounding claims that we can establish is at least tethered in some clear way to what is already known and is further filtered by explanatory virtues.

[38] Rationalists might object that IBE can only provide knowledge of grounding claims if we first demonstrate its reliability and it is hard to see how we would do this without employing some kind of rational faculty. But this objection assumes that we can only acquire knowledge through the use of a method if we have already demonstrated the reliability of the given method. This assumption is doubtful. For if we apply it across the board it will follow that no method can provide knowledge at all. This is because in order to acquire knowledge through the use of any particular method we would first need to demonstrate the method's reliability using some other method whose reliability would require prior demonstration as well.

that some form of rationalism is the only viable option.[39] Rationalists might argue that their view therefore has the following advantage: it permits a uniform epistemology for transparent and non-transparent grounding claims. But I believe that we have important reasons to prefer a bifurcated epistemology. First, a uniform rationalist epistemology does little to illuminate the greater difficulty involved in knowing non-transparent grounding claims.[40] By contrast, suppose that we adopt a bifurcated epistemology on which a) IBE is the *only* means of knowing non-transparent grounding claims while b) transparent grounding claims are knowable in some other (presumably a priori) way. This overall epistemology accounts straightforwardly for the difficulty involved in knowing non-transparent grounding claims given that IBE is a difficult process and often requires empirical input. Importantly, this payoff requires that we treat IBE as the *only* means of knowing such claims rather than one method among many. Second, the bifurcated account has another important advantage. Specifically, chemical grounding claims such as "something is water because it is composed of H_2O" were clearly not known prior to empirical investigation (I take no stand here, though, on whether or not such claims are presently known). The bifurcated epistemology provides us with a viable explanation for this pattern of ignorance: namely, that there were no known scientific truths for such grounding claims to explain at the time. Consequently, IBE could not give us knowledge of these kinds of chemical grounding claims. The rationalist approach, by contrast, offers no apparent explanation. This gives us a further reason to prefer the bifurcated account over a uniform rationalist approach.

I will now consider whether or not the unionist thesis – namely, the thesis that grounding facts are identical to explanatory facts – provides us with an obvious epistemology for grounding claims.[41] For at first glance it may seem to do so given that explanatory facts appear to be knowable a priori. However, this appearance stems from a conceptual confusion. Specifically, it may be knowable a priori that a claim of the form "[φ] < [ψ]" is an explanation for ψ in the human giving sense of "explanation" discussed at the start of section II. However, this does not imply that

[39] Siscoe examines and argues against alternative epistemologies that appeal solely to conceptual competence.

[40] Siscoe himself accepts a bifurcated epistemology and only applies his rationalism to transparent grounding claims.

[41] See Thompson (2016) for a discussion of some of the challenges that the unionist view and its separatist rival respectively face. See Maurin (2019) for a critique of unionism and a discussion of the view's relevance to our knowledge of ground.

the truth of the claim "[φ] < [ψ]" is knowable a priori. For while grounding facts may be explanatory facts in the worldly sense discussed earlier on, there is no reason to think that statements that describe explanatory facts in this worldly sense are knowable a priori. By extension, the fact that grounding claims describe explanatory facts gives us no reason to believe that all grounding claims are knowable a priori.[42] Consequently, adopting unionism does not provide us with an alternative to my IBE based epistemology.

I will pause here to say something about my overall argumentative strategy. Specifically, I am defending the use of IBE in the metaphysical realm on the basis that it best explains our knowledge of non-transparent grounding claims. My argument therefore employs a form of IBE itself – specifically, IBE in the epistemological realm. However, this is a form of argument with significant precedent. For epistemological IBE is sometimes used to support the claim that we have a faculty of moral intuition on the basis that this view best explains our moral knowledge. Others use it as a way to support the view that imaginative reasoning yields knowledge on the basis that this view best explains our counterfactual knowledge (e.g. Williamson (2007, ch. 5)). I am therefore applying a well-worn argumentative strategy to support the use of IBE in the metaphysical realm.

I will end, as promised, with a discussion of the epistemic significance of the logic of ground. My proposed epistemology highlights the epistemic role of grounding rules. Specifically, they are needed to devise some of the explanations over which metaphysical IBE ranges. This observation in turn implies that debates over the logic of ground are not epistemically innocent. For conclusions concerning the logic of ground will presumably affect our assessment of the strength of rival type B explanations. An explanation which employs a rule that admits of exceptions, for example, will presumably be weaker than an explanation that employs no such rules.

[42] Skiles and Trogdon (2021) argue that separatists may infer that ground has certain general features (e.g. non-reflexivity) from the fact that explanations have such features together with the fact that the notion of ground backs metaphysical explanations. Whether or not this is true, their claim is not intended to serve as an account of how we know *particular* grounding claims – nor is it clear how it could be developed into such an account. I will therefore not dwell on it further.

Conclusion

I have assumed that non-transparent grounding claims are knowable and used this assumption to motivate the use of IBE in metaphysics and thereby demonstrate the epistemic role of certain grounding rules. Is there any way to justify my initial assumption? While I will not attempt to treat this question properly here, I will make the following observation: if metaphysical IBE is, in fact, a legitimate method then we may use it to acquire knowledge of non-transparent grounding claims and subsequently reflect on the fact that we have this knowledge. While this method will not convince skeptics of metaphysical IBE, it may convince those who simply harbor doubts about the possibility of knowing non-transparent grounding claims.

It is worth asking why the relationship between the logic and epistemology of ground has been overlooked. The answer, I believe, is that grounding theorists have focused more on logical and ontological issues than epistemology. I therefore hope to have shown that issues in the logic of ground cannot be fully separated from epistemic concerns.

Chapter 2: Meta-Semantic Explanation and 'Tonk'

According to inferentialists, the meanings of the logical constants ("constants" for short) are given by their rules of inference.[43] For example, the meaning of 'and' is given by the introduction rule that lets us infer 'P and Q' from P and Q as well as the elimination rule that lets us infer P and Q from 'P and Q'. What does it mean to say that these rules "give the meaning" of 'and'? While different authors answer this question in different ways, I will use the term "inferentialism" to describe any view on which understanding a constant consists in being disposed to follow certain deductive rule(s). These rules may then be said to give the meaning of the relevant constant. Going forward, I will assume that these rules are all formulated in a natural deduction setting. Moreover, I will confine my attention to inferentialist views that accept a realist (i.e. non-epistemic) view of truth as well as truth-conditional semantics.[44] Following Murzi and Steinberger (2015), I will call these views "orthodox".[45]

As Prior (1960) first observed, the inferentialist position faces a significant challenge. Specifically, if the meanings of the logical constants are given by their rules of inference then we may (seemingly) introduce a constant 'tonk' into our

---

[43] Importantly, I am only concerned with *logical* inferentialism (i.e. inferentialist views that concern the logical constants and nothing more).

[44] By "truth-conditional semantics" I refer to familiar Tarskian style truth theories that employ the notions of reference and satisfaction. My use of the term "realist", meanwhile, encompasses any non-epistemic view of truth and therefore includes deflationary views such as those espoused by Horwich (1982) and Field (1994). Importantly, one may accept both truth-conditional semantics and a realist view of truth while rejecting truth-conditional theories of meaning. See Soames (1998), for example, for a book-length development of this general outlook.

[45] Murzi and Steinberger (2015) use this label in order to contrast the kind of inferentialism I have just described with the kind espoused by writers such as Dummett (1982). Inferentialists such as Dummett identify truth with provability and reject truth-conditional semantics in favor of theories that state the verification conditions for each sentence in our language. This combination of views is often motivated by the following (alleged) requirement: that our understanding of any sentence must be manifestable to others (see Dummett (1993) for a defense of this requirement and Miller (2002) for a critical discussion). However, unlike Dummett, inferentialists in recent years have often been motivated by the desire to rehabilitate a form of the analytic-synthetic distinction and explain how a priori knowledge is possible (e.g. Peacocke (1993), Boghossian (1996), Warren (2015a)). This aim does not seem to require rejecting realist views of truth.

vocabulary that is governed by the following rules: from P infer 'P tonk Q' and from 'P tonk Q' infer Q. Together, these rules let us infer any claim from any other and thereby trivialize the consequence relation that characterizes the language to which they are added. As Belnap (1962) notes, an inferentialist may respond to Prior's challenge by restricting the rules that may give the meaning of a logical constant to those that yield a conservative extension of our original language. The rules that govern a constant C yield a conservative extension of our original language L if they meet the following criterion. Let Γ be an arbitrary collection of sentences in L and let φ be an arbitrary sentence in L. Now suppose that neither φ nor any member of Γ includes C. Then $L \cup \{C\}$ is a conservative extension of L if the rules that govern C do not let us derive φ from Γ *unless* we can derive φ from Γ without the rules that govern C.[46]

However, Belnap's proposal is not the only response on offer to Prior's challenge. Dummett (1991), for example, instead suggests imposing a proof-theoretic "harmony constraint" on the rules that give the meaning of any constant. This constraint requires that a constant's introduction and elimination rules exhibit a kind of proof-theoretic balance called "logical harmony".[47] The notion of logical harmony is formulated in different ways by different authors, though, and I will not attempt to settle its formulation here.[48] Nevertheless, on any viable formulation, the tonk-rules are not harmonious.[49]

More recently, Warren (2015b, 2020) has argued in favor of a view that I will call 'Unrestricted Inferentialism' (henceforth "UI"). On this view, any collection of

[46] The (alleged) need to impose a conservativity constraint is often used as the basis for an argument against classical logic. This is because the sentence '$((P \rightarrow Q) \rightarrow P) \rightarrow P$' – known as 'Peirce's Law' – is a theorem of classical logic. However, the theorem is not provable in standard formulations of classical logic without using the rules for negation. It follows that the classical negation operator yields a non-conservative extension of the propositional language that only contains the connectives $\wedge$, $\vee$ and $\rightarrow$ in the event that these connectives are all governed by their standard classical rules.

[47] Dummett conjectures that satisfaction of the harmony constraint guarantees satisfaction of the conservativity constraint and vice versa. See Steinberger (2013) for a critical discussion of this conjecture.

[48] See Read (2010), Tennant (2020) and Weir (1986) for different formulations of the notion of harmony.

[49] See Cook (2005) for another response to Prior's challenge. Cook presents his reply in a sequent calculus setting. Specifically, he argues that the addition of 'tonk' to a language is only problematic if we adopt certain structural rules such as transitivity.

rules whatsoever may give the meaning of a logical constant. Warren therefore rejects the widely-held assumption that inferentialists must rule out constants such as 'tonk'. Moreover, he claims that the tonk-rules preserve truth in any language that contains them. This claim may seem false when we observe that adding 'tonk' to the language of Peano Arithmetic (PA) lets us derive '1=3' from '1=1'. However, the claim that the 'tonk' rules preserve truth in the new language is viable, Warren argues, because the addition of 'tonk' to the language of PA may change the meanings of the sentences in the original language. If this is the case, then we have no obvious reason to deny that the tonk-rules preserve truth in the new language. Warren's view therefore resembles Carnap's (1937) Principle of Tolerance, according to which, "In logic, there are no morals. Everyone is at liberty to build up his own logic, i.e. his own form of language, as he wishes" (pg 52).

Given the available options, how should inferentialists respond to Prior's challenge? Warren's observation regarding potential meaning change complicates the assumption that we must prohibit 'tonk'. For it shows that this assumption requires more justification than has been typically given. Consequently, we cannot justify proof-theoretic constraints simply on the basis that they rule out 'tonk'. The purpose of this chapter is to argue that orthodox inferentialists (henceforth "inferentialists") should impose a constraint that I will call the "referential constraint" on the range of possible languages. This constraint rules out a wide range of languages that contain 'tonk' and is therefore incompatible with Warren's UI. Importantly, it is motivated on independent explanatory grounds.

I will pursue this task as follows. In section I, I will argue that inferentialists face an explanatory challenge called the "truth-table challenge" and I will distinguish this challenge from a related problem that Carnap (1934) raises.[50] In sections II and III, I will develop my own solution to this challenge. My solution requires that we draw a distinction between the sense and referent of a logical constant and adopt a non-deflationary view of truth. In section IV, I will examine potential alternatives to my solution and argue that they fall short in various ways. This gives inferentialists a reason to adopt my solution. Lastly, in section V, I will argue that my solution supports the referential constraint and therefore provides an independently motivated response to Prior's challenge.

[50] Roughly speaking, the truth-table challenge consists in explaining why our connectives exhibit their truth-functional behavior. I go to significant lengths to distinguish my challenge from Carnap's problem. This is because, unlike Carnap's problem, I do not believe that the challenge I have in mind has received much (if any) explicit attention.

Section I – *The Truth-Table Challenge.*

Inferentialists seek to characterize the meaning of a logical constant in terms of its inferential rules rather than its contribution to the truth conditions of the sentences in which the constant appears. Nevertheless, (orthodox) inferentialists must acknowledge the fact that our ordinary connectives (i.e. $\wedge$, $\vee$, $\neg$ and $\rightarrow$) obey the truth tables that we all encounter in introductory logic.[51] For example, the connective $\vee$ obeys the following truth table:

P	**Q**	**P $\vee$ Q**
T	T	T
T	F	T
F	T	T
F	F	F

What does it mean to say that the connective $\vee$ obeys the table above? Specifically, it means that it obeys the conditional claims encoded in each row of the table.[52] For example, the first row encodes the claim that $\forall P \forall Q$((P is true and Q is true) $\rightarrow$ $\ulcorner$ P$\vee$Q $\urcorner$ is true). The last row encodes the claim that $\forall P \forall Q$((P is false and Q is false) $\rightarrow$ $\ulcorner$ P$\vee$Q $\urcorner$ is false). It is worth highlighting that we can separate these conditional claims from the principle that every sentence is true or false as well as the principle that no sentence is both true and false. Without these further principles, the conditional claims that the truth tables encode do not incur a commitment to classical logic and may therefore be accepted by proponents of weaker logics as well. Inferentialists that endorse weaker logics (e.g. intuitionistic logic) therefore

[51] I confine my claim to orthodox inferentialists because it has been argued that anti-realists may eschew any commitment to the familiar truth tables (e.g. Murzi and Hjortland (2009)). I take no stand on this issue here. Additionally, I will set aside controversies regarding the behavior of the material conditional for the sake of this paper. If the reader wishes, they may therefore interpret my discussion as pertaining only to the connectives $\wedge$, $\vee$ and $\neg$.

[52] Importantly, it does *not* mean that the meaning of $\vee$ is given, in any sense, by its truth-table rather than its inferential rules.

have no obvious reason to reject these claims.[53] Moreover, every instance of the given claims is intuitionistically derivable from the T-schema together with the assumption that a sentence is false if and only if its negation is true.[54] Lastly, it is worth highlighting that proponents of intuitionistic logic may formulate their preferred account of logical validity in terms of Kripke trees, if they wish, and yet still accept the conditional claims that our familiar truth tables encode. For these claims do not concern the notion of validity. Rather, they concern the truth-theoretic behavior of our connectives and nothing more.

The truth tables encode highly general claims about our connectives. Moreover, let us suppose that a speaker uses the connectives $\wedge$, $\vee$, $\neg$ and $\rightarrow$ in such a way that their meanings are given by their standard rules. Then the connectives in the speaker's language will obey the conditional claims that our familiar truth-tables encode. A general connection of this kind requires an explanation. Inferentialists are therefore confronted with an explanatory challenge: namely, they must explain why the connectives $\wedge$, $\vee$, $\neg$ and $\rightarrow$ obey their familiar truth tables whenever their meanings are given by their standard rules. I will call this challenge the "truth-table challenge". Going forward, I will develop a solution to this challenge and argue that my solution avoids pitfalls that others do not. I will then argue that my solution supports rejecting UI.

I will pause here in order to say something about my use of the term 'explain'. First, I am assuming that the things we need to explain are true claims rather than facts in any worldly metaphysical sense. My use of the word 'fact' is therefore a shorthand for 'true claim'. Second, it is important that we distinguish explanations from justifications. For I am not claiming that inferentialists must justify the claims that our truth tables encode – rather, I am claiming that they must explain why these claims are true whenever the meanings of our connectives are given by their familiar rules. Third, I am not using the term 'explain' to signify any kind of *causal* explanation. Rather, I am using it to signify explanations of a *metaphysical* kind – specifically, the type B explanations discussed in the previous chapter.[55] We may illustrate this sense of "explain" via the following example. Suppose we wish to

[53] I separate the adoption of such logics from anti-realist views of truth. For it is possible that orthodox inferentialists may wish to revise classical logic for solely proof-theoretic reasons.

[54] The T-schema is the schema 'Tr('P')$\leftrightarrow$P' – where 'Tr' is the truth predicate and 'P' is a sentence variable.

[55] See Rubenstein (2024) for a detailed discussion of different approaches to metaphysical explanation.

explain the fact that any water molecule is composed of more atoms than any oxygen molecule. We can explain this fact by first observing that a molecule is a water molecule in virtue of the fact that it is composed of two hydrogen atoms and one oxygen atom. Next, we may observe that a molecule is an oxygen molecule in virtue of the fact that it is composed of two oxygen atoms.[56] It follows from these two observations that any water molecule is composed of more atoms than any oxygen molecule. We have therefore explained the fact at hand (in the relevant sense of "explain").

When addressing the truth-table challenge, I will take it for granted that certain facts either do not require or, at the very least, stand less in need of explanation than other facts. Importantly, I will assume that this class of facts includes any (true) claim to the effect that something is the case *in virtue of* something else (I will call such claims 'ontological reductions').[57] For example, the earlier claim that a molecule is an oxygen molecule in virtue of the fact that it is composed of two oxygen atoms is an ontological reduction. Reductions of semantic facts to more basic facts (i.e. meta-semantic facts) belong to this class of facts as well.[58] What form should an adequate solution to the truth-table challenge take? An adequate solution, on my view, should show how we can derive the claims that our truth tables encode from a) the assumption that the meanings of our connectives are given by their standard rules together with b) ontological reductions (or other principles that do not require further explanation).

It is worth distinguishing the truth-table challenge from a related issue that Carnap (1943) raises known as the "categoricity problem". This problem has received increased attention in recent years due to Raatikainen (2008). In order to explain the problem, I will introduce some terminology:

(1) A language L is a "propositional language" if it meets the following conditions. First, the vocabulary of L consists of sentence letters as well as sentential connectives that are governed by various inferential rules. Second, the set of sentences in L consists of the smallest set of expressions that

[56] This claim should not be misconstrued as a conceptual analysis. For it concerns oxygen molecules themselves rather than the concepts *oxygen* and/or *molecule*. Those who wish to use the terminology of grounding may replace the "in virtue of" locution with a grounding operator. See Audi (2012) for a general discussion of the "in virtue of" relation.

[57] See Dasgupta (2016) for a discussion of this issue.

[58] Kripke's (1980) well-known claim that a proper name refers to its referent in virtue of standing in a certain causal relation to it is an example (if true) of a meta-semantic fact.

includes all of L's sentence letters and that is also closed under the application of each connective in L.

(2) A language is a "standard propositional language" if it is a propositional language that contains all and only the connectives $\wedge$, $\vee$, $\neg$ and $\rightarrow$ (where these connectives are governed by their usual classical rules).

(3) A function v is a "valuation for L" – where L is a propositional language – if v's domain consists of all the sentences in L and v assigns a truth value to every sentence in its domain.

(4) A valuation v for a propositional language L is "bivalent" if it assigns either T or F (and only one of the two) to every sentence in L.

(5) A valuation v for a propositional language L is "admissible" if it meets the following criterion: for any set of sentences S in L and any sentence p in L, if S$\vdash$p then v assigns T to p if v assigns T to every sentence in S.[59]

(6) A bivalent valuation v for a standard propositional language L is a "standard valuation" if it obeys the following clauses (let φ and ψ be arbitrary sentences in L): v($\ulcorner \neg\varphi \urcorner$) = T iff v($\varphi$) = F; v($\ulcorner \varphi\wedge\psi \urcorner$) = T iff v($\varphi$) = T and v($\psi$) = T; v($\ulcorner \varphi\vee\psi \urcorner$) = T iff v($\varphi$) = T or v($\psi$) = T; v($\ulcorner \varphi\rightarrow\psi \urcorner$) = T iff v($\psi$) = T or v($\varphi$) = F.[60]

As Carnap observed, for any standard propositional language L, there are non-standard bivalent valuations that are also admissible (henceforth "NBA valuations"). Most obvious among these is the valuation that assigns T to every sentence in L. Another such valuation is that which assigns T to all of L's theorems and F to every other sentence.[61] The existence of such valuations is known as the "categoricity problem" for propositional logic. Their existence is thought to pose a problem for inferentialists because it shows that they (seemingly) have no principled reason for restricting their attention to standard valuations.[62]

Solutions to the categoricity problem generally fall into one of three categories. Solutions in the first category strengthen the rules that give the meanings of our connectives so that non-standard valuations are no longer

[59] 'S$\vdash$p' means that p follows from S via the rules that give the meanings of the connectives in L.

[60] The model-theoretic account of logical truth for propositional logic thus amounts to the following: that a sentence s in a standard propositional language is a logical truth iff all standard valuations for the relevant language assign T to s.

[61] See Belnap and Massey (1990) for more examples of NBA valuations.

[62] This is a problem, for example, if inferentialists wish to devise a model-theoretic account of validity. For such accounts restrict their attention to standard valuations.

admissible. This is Carnap's approach.[63] It is also the approach that bilateralists such as Rumfitt (2000) and Smiley (1996) take.[64] Solutions in the second category alter the definition of "admissible" so that only standard valuations are rendered admissible. This is the approach taken by Valle-Inclan (2024). Lastly, solutions in the third category impose an initial constraint on the space of valuations that are deemed worthy of consideration. This is the approach that Bonnay and Westerståhl (2016) take and I will introduce some terminology in order to explain their solution. First, let L be a standard propositional language. The authors use the term "non-trivial" as a label for any valuation that assigns F to at least one sentence in L. They use the term "compositional" as a label for any valuation v that meets the following criterion: for any n-ary connective C in L, there is a function F_C such that $v(\ulcorner C(p_1, \dots, p_n) \urcorner) = F_C(v(p_1), \dots, v(p_n))$ for every sentence of the form $\ulcorner C(p_1, \dots, p_n) \urcorner$ (i.e. every sentence with principal connective C). We may summarize this condition as follows: that the truth-value that v assigns to any sentence of the form $\ulcorner C(p_1, \dots, p_n) \urcorner$ is determined by the truth-values that v assigns to $p_1, \dots, p_n$. Bonnay and Westerståhl then argue that any standard propositional language contains at least one false sentence. Additionally, they argue that the truth value of any compound sentence $\ulcorner C(p_1, \dots, p_n) \urcorner$ in any standard propositional language is determined by the truth values of $p_1, \dots, p_n$.[65] These facts entitle us, in their view, to ignore valuations for standard propositional languages that are either trivial or that violate the compositionality constraint. Given that every NBA valuation is either trivial or non-compositional (a fact that they prove) it follows that we are entitled to ignore all non-standard valuations.

[63] Specifically, Carnap suggests adopting a multiple conclusion sequent calculus. I align with Steinberger's (2011) view, however, that such calculi are unsuited for the task of capturing the meanings of the logical constants.

[64] See Murzi and Topey (2021) for a third approach in this category that appeals to higher order rules. Murzi and Topey extend their approach to handle first and second order quantifiers as well.

[65] Specifically, they argue that this claim is implied by the fact that standard propositional languages are learnable together with the fact that such languages contain infinitely many sentences. In order for such languages to be learnable by finite beings such as ourselves, they argue, the semantic value of any sentence in the relevant language must be determined by the semantic values of the sentence's parts. This is tantamount (or so they argue) to the claim that the truth value of any compound sentence of the form $\ulcorner C(p_1, \dots, p_n) \urcorner$ is determined by the truth values of $p_1, \dots, p_n$.

The categoricity problem is clearly relevant to the truth-table challenge. For it highlights the following important fact: the assumption that the classical rules preserve truth does not singlehandedly imply that the connectives obey their familiar truth-tables. This observation constrains the way(s) in which we can solve the truth-table challenge. Nevertheless, the categoricity problem and the truth-table challenge are clearly distinct. For unlike the categoricity problem, the truth-table challenge is an explanatory challenge. Additionally, the categoricity problem is explicitly concerned with valuations whereas the truth-table challenge may be formulated without mentioning valuations at all. Lastly, solutions to the categoricity problem do not furnish immediate solutions to the truth-table challenge. To illustrate this fact, let us suppose that we have formulated a proof system that rules out all non-standard valuations (i.e. renders them non-admissible) thereby solving the categoricity problem.[66] Do we now have an immediate solution to the truth-table challenge? We do not. For by formulating such a proof system we have only shown that the connectives in our language obey their familiar truth tables in the event that a) our rules preserve the value T and b) every sentence in the language has the value T or F (and not both). This second limitation stems from the fact that the non-standard valuations that are ruled out by any such proof system are all bivalent (i.e. they each assign T or F to every sentence in the relevant language and only one of the two). It is possible, though, that certain sentences have other truth values or even multiple truth values.[67] If we wish to rely on the assumption that this is not the case when solving the truth-table challenge then this assumption requires an explanation.[68] Moreover, we cannot

[66] I take it as obvious that solutions to the categoricity problem that fall into the second and third categories do not provide any immediate solution to the truth-table challenge.

[67] For example, Strawson (1950) argues that certain statements that make use of definite descriptions are neither true nor false. Priest (2005), meanwhile, argues that sentences such as the liar-sentence are both true and false.

[68] We might wonder whether or not we can circumvent this issue by a) broadening the space of valuations under consideration to those that assign T, F or neither to any given sentence and then b) formulating a proof system that rules out all non-standard valuations within this broader space. Hjortland (2014) accomplishes this task by building upon the bilateralist framework developed by Rumfitt (2000) and Smiley (1996). However, performing this feat still does not provide us with an immediate solution to the truth-table challenge. This is because the feat only demonstrates that our connectives obey their familiar truth tables in the event that every sentence has one of the three given truth values. To solve the truth-table challenge, though, we still need to explain why every

explain the fact that a sentence is never both true and false by simply assuming that a sentence is false whenever its negation is true and vice versa. For this assumption only implies that a sentence is never both true and false if we also assume that a sentence and its negation are never both true. However, an inferentialist account of the logical constants should explain this fact as well.[69] For as Rumfitt (2000) notes, this fact is "an important part of what any competent user of a negation sign needs to know" (pg 792).

Section II – *Sense and Reference*.

I will now argue that inferentialists should introduce a distinction between the sense and referent of a logical constant in order to devise a satisfactory solution to the truth-table challenge.[70] It is important, going forward, that we do not characterize the notion of reference in terms of truth.[71] For example, we should distinguish the notion of reference from Tugendhat's (1970) notion of "significance". Tugendhat characterizes the significance of a proper name as "whatever property of names remains the same when we exchange them in otherwise identical sentences without changing their truth-value." We can reformulate this definition as follows: first, let L be a (first-order) language and N be a name in L. Now let the "significance class" of a name N consist of the set of names that can be substituted in place of N in any sentence in L without changing the truth value of the relevant sentence. We may now characterize the "significance" of N as the property that N shares with every name in its significance class.[72] We can extend Tugendhat's notion of significance, if we wish, to the logical constants by replacing the term 'names' in the definition just given with the phrase 'logical constant'. However, if we conflate the notion of reference with that of

---

sentence has exactly one of these values. A problem of this sort will clearly persist no matter how much we broaden the space of valuations that we are willing to consider.

[69] See Hand (1999) for a discussion of this issue.

[70] This distinction is often attributed to Frege (1892). Roughly speaking, an expression's sense consists of its cognitive content while an expression's referent is the object to which the expression refers. Frege famously identifies the referent of a sentence with its truth value.

[71] I use the term "reference" to describe the phenomenon by which an expression refers to an entity (i.e. the phenomenon by which an expression acquires a referent).

[72] This definition implies that the significance of a name is the object to which it refers (or more accurately, its reference to the given object).

significance we will then encounter the following issue: namely, in order to demonstrate that any of our ordinary constants has a particular referent we will need to rely on the claims that our familiar truth tables encode.[73] This will prevent us – on pain of circularity – from relying on any such results when solving the truth-table challenge. It is therefore worth highlighting that I am employing our pre-theoretic notion of reference rather than the notion of significance when I suggest that we draw a distinction between the sense and referent of a logical constant.

When developing my account of sense and reference, I will assume that the meanings of our familiar connectives are given by all their classical rules. This assumption will play an important role in the account I develop and so proponents of weaker logics (e.g. intuitionistic logic) may therefore reject my solution to the truth-table challenge. However, while inferentialism is often associated with a preference for intuitionistic logic over its classical rival, this preference is often motivated by the (perceived) need to impose a constraint on the range of possible connectives that rules out connectives like 'tonk'.[74] As I will show later on, though, we may impose a constraint of this kind that does not require any revision of classical logic.

I will now devise an account of sense and reference for the logical constants that is similar to the account that Peacocke (1987) develops.[75] For the sake of simplicity, I will restrict my attention to the sentential connectives. First, let a connective's sense consist of the collection of rules that give the connective's meaning. I will now introduce some definitions. Let L be a propositional language and let C_L be the set of connectives in L:

(7) A "n-ary Boolean function" is a function f that assigns a binary truth value (i.e. T or F) to every n-tuple of binary truth values.

(8) A "Boolean assignment for L" is a function B that assigns a n-ary Boolean function to each n-ary connective in C_L.

[73] This is because we will need to make use of these claims in order to demonstrate that substituting certain logical constants in place of others never alters the truth values of the sentences in which the substitutions occur.

[74] This preference is sometimes motivated by other considerations as well. See Dummett (1975) for arguments in favor of intuitionistic logic and Burgess (1984) for further discussion. Nevertheless, if proponents of intuitionistic logic are unable to solve the truth-table challenge then this may give orthodox inferentialists a reason to prefer classical logic.

[75] While I do not claim that this account is novel, my application of it is new.

(9) A "truth assignment for L" is a function M that assigns a binary truth value to each atomic sentence letter in L.

(10) For any Boolean assignment B for L and any truth assignment M for L, let us inductively define the valuation function "v_{BM}" as follows. First, for any sentence letter s in L, $v_{BM}(s) = M(s)$. Now let N be an arbitrary n-ary connective in C_L and let ⌜ $N(p_1, \dots, p_n)$ ⌝ be an arbitrary sentence in L with principal connective N – where $p_1, \dots, p_n$ are also sentences in L. Then $v_{BM}(⌜ N(p_1, \dots, p_n) ⌝) = B[N](<v_{BM}(p_1), \dots, v_{BM}(p_n)>)$ – where B[N] is the n-ary Boolean function that B assigns to N.

(11) A Boolean assignment B for L "fits L" if for any truth assignment M for L, v_{BM} is admissible (in the sense of "admissible" defined in (5)).

We may now introduce the following determination condition:

(12) Let N be a connective in L. Then N refers to o if and only if a) there is a unique Boolean assignment that fits L and b) this Boolean assignment assigns o to N.

There are several points worth noting with regard to (12). First, (12) should be understood as a meta-semantic account of the relation that obtains between a connective and its referent. Put otherwise, (12) tells us what makes it the case that N refers to o. Second, (12) is a holistic condition insofar as it implies that the possession of a referent (or lack thereof) by any connective in L depends not only on the connective's sense but also on the senses of all the other connectives in L. Third, (12) implies that either all the connectives in L have referents or none of them do. Fourth, (12) implies that each of the connectives in any standard propositional language refers to its expected Boolean function (i.e. the Boolean function that matches the connective's familiar truth table).[76] Lastly, (12) restricts the space of a connective's potential referents to the class of all Boolean functions.

This last observation raises the following question: namely, by confining the space of a connective's potential referents in this way have I thereby assumed that the connectives are truth-functional or that every sentence has the value T or F? If so, this would be problematic – for I have already argued that both claims should be explained rather than simply assumed when solving the truth-table challenge. It is therefore worth emphasizing that I have not made assumptions of either kind. For the assumptions just listed are semantic assumptions whereas (12) is a *meta*-semantic claim concerning the relation between connectives and their referents.

[76] For example, the expected Boolean function for ¬ consists of the function that assigns F to T and T to F. The claim that our connectives refer to their expected Boolean functions if we adopt my account of reference is easily provable and so I will not prove it here.

As such, it requires no further explanation (as per my discussion in section I). To highlight this distinction, we should observe that (12) implies nothing at all about a connective's truth-functional behavior.[77] To illustrate this point, let L be a standard propositional language and let P and Q be sentences in L. Let us suppose that P and Q both have the truth value T. Does it follow from the account of reference developed so far that ⌜ P∧Q ⌝ has the value T too? The answer is no. For the most we can infer from my account is that the referent of ∧ assigns the value T to the ordered pair of truth values that P and Q possess. Absent some further bridge principle(s), though, we cannot infer anything about the truth value of ⌜ P∧Q ⌝ itself. In light of this observation, (12) and its consequences should be clearly distinguished from any assumptions concerning the truth-theoretic behavior of the sentences in which the connectives occur.[78] Of course, those who are skeptical either that a) every sentence has the value T or F or that b) our connectives are truth-functional will likewise be skeptical of (12). However, this observation does not imply that we cannot appeal to (12) when *explaining* why a) and b) are true. Nor does it imply that we cannot rely on (12) when solving the truth-table challenge. For the truth-table challenge is an explanatory challenge and nothing more. Specifically, it does not include convincing skeptics of a) and/or b) that these principles are correct.[79]

Section III – *Solving the Truth-Table Challenge.*

I have now devised an account of sense and reference for the connectives. This account implies that each of our familiar connectives refers to its expected Boolean function (i.e. the Boolean function that matches the connective's familiar truth table). Does it follow now that our connectives obey their familiar truth tables? It does not. For as I have already observed, it is possible that a sentence's truth value

[77] On a related note, it is important that we distinguish Boolean functions from truth-tables. For a Boolean function is a mathematical object whereas a truth-table is a graphical representation of a series of claims.

[78] For example, we should distinguish (12) from Bonnay and Westerståhl's assumption that the connectives in any (learnable) propositional language are truth-functional.

[79] It is worth comparing the situation here with Dummett's (1974) assessment of soundness proofs. As Dummett observes, a soundness proof will not convince a skeptic of modus ponens that modus ponens is valid – for the proof employs modus ponens. Nevertheless, it may provide those who accept it already with an *explanation* for why the rule is valid.

is unrelated to the behavior of the Boolean functions to which the sentence's connectives refer. In order to solve the truth-table challenge I will therefore introduce some additional principles. In order to formulate principle (13) let ⌜ $N(p_1, \dots, p_n)$ ⌝ be an arbitrary compound sentence in a propositional language L – where N is the sentence's principal connective and $p_1, \dots, p_n$ are also sentences in L. Second, let R_N be the referent of N and let ⌜ $Tr(p_i)$ ⌝ denote the truth value of p_i for any $1 \leq i \leq n$. Then,

(13) ⌜ $N(p_1, \dots, p_n)$ ⌝ has the value T (or F) if and only if $R_N(<Tr(p_1), \dots, Tr(p_n)>)$ = T (or F).[80]

(14) A sentence is true (or false) iff it has the value T (or F).

Principle (13) states that ⌜ $N(p_1, \dots, p_n)$ ⌝ has the value T (or F) if and only if the referent of N assigns the value T (or F) to the n-tuple of truth-values possessed by $p_1, \dots, p_n$. Given that each connective in a standard propositional language refers to its expected Boolean function, it follows – assuming (13) and (14) are true – that our ordinary connectives all obey their familiar truth tables.[81] This solves the truth-table challenge and thereby vindicates the sense/referent distinction for logical constants that was proposed in the previous section.

One might wonder why inferentialists are allowed to rely on principles (13) and (14) when they are required to derive the connectives' truth tables from more fundamental principles. For it is worth recalling my original claim that an adequate solution to the truth-table challenge must include a derivation of the connectives' truth tables from principles that stand less in need of any explanation. It may seem unclear, though, why (13) and (14) meet this description. A viable response to this worry is that we can explain the truth of (13) by appealing to the following meta-semantic principle:

(15) Any compound sentence ⌜ $N(p_1, \dots, p_n)$ ⌝ has the value T (or F) *in virtue of* the fact that $R_N(<Tr(p_1), \dots, Tr(p_n)>)$ = T (or F).

---

[80] Those concerned about the semantic paradoxes may wish to limit 13) to languages that do not contain their own truth predicate. For on certain approaches to the paradoxes (e.g. Kripke (1975)) truth-functionality may not apply everywhere otherwise. A proper treatment of the paradoxes, however, lies outside the scope of this chapter.

[81] We may illustrate this claim with the following example. First, suppose that P and Q are both true. Then it follows via (14) that P and Q both have the truth value T in which case the referent of $\wedge$ maps the ordered pair $<Tr(P), Tr(Q)>$ to T. It then follows via (13) that ⌜ $P \wedge Q$ ⌝ has the truth value T. This implies via (14) that ⌜ $P \wedge Q$ ⌝ is true and so we have therefore established the first line of the truth table for $\wedge$.

Put otherwise, ⌜ N(p_1, … , p_n) ⌝ has the value T (or F) in virtue of the fact that the sentences p_1, … , p_n have the truth values they do and the referent of N maps the n-tuple of these values to T (or F). We can explain (14), meanwhile, by claiming that any sentence is true (or false) in virtue of the fact that the sentence has the value T (or F). Taken together, these meta-semantic claims constitute a partially developed non-deflationary theory of truth.[82] My solution to the truth-table challenge therefore requires that inferentialists reject deflationary theories of truth in favor of the account I provide.[83]

My explanation for principle (13) may prompt the following objection. Specifically, one might argue that (15) faces counterexamples in the form of intensional operators. For example, sentences of the form ⌜ ◊P ⌝ – where '◊' is the modal possibility operator – are compound sentences. However, the truth value of any such sentence is not determined by the truth value of P. Such sentences therefore do not have their truth values in virtue of the fact that the referent of '◊' assigns a particular truth-value to the truth-value of P.[84] How might we address these counter-examples to principle (15)? We might consider modifying the principle by restricting it to cases in which the relevant sentence's principal connective is a truth-functional operator. However, modifying (15) in this way would undermine my solution to the truth-table challenge. For in order to make use of the principle when solving the challenge, we would need to rely on the further assumption that the connectives $\wedge$, $\vee$, $\neg$ and $\rightarrow$ are truth-functional. However, this is a fact that we must explain rather than simply assume. I will therefore pursue an alternative solution to the problem posed by intensional operators. I will begin by distinguishing between two kinds of rules: "language-language" rules and "language-entry" rules.[85] A "language-language" rule is a rule that lets us infer certain sentences from other sentences (e.g. the conjunction elimination rule). A "language-entry" rule, by contrast, is a rule that lets us infer

[82] The account resembles the account of truth that Wittgenstein (1922) provides for molecular propositions.

[83] While deflationary accounts of truth come in various forms, most such views assert that there is nothing substantive to be said about truth beyond what is captured by instances of the T-schema. On one version, for example, any sentence of the form ⌜ 'S' is true ⌝ is said to be synonymous with the sentence named by 'S'. On this view, the truth predicate is therefore nothing more than a disquotational device.

[84] I consider it an open question whether or not such operators have referents.

[85] This terminology is often attributed to Sellars (1954). I do not mean to imply, however, that this attribution is correct or that Sellars' use of the terms is identical to mine.

certain sentences from certain non-linguistic items (e.g. perceptual states).[86] I will now modify principle (15) by restricting it to cases in which the meaning of the sentence's principal operator is given entirely by language-language rules. On the inferentialist view, the meanings of the connectives $\wedge$, $\vee$, $\neg$ and $\rightarrow$ meet this description. It is not clear, however, that the same may be said for intensional operators. For example, it is plausible that the meaning of the modal possibility operator '$\Diamond$' is given, in part, by a language-entry rule that lets us infer sentences of the form $\ulcorner \Diamond P \urcorner$ from certain acts of imagination (e.g. the act of imagining a world in which P is the case). If this is true, then the modal operators do not constitute a counter-example to the modified version of principle (15).[87] I leave open the question of whether or not a similar line of reasoning may be used to address the apparent counter-examples posed by other intensional operators.

The foregoing solution to the truth-table challenge also lets us explain why our familiar rules of inference preserve truth (i.e. the rules that give the meanings of $\wedge$, $\vee$, $\neg$ and $\rightarrow$).[88] We may explain this fact in the following way. First, let L be a standard propositional language and let B be the unique Boolean assignment that fits L. Let R be the truth assignment for L that assigns each atomic sentence's truth value to each atomic sentence.[89] Given that B fits L it follows that the valuation v_{BR} is admissible. Now let Γ be a set of true premises in L and let $\ulcorner N(s_1, \ldots, s_n) \urcorner$ be a sentence in L that is provable from Γ. Let us assume for the sake of simplicity that

[86] For example, such a rule might let us infer any sentence of the form $\ulcorner$ O is red $\urcorner$ – where O is a demonstrative – from any observation in which the referent of O appears to be red.

[87] While there are deductive systems that govern the modal operators (e.g. S5), it would be question begging to assert without argument that these rules fully capture the meanings of the relevant operators. Absent such an argument, though, it is not clear that the modal operators constitute a counterexample to principle (15).

[88] The proof I am about to give differs from a soundness proof. For unlike a soundness proof, it does not appeal to the truth-tables.

[89] I am assuming here that every atomic sentence has the value T or F (and not both). For without this assumption there is no guarantee that R exists. While this assumption may require an explanation, it is not clear that an account of the logical constants should be expected to provide it (as we shall soon see, the bilateralist solution relies on this assumption as well). Moreover, if we grant the assumption it then follows via my solution to the truth-table challenge that every compound sentence has the value T or F (and not both).

Γ only contains the single premise $\ulcorner M(p_1, \ldots, p_m) \urcorner$.[90] Given that $\ulcorner M(p_1, \ldots, p_m) \urcorner$ is true it follows via (14) that it has the value T and so it follows via (13) that the referent of M assigns T to $<Tr(p_1), \ldots, Tr(p_m)>$. I will now appeal to the following claim (henceforth "claim C"): that for any compound sentence $\ulcorner C(i_1, \ldots, i_k) \urcorner$ in L, the valuation v_{BR} assigns T to $\ulcorner C(i_1, \ldots, i_k) \urcorner$ if and only if the referent of C assigns T to $<Tr(i_1), \ldots, Tr(i_k)>$.[91] Given that the referent of M assigns T to $<Tr(p_1), \ldots, Tr(p_m)>$ it follows via claim C that v_{BR} assigns T to $\ulcorner M(p_1, \ldots, p_m) \urcorner$. Given that $\ulcorner N(s_1, \ldots, s_n) \urcorner$ is provable from $\ulcorner M(p_1, \ldots, p_m) \urcorner$ and v_{BR} is admissible it therefore follows that v_{BR} assigns T to $\ulcorner N(s_1, \ldots, s_n) \urcorner$. It follows again via claim C that the referent of N assigns T to $<Tr(s_1), \ldots, Tr(s_n)>$ in which case it follows via (13) that $\ulcorner N(s_1, \ldots, s_n) \urcorner$ has the value T. This implies, via (14), that $\ulcorner N(s_1, \ldots, s_n) \urcorner$ is true. We have now shown that if $\ulcorner M(p_1, \ldots, p_m) \urcorner$ is true then $\ulcorner N(s_1, \ldots, s_n) \urcorner$ is true as well. Given that $\ulcorner M(p_1, \ldots, p_m) \urcorner$ and $\ulcorner N(s_1, \ldots, s_n) \urcorner$ were arbitrary we have therefore explained why our familiar rules of inference preserve truth.[92]

Section IV – *Alternative Solutions*.

Let us take stock. I have developed a solution to the truth-table challenge and used it to explain why our familiar rules preserve truth. On its own, this does not give inferentialists a reason to adopt the claims I have made. For there may be other ways of solving the truth-table challenge. I will therefore examine two alternative solutions and argue that they fall short in various ways. This gives inferentialists a reason to adopt my solution.

---

[90] The following argument may clearly be modified to accommodate cases where Γ contains multiple premises. However, the modifications involved require the use of cumbersome notation. The argument may also easily be modified to address the cases in which the premise(s) and/or conclusion are atomic.

[91] This claim is provable via induction on the complexity of formulas.

[92] One might object to this explanation on the following grounds: that it appeals to my account of reference for the logical constants and this account already incorporates the assumption that our rules preserve truth. However, this objection rests on a mischaracterization of my account. For the account of reference that is developed in (11) and (12) does not mention truth preservation anywhere at all – rather, it is formulated solely in mathematical terms. Specifically, my account mentions the assignment of truth values to sentences by valuation functions (e.g. in (5)) but says nothing about the truth values of sentences themselves.

Rumfitt (2000) and Smiley (1996) devise a logical framework known as "bilateralism" that may seemingly be used to solve the truth-table challenge.[93] Its main point of departure consists of the following claim: that in addition to the speech act of assertion there is a distinct speech act of rejection. More specifically, on the bilateralist view the act of rejecting a sentence is distinct from the act of asserting its negation. For example, the act of rejecting the sentence "it is raining" is distinct from that of asserting the sentence "it is *not* raining." Bilateralists may then define the term 'correct' as follows: an assertion is "correct" if and only if the asserted sentence is true while a rejection is "correct" if and only if the rejected sentence is false.[94] With this framework in place, bilateralists develop logical systems that govern assertions and rejections. The rules in such systems are said to preserve the correctness of the speech acts they govern (in the sense of "correct" that was just defined). The act of asserting a sentence S is designated in such systems by the string of symbols '+S' while the act of rejecting S is designated by the string of symbols '–S'. Given that the symbols '+' and '–' are force markers they cannot embed within a sentence (i.e. unlike logical connectives, they can only appear before a sentence). Moreover, they are not a part of the sentences that we assert or reject and it would thus be mistaken to think that they have meanings that ought to be captured by our inferential rules (or in some other way).

Rumfitt and Smiley both propose the following system of bilateral rules:

$+\&I$:	$+A, +B \vdash +A\&B$	
$+\&E$:	$+A\&B \vdash +A$	$+A\&B \vdash +B$
$-\vee I$:	$-A, -B \vdash -A \vee B$	
$-\vee E$:	$-A \vee B \vdash -A$	$-A \vee B \vdash -B$
$+\supset E$:	$+A \supset B, +A \vdash +B$	
$-\supset E$:	$-A \supset B \vdash +A$	$-A \supset B \vdash -B$
$+\neg I$:	$-A \vdash +\neg A$	
$+\neg E$:	$+\neg A \vdash -A$	
Reductio:	If $\Gamma, \alpha \vdash \perp$, then $\Gamma \vdash \alpha*$	
Non-Contradiction:	From $\alpha, \alpha*$, infer $\perp$	
Structural Rules:	Thinning, Cut, Reflexivity	

[93] This framework is devised, in part, to solve the categoricity problem. As I have already argued, though, we should distinguish the categoricity problem from the truth-table challenge.

[94] This definition of 'correct' is not intended to capture the term's ordinary use. Rather, it is simply meant to characterize the feature of assertions/denials that is preserved by the bilateral rules. Importantly, it should not be taken to imply that true sentences are always assertible – for this is a claim that Rumfitt (2002) rejects in his reply to Dummett (2002).

The system above consists of a) operational rules that give the meaning of each connective, b) the so-called "coordination principles" of Reductio and Non-Contradiction and c) the familiar structural rules of thinning, cut and reflexivity. If 'α' represents an act of asserting or rejecting a sentence then 'α*' represents the opposite act. For example, if 'α' represents the act of asserting a sentence S then 'α*' represent the act of rejecting S (and vice versa). The coordination principles govern the interaction of the force markers '+' and '–' and it is worth noting that they are restricted to atomic sentences (i.e. one only obtains an instance of either rule by replacing 'α' with the assertion or rejection of an atomic sentence).[95] Lastly, following Tennant's (1999) suggestion – and contra Dummett (1991) – Rumfitt treats the symbol '$\perp$' as a logical punctuation sign that marks the point in a derivation where a contradiction has been reached.[96] The symbol '$\perp$' is therefore devoid of any content.

Rumfitt's system first lets us explain why a sentence and its negation are never both true. As Rumfitt observes, we may use an inductive proof to demonstrate that the Non-Contradiction rule applies to every sentence in our language (i.e. to demonstrate that we may derive '$\perp$' in our system from α and α* even when 'α' designates the assertion or rejection of a compound sentence).[97] We may then reason as follows. First, let us suppose for the sake of contradiction that a sentence S and its negation are both true. Then it follows that the assertion and rejection of S are both correct. Given that '$\perp$' is derivable from the combination of ⌜ +S ⌝ and ⌜ –S ⌝ and the bilateral rules preserve correctness it follows that '$\perp$' is correct. However, '$\perp$' is devoid of any content and is therefore not the sort of thing

[95] Rumfitt also develops an alternative bilateral natural deduction system. However, the system is unwieldy due to its large number of rules and so Rumfitt therefore provides the more compact system given here instead. While it may seem as though Rumfitt's natural deduction system does not require the use of any coordination principles, this is not the case. For as Kurbis (2016) observes, Rumfitt's natural deduction system does not yield the full extent of classical logic unless it is supplemented with coordination principles such as those employed in the system given above.

[96] Dummett, by contrast, treats the symbol '$\perp$' as a zero-place logical connective with an elimination rule that lets us infer any atomic sentence from it. However, Rumfitt finds this treatment unsatisfactory because he thinks that it does not let us explain why '$\perp$' is always false and by extension why a sentence and its negation are never both true.

[97] Ferreira (2008) demonstrates that no such inductive proof is possible for the Reductio rule. As a result, obtaining classical logic requires that the rule not be restricted to atomic sentences.

that can be correct. We must therefore reject our supposition that S and its negation are both true.

The bilateralist may now attempt to solve the truth-table challenge in two steps. First, the bilateralist may inductively prove that every sentence in a standard propositional language is true or false by first assuming that every *atomic* sentence is true or false and then making use of the further claim that the bilateral rules preserve correctness. Second, the bilateralist may then derive the conditional claims encoded in the familiar truth tables from the assumption that the bilateral rules preserve correctness together with the claim that every sentence is true or false. For example, let P and Q be sentence letters in a standard propositional language and let us assume that P is true and Q is false. Then it follows via the earlier definition of 'correct' that ⌜ +P ⌝ and ⌜ –Q ⌝ are both correct. Now let us suppose for the sake of contradiction that ⌜ P&Q ⌝ is true. Then ⌜ +P&Q ⌝ is correct and so it follows from the assumption that the +&E rule preserves correctness that ⌜ +Q ⌝ is correct as well. However, we have already established that ⌜ –Q ⌝ is correct and given that no sentence is both true and false it follows that ⌜ +Q ⌝ and ⌜ –Q ⌝ are *not* both correct. We must therefore reject our supposition that ⌜ P&Q ⌝ is true. Given that every sentence is true or false it follows that ⌜ P&Q ⌝ is false and so we have therefore established one of the lines in the familiar truth table for '&'. Similar reasoning lets us establish every line in each of the familiar truth tables. This concludes the bilateralist solution to the truth-table challenge.

While the bilateralist solution may seem promising, it faces several issues. First, the meanings of the connectives are more plausibly given by Rumfitt's natural deduction rules than the sequent calculus rules that are shown in the system above. However, the introduction rule for $\rightarrow$ in Rumfitt's natural deduction system only lets us infer an assertion of ⌜ P$\rightarrow$Q ⌝ after we have derived an assertion of Q from an assertion of P. As Kürbis (2017) notes, though, it does not seem as though we typically assert P in the course of establishing ⌜ P$\rightarrow$Q ⌝. Rather, we usually "suppose" P for the sake of argument and then derive Q from P. Rumfitt's system does not seem capable of capturing this kind of reasoning, though, and therefore seems to omit a key part of the meaning of $\rightarrow$.[98]

More importantly, the bilateralist solution does not explain why any of our familiar rules (i.e. those that give the meanings of $\wedge$, $\vee$, $\neg$ and $\rightarrow$) preserve truth (or correctness) at all. For the fact that these rules give the meanings of our

[98] See Kürbis (2023) for further discussion of this issue.

connectives does not explain, in and of itself, why they preserve truth.[99] I wish to emphasize here that I am not questioning whether or not these rules *do* preserve truth. Rather, I am merely observing that it is not at all obvious *why* this is so. If the truth functional behavior of the connectives requires an explanation, though, then the fact that our rules preserve truth surely requires an explanation as well. Unlike the bilateralist solution, my solution explains this fact (as I have already shown).

Alternatively, deflationists might argue that we can solve the truth-table challenge by appealing to the T-schema. For example, we might try to show that any conjunctive sentence is true if both of its conjuncts are true via the following argument:

1. Tr('P$\wedge$Q') $\leftrightarrow$ (P$\wedge$Q). [*Instance of the T-schema*]
2. Tr('P') $\leftrightarrow$ P. [*Instance of the T-schema*]
3. Tr('Q') $\leftrightarrow$ Q. [*Instance of the T-schema*]
4. (Tr('P')$\wedge$Tr('Q')) $\rightarrow$ Tr('P$\wedge$Q'). [*Follows from 1., 2. and 3.*]

We appear to have shown that if 'P' and 'Q' are true then so is 'P$\wedge$Q'. The problem with this argument, though, is that it is an argument schema rather than a genuine argument given that 'P' and 'Q' are sentence variables. The conclusion is therefore schematic and so we have not established the desired result. Moreover, there does not appear to be any way of modifying the schema into a genuine argument for the desired conclusion. For while we can replace the variables 'P' and 'Q' with particular sentence letters, doing so only lets us establish that the resulting sentence is true if both its conjuncts are true. This claim is distinct, though, from the general claim that *any* conjunctive sentence is true if both its conjuncts are true.[100]

We might try to overcome this obstacle in one of two ways. First, we might try to establish the desired result by using induction on the complexity of formulas and appealing to the assumption that every step in the argument schema preserves truth whenever the variables 'P' and 'Q' are replaced by particular sentences. The problem with this approach, though, is that it relies on the explicit assumption that our rules preserve truth. As I have already argued, though, the fact that our rules preserve truth requires an explanation as well.[101]

---

[99] For example, the fact that our understanding of $\vee$ consists (partly) in our disposition to follow the $\vee$ introduction rule does not explain, in and of itself, why this rule preserves truth.

[100] This limitation has often been observed in the deflationist literature. However, it has rarely been connected explicitly to explanatory concerns.

[101] We might try appealing to the T-schema in order to show that our rules preserve truth. However, the resulting arguments will again be schematic and will therefore not establish

Alternatively, we might try to salvage the present solution by following Field's (1994) lead and using substitutional quantifiers. Specifically, we might seek to derive the claims that our truth tables encode from the premise that 'Πp(Tr('p') $\leftrightarrow$ p)' – where 'Π' is a substitutional quantifier. The given premise is true if and only if every instance of the schema 'Tr('P') $\leftrightarrow$ P' is true – where we obtain instances of the schema by replacing 'P' with a sentence in the relevant language. If we rely on this premise, though, we must then explain why it is true. Following Field, we might appeal to a deflationary conception of truth in order to explain why every instance of 'Tr('P') $\leftrightarrow$ P' is true. We might then attempt to explain why the premise 'Πp(Tr('p') $\leftrightarrow$ p)' is true by appealing to the assumption that any sentence of the form $\ulcorner \Pi s(\varphi(s)) \urcorner$ is true whenever every substitution instance of φ(s) is true. However, this latter fact is clearly harder to explain than the truth functional behavior of our ordinary connectives. This difficulty is compounded by the lack of clarity surrounding the inference rules that govern substitutional quantifiers. It is also worth noting that we cannot appeal to the principle 'Πp(Tr('p') $\leftrightarrow$ p)' when explaining the truth-conditional behavior of substitutional quantifiers. For we would otherwise end up relying on the principle 'Πp(Tr('p') $\leftrightarrow$ p)' when explaining why the given principle is true. We therefore do not seem to gain much, if anything, by appealing to substitutional quantifiers as a way of addressing the truth table challenge.[102]

Lastly, Field (2006) proposes that we posit rules that let us infer general (objectual) claims from schemas – thereby circumventing the original issue. However, derivations that employ such rules together with schematic reasoning depart significantly from familiar forms of explanation in which the relevant derivations rely solely on standard logical resources and grounding principles. Consequently, it is not obvious that derivations of the kind that Field proposes are, in fact, explanatory. I will therefore assume that explanations (such as my own) in which the derivations do not involve such departures are preferable to those that do.

---

the desired result. Moreover, if we try to use induction on the complexity of formulas in order to overcome this issue we will again need to rely on the assumption that our rules preserve truth. However, this is the very fact that we are trying to explain.

[102] It is worth noting that the intelligibility of substitutional quantification is also a matter of controversy (see Van Inwagen (1981)).

Section V – *Tonk*.

I will now argue that my solution to the truth-table challenge is incompatible with Warrenʾs Unrestricted Inferentialism. On Warren's view, any collection of rules whatsoever may give the meanings of a language's logical constants and preserve truth as well. Importantly, these rules may include the notorious 'tonk' rules. While it may seem as though the 'tonk' rules do not preserve truth, Warren rejects this claim on the grounds that the addition of 'tonk' to a language L may change the meanings of the original sentences in L. Warren observes that the same may be said for the addition of any new constant to L. If this is true, then we do not have any obvious reason to conclude that the 'tonk' rules do not preserve truth in the new language.

Warrenʾs observation undermines straightforward arguments against the intelligibility of 'tonk'. Nevertheless, I will argue that we have good reason to prohibit a wide range of languages that contain it. Going forward, let a connective N be a "language-language" connective if its meaning is given entirely by language-language rules. I will now argue that we should prohibit languages that contain both a) the connective 'tonk' as well as b) true compound sentences of the form ⌜ $N(p_1, \dots, p_n)$ ⌝ – where 'N' is a language-language connective. I will call such languages "tonk languages". To show why we should prohibit tonk languages, let L be a tonk language. Then L contains a true compound sentence of the form ⌜ $N(p_1, \dots, p_n)$ ⌝ – where 'N' is a language-language connective. It follows via (13) and (14) that the referent of 'N' assigns T to the n-tuple of truth values of $p_1, \dots, p_n$. This in turn implies that 'N' has a referent. However, given that L contains 'tonk' it follows via my account of reference that none of the language-language connectives in L have referents. We should therefore prohibit tonk languages.

The argument just given clearly generalizes to any language that a) contains true sentences of the form ⌜ $N(p_1, \dots, p_n)$ ⌝ – where 'N' is a language-language connective – and in which b) the rules that give the meanings of the language-language connectives do not determine referents. We should therefore prohibit such languages and I will call this constraint the "referential constraint". This constraint provides us with an independently motivated response to Prior's original challenge. It implies, for example, that there is no language that contains both 'tonk' as well as our familiar connectives and in which the meaning-constitutive rules all preserve truth.

Conclusion

I have argued that orthodox inferentialists must solve the truth-table challenge and that doing so motivates the referential constraint. Do the conclusions that I have drawn apply to all inferentialists (i.e. non-orthodox ones included)? A proper treatment of this question requires an assessment of the debate between realists and anti-realists and this issue lies outside the scope of this chapter.[103] Nevertheless, I will briefly remark on why I do not wish to adopt an epistemic view of truth. First, proponents of the epistemic view must confront Fitch's (1963) paradox – namely, the demonstration that if all truths are knowable then all truths are known. Second, it is unclear how exactly we should make sense of the modality involved in our use of the term 'knowable'. For it is unclear whether or not claims about infinite totalities, other minds or the distant past are "knowable" in the relevant sense.[104] Lastly, while some proponents of the epistemic conception of truth equate the truth of a claim with the existence of a proof of it, others equate the truth of a claim with the possibility of obtaining such a proof.[105] While these difficulties may or may not be surmountable, they provide significant (prima facie) reasons for favoring orthodox views. Whatever the final verdict on these issues may be, however, I hope to have shown that explanatory concerns are of greater relevance to logical issues than has been generally recognized so far.

[103] See Loux (2003) for a useful overview of this debate.

[104] See Dummett (1968) for a discussion of past-tense claims. Dummett's epistemic conception of truth leads him to the radical conclusion that the past is not fixed but changes over time.

[105] See Raatikainen (2004) for a discussion of this issue.

Chapter 3: An Analysis of Logical Harmony

As we have already seen, inferentialists claim that the meanings of the logical constants ("constants" for short) are given by their rules of inference. But which rules are these? On one approach, the rules in question are those that we are disposed to follow or some subset thereof (I will call proponents of this view "dispositionalists").[106] Importantly, we may distinguish these rules from those that "govern" the logical constants – where a rule "governs" a constant if a) the schema that represents the rule includes the constant in question and b) any inference that is carried out in accordance with the rule transmits justification from its premises to its conclusion.[107]

Various inferentialists (e.g. Dummett (1991)) have proposed imposing a so-called "harmony requirement" on any logical constant. A constant meets this requirement when its introduction and elimination rules satisfy a certain proof-theoretic constraint called "logical harmony" ("harmony" for short). Rumfitt (2016) calls this requirement the "harmony thesis" and characterizes it as follows: "a connective will not possess a sense – it will not have a coherent meaning – unless its logical behavior is regulated only by a pair of harmonious introduction and elimination rules."[108] While the notion of harmony is formulated in different ways by different authors, it is generally regarded as a local rather than a global feature insofar as it applies to particular rules in our deductive system rather than our system as a whole.[109] Moreover, the notion is meant to ensure that a connective's introduction and elimination rules exhibit a suitable balance. Importantly, on many formulations, the intuitionistic rules for negation are deemed harmonious whereas the classical rules are not. If the rules that give a connective's meaning must satisfy a harmony requirement, then this outcome seemingly poses a problem for the view defended in the previous chapter. For the view required that the meaning of the negation operator is given by all its classical rules.

---

[106] I will later distinguish this form of inferentialism from another form that I will call "normative inferentialism".

[107] For example, the conjunction introduction rule has premises φ and ψ and conclusion $\varphi \wedge \psi$. The rule governs '$\wedge$' given that '$\wedge$' appears in the schema that represents it and whenever we have justification for both φ and ψ and infer $\varphi \wedge \psi$ from the given premises we obtain justification for $\varphi \wedge \psi$.

[108] Throughout the rest of his paper, Rumfitt examines various arguments in support of this thesis and finds them all lacking.

[109] We may therefore contrast the notion of harmony with proof-theoretic notions such as normalizability – which apply to whole deductive systems.

In section I of this chapter, I will distinguish between several versions of the harmony requirement and will consider possible reasons for adopting them each. These versions are rarely distinguished from one another in the relevant literature. I will then argue that inferentialists have no reason to adopt any such requirement. It follows that such requirements do not provide an independently motivated response to Prior's original challenge. In section II, I will argue that despite this outcome, the notion of harmony nevertheless retains a valuable role. In section III, I will assess existing accounts of the notion in light of this particular role. In section IV, I will discuss Dummett's (1991) proof-theoretic account of logical validity. While this discussion may seem tangential at first, it will serve as the basis for the account of harmony that I develop in section V.

Section I – *The Harmony Requirement.*

Before assessing whether or not we should adopt the harmony requirement, we must clarify the requirement itself. All rules, going forward, will be formulated within a Gentzen-style natural deduction system. An introduction rule for a constant C is a rule that may be represented by a schema whose conclusion has principal operator C (e.g. conjunction introduction).[110] An elim rule for a constant C, meanwhile, is a rule that may be represented by a schema whose major premise has principal operator C but whose conclusion does not include C (e.g. modus ponens). It appears as though there are different requirements that inferentialists may have in mind when they speak of the so-called "harmony requirement". The first such requirement consists of the following – where C is a constant and R is the set of rules that we are disposed to follow with respect to C:[111]

1) If C is a meaningful constant then the rules in R are all harmonious.

This principle consists of a constraint on the rules that we are *disposed* to follow with respect to any meaningful constant. Importantly, it commits inferentialists to the following claim:

2) The classical negation operator is not a meaningful constant.[112]

---

[110] I will say more about the distinction between rules and their schematic representations in section II.

[111] It is worth highlighting that while "being harmonious" is a property of individual rules, the possession of this trait (or lack thereof) by a particular rule will depend on the other rules in the system.

[112] I will call a constant the "classical negation operator" if we are disposed to follow the classical negation rules with respect to it.

This is because the (standard) classical rules for negation are not harmonious on any viable conception of harmony.[113] Proponents of 1) with classical sympathies may try to avoid 2) in one of two ways. First, they might adopt a bilateral formulation of classical logic seeing as the rules that such systems employ are all harmonious.[114] However, I have already discussed the problems with bilateralism in the previous chapter. Alternatively, they might devise a non-standard unilateral formulation of classical logic in which all the rules are harmonious (e.g. as seen in Milne (2015)). However, the rules employed in such systems are highly unnatural and it is therefore implausible to claim that ordinary speakers are disposed to follow them. Claim 2) therefore seems to be an inescapable consequence of 1). The problem with 2), though, is that it employs the negation operator. As a result, 2) implies the following: that it is unintelligible in the event that we are disposed to follow all the standard classical rules. Given that proponents of classical logic are disposed to follow these rules, they will therefore be justified in concluding that 2) implies its own unintelligibility and is therefore not true.[115] Moreover, given that 2) is a consequence of 1) they will be justified in concluding that 1) is not true either. Proponents of 1) might try to defend it by observing that we can avoid the force of the given argument by abandoning our disposition to follow certain classical rules. For if we are not disposed to follow all the classical rules then 2) will no longer imply its own unintelligibility. While this may be true, this response will not be convincing to proponents of classical logic unless they have *already* been given a convincing reason to modify their logical practices. However, no such reason has yet been given.[116] Moreover, absent such a reason, proponents of classical logic have every reason to treat 2) and 1) as untrue. Given that my starting point is classical, I will therefore reject 1) for now.

---

[113] The truth of this claim will become apparent in sections III–V.

[114] See Rumfitt (2000) and Smiley (1996) for a presentation of the bilateralist view.

[115] This argument is a form of self-refutation. While it will not give intuitionists any reason to reject 1), this does not make it any less compelling to proponents of classical logic. It is also worth highlighting that proponents of classical logic need not employ any form of reasoning that intuitionists reject in order to reach the conclusion that 2) implies its own unintelligibility. Rather, they need only recognize that they are disposed to follow all the classical rules.

[116] Generally speaking, proponents of classical logic have no reason to modify their logical practices when presented with a (supposed) demonstration of 2). For 2) is self-refuting from the vantage point of a proponent of classical logic. Consequently, so is any premise that implies 2).

Inferentialists might instead formulate the harmony requirement as follows:

3) If C is a meaningful constant then the rules that give the meaning of C are all harmonious.

Why should inferentialists adopt 3)? One might argue that the principle is needed in order to rule out the existence of any constant the meaning of which is given by Prior's (1960) rules for 'tonk'. The constant 'tonk' has an intro rule that lets us infer 'P tonk Q' from P and an elim rule that lets us infer Q from 'P tonk Q'. Together these rules let us infer any claim from any other and thereby trivialize the language to which they are added. However, connectives such as 'tonk' are already ruled out by the referential constraint that I defended in the previous chapter and so principle 3) is not needed to accomplish this task.[117]

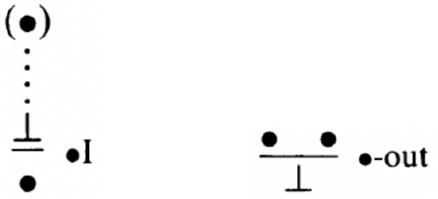
Steinberger (2011) instead defends the harmony requirement on the grounds that it upholds his so-called principle of innocence ('POI' for short). He formulates this principle as follows: "it should not be possible, solely by engaging in deductive logical reasoning, to discover hitherto unknown (atomic) truths that we would have been incapable of discovering independently of logic." However, as Rumfitt (2016) observes, even if we accept this principle, it is unclear why the constraint it imposes on our deductive system must be upheld through the use of a local requirement such as that of logical harmony. One might therefore attempt to bolster Steinberger's reasoning into a defense of principle 3) in the following way. First, let us assume that the POI is correct and let L be a language that is governed by the rules in a deductive system D – where D satisfies the POI. Suppose now that we add a new constant C to L – where R is the set of rules that would give the meaning of C *in the event that* C was meaningful.[118] I will call the expanded language L+. Suppose as well that the addition of R to D lets us prove atomic sentences in L that are not provable via D alone. What happens after we add C to

[117] The referential constraint consists of the following requirement: that the rules that govern our logical constants must determine a referent for each and every one of the relevant constants. If we assume, as I will, that the rules that give the meaning of a constant are always among the rules that govern it then this constraint implies that the 'tonk' rules do not give the meaning of any connective.

[118] I will assume throughout this paper that the rules that give the meaning of a logical constant are always among the rules that govern it. Those worried about the semantic paradoxes may restrict this assumption to languages that do not contain their own truth predicate. This restriction will help ensure that whenever a rule governs a constant it preserves truth as well. Moreover, it will help us avoid any commitment to the claim that inferences involving paradoxical sentences transmit justification.

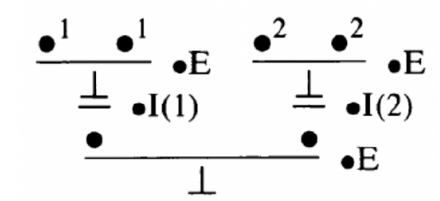
L? As proponents of the POI, we are (seemingly) faced with two options. First, we may claim that C is *not* a meaningful constant while maintaining that the original constants in L remain meaningful following the addition of C to L. The trouble with this option, though, is that it is unclear why we should single out C as defective rather than some other constant(s) in L+. For while the addition of R to D leads to a violation of the POI, it may nonetheless be possible – given that the POI is a global constraint – to restore the integrity of the POI by removing some constant(s) from L+ other than C. Alternatively, rather than singling out C as defective, we might claim instead that the constants in L+ are *all* rendered defective by the addition of C to L. However, this second option introduces an undesirable vulnerability into our logical practices. We therefore seem to face a dilemma. In light of this fact, it is worth highlighting that we can (seemingly) avoid the given dilemma by adopting the harmony requirement. For if the rules in R are harmonious then the addition of R to D will (presumably) not violate the POI. On the other hand, if the rules in R are *not* harmonious then the harmony requirement gives us a principled reason to single out C as defective rather than some other constant in L. In either case, we avoid the dilemma that we previously faced and this safeguard gives us a reason to adopt the harmony requirement.

The problem with this argument, though, is that it relies on the following (questionable) assumption: that the addition of harmonious rules to D can never result in a violation of the POI. However, we can only properly evaluate this assumption once we have an adequate formulation of the notion of harmony. Moreover, there is a reason to doubt the assumption's correctness. In particular, the rules for the following zero-place connective appear to be harmonious (this connective is taken from Read (2010)):[119]



These rules clearly lead to a violation of the POI, though, when they are added to D. This is because they let us prove any claim whatsoever when taken together – as we can demonstrate via the following derivation:

[119] The reason for claiming that these rules seem harmonious will become evident when I discuss existing accounts of harmony in section III.



This result casts doubt on the defense of 3) that I have been considering.

I will now consider one final version of the harmony requirement. To explain the requirement, I will introduce some terminology. Let L be a propositional language and let an "argument" consist of a single step derivation (roughly speaking – a collection of premises and a conclusion). Then:

4) An argument in L is a "justificatory" argument iff inferring the argument's conclusion from its premises would transmit justification for the conclusion in the event that we had justification for the argument's premises.[120]

It follows from the definition above that if a rule governs a constant in the sense described in the introductory paragraph then all instances of the rule are justificatory arguments. Additionally, it is clear that not all valid arguments are justificatory. For an argument that has as its conclusion Fermat's Last Theorem and has as its premises the axioms of ZF that are needed to prove the theorem is valid. Simply inferring the conclusion from the premises, however, does not transmit justification from the latter to the former and so the argument is not justificatory.

Now let P be a proof system for L – where a proof system is a set of schematically describable rules. Let A be a set of arguments in L. Then:

5) "P codifies A" iff any instance of any member of P belongs to A and any member of A is an instance of some member of P.

We may now introduce a third version of the harmony requirement:

6) Let J be the set of justificatory arguments in L. Then there is a proof system that codifies J in which all the rules are harmonious.

How might we argue in favor of 6)? To do so, we might appeal to a view that I will call "Gentzen's thesis". This thesis consists of the claim that the meanings of the logical constants are given solely by their introduction rules. The thesis has its origins in Gentzen's (1934) famous remark concerning the logical constants that, "the introductions constitute, so to speak, the definitions of the signs in question, and the eliminations are actually just consequences thereof." Gentzen's thesis motivates claim 6) for the following reason: specifically, we need some criterion

[120] I assume that an argument's premises may include sub-derivations. Having justification for a sub-derivation requires having justification for all its undischarged premises and performing all the inferences within it as well.

that lets us determine when an argument in L is justificatory. If a constant's meaning is given solely by its introduction rules, though, then we cannot simply identify the justificatory arguments in L with those that are instances of the rules that give the meanings of the logical constants in L. This is because instances of elimination rules will clearly be justificatory as well. We must therefore devise some other criterion.[121] Moreover, the most obvious candidate consists of the following:

7) An argument is justificatory iff it is an instance of either a) one of the intro rules that give the meaning of a logical constant or b) an elim rule that is in harmony with the aforementioned intro rules.

Now let P be the proof system that contains the intro rules that give the meanings of the logical constants as well as the elim rules that are in harmony with those intro rules. It then follows from 7) that P codifies the set of justificatory arguments in L. In other words, it follows that there is a proof system that codifies J in which every rule is harmonious. Gentzen's thesis therefore provides us with a reason to adopt the harmony requirement described in 6).

There is an obvious issue, however, with the argument just given in support of 6) – namely, it is unclear why we should endorse Gentzen's thesis. I will therefore formulate an argument that dispositionalists might try to use in its favor. Specifically, there are several (seemingly) viable answers, on the dispositional view, to the general question of what makes a constant meaningful.[122] These answers consist of the following:

a) A constant C is meaningful in virtue of the fact that there are rules we are disposed to follow that pertain to C.

b) A constant C is meaningful in virtue of the fact that there are *introduction* rules we are disposed to follow that pertain to C.

c) A constant C is meaningful in virtue of the fact that there are rules we are disposed to follow that pertain to C and that also meet certain technical constraints.

---

[121] See Boghossian (2003) for a discussion of the relationship between meaning constitutive rules and inferences that transmit justification.

[122] We should distinguish the question of what gives a specific constant its particular meaning from the general question of what makes a constant meaningful. For the former question, unlike the latter, can only be answered on a case by case basis by specifying the rules that give the meaning of the relevant constant.

Option a) is only plausible if the meaning of any constant is given by *all* the rules we are disposed to follow with respect to the given constant.[123] However, this claim leads to problems when taken together with a). We may show this as follows. First, suppose that we introduce a connective 'tonk' into our language with respect to which we are disposed to follow the 'tonk' rules. Option a) then commits us to the claim that 'tonk' is meaningful. Moreover, given that we are also committed to claiming that the meaning of any constant is given by all the rules that we are disposed to follow with respect to the relevant constant, we must therefore conclude that the meaning of 'tonk' is given by both of the 'tonk' rules. However, this outcome is unacceptable and so the dispositionalist must therefore reject option a). If option c) is correct, though, then it is unclear how most individuals are able to know that their ordinary constants (e.g. 'and') are meaningful.[124] This is because few people (if any) know what the relevant technical constraints are. Given that we clearly *do* know that constants such as 'and' are meaningful, though, this leaves option b) as the only viable option. Importantly, this option is only plausible if the meaning of each constant is given solely by its intro rules. In other words, it is only plausible if Gentzen's thesis is true.[125]

---

[123] We may defend this claim in the following manner. First, let C be a constant and let R be the set of rules we are disposed to follow with respect to C. Now suppose for the sake of contradiction that a) is true but the meaning of C is *not* given by all the rules in R. Then there is a rule r in R that does not contribute to giving the meaning of C *even though* C is meaningful in virtue of the fact that we are disposed to follow r. However, this claim is implausible. For if a constant is meaningful then this is due entirely to the fact that it has some particular meaning. However, r does not contribute to giving the meaning of C and so our disposition to follow r is irrelevant to whether or not C is meaningful.

[124] It is worth emphasizing that I am talking here about our *explicit* knowledge that our constants are meaningful rather than the *implicit* knowledge that constitutes our understanding of the given constants. See Dummett (1978) for a discussion of the distinction between explicit and implicit knowledge.

[125] It may seem as though I have overlooked the possibility that a constant C is meaningful in virtue of the fact that there are *elimination* rules we are disposed to follow that pertain to C. I have omitted this possibility because it is only plausible if the meaning of any constant is given solely by its elim rules. This claim is only plausible, in turn, if we can devise a notion of harmony that lets us extract the intro rules that govern our constants from their elim rules. However, few attempts, if any, have been made to devise an account of harmony that would let us accomplish this task – nor is it clear how such an account would proceed.

It may seem as though we now have an argument for Gentzen's thesis and by extension a reason to adopt the harmony requirement described in 6). However, there is a problem with the argument just given. Specifically, it relies on the following assumption: that if c) is correct then we cannot know that 'and' is meaningful unless we know that its rules satisfy the relevant technical constraints. It is not clear, though, why we should accept this assumption. For it is not the case, generally speaking, that if p is true in virtue of q then we cannot know q without knowing p.[126] The proponent of the present argument (i.e. the argument for Gentzen's thesis) might concede this point but nevertheless claim that it is hard to see how we could know that 'and' is meaningful, if c) is true, other than by knowing that its rules satisfy the relevant technical constraints. I will therefore (briefly) sketch a potential answer to this worry. Let S be a sentence containing 'and' that we know is true (e.g. a tautology). Given that only meaningful sentences are true we may conclude that S is meaningful. If S is meaningful then so, presumably, are the words that it contains – including 'and' – and so we have therefore addressed the worry at hand.

The argument that was given in support of Gentzen's thesis faces further problems if we try to formulate a version of it that applies to so-called "normative inferentialism". On this view, the meaning of a logical constant is given by some of the rules that *govern* it rather than the rules that we are disposed to follow with respect to the constant at hand. In light of this difference, the phrase "given by" is construed differently too. Specifically, the claim that certain rules "give the meaning" of a constant C is characterized as follows: that C has the meaning it does in virtue of the fact that it is governed by the relevant rules.[127] In order to formulate a version of the argument for Gentzen's thesis that applies to the normative view, the normative inferentialist must first identify the available answers to the question of what makes a constant meaningful. They appear to consist of the following:

a') A constant C is meaningful in virtue of the fact that there are rules that govern C.

b') A constant C is meaningful in virtue of the fact that there are *intro* rules that govern C.

c') A constant C is meaningful in virtue of the fact that there are rules that govern C and that also satisfy certain technical constraints.

---

[126] For example, many people know that the liquid they drink is water even though they do not know that it is composed of H_2O. This is true even though the liquid is water in virtue of the fact that it is composed of H_2O.

[127] Gentzen's thesis must be understood accordingly.

Option a') is only plausible if the meaning of a constant is given by all the rules that govern it. Option b'), meanwhile, is only plausible if the meaning of a constant is given entirely by the *intro* rules that govern it. Just as in the dispositional case, normative inferentialists will therefore have an argument for Gentzen's thesis if they can rule out options a') and c'). Is option a') ruled out by the threat of 'tonk', though, in the same way as option a)? I do not believe that it is. For while we can clearly introduce a connective into our language with respect to which we are *disposed* to follow the 'tonk' rules, it is not at all obvious that we can introduce a connective that is *governed* by these rules. This is because we lack the control over normative facts that we have over our own dispositions.[128] If we cannot introduce such a connective, though, then we do not have an argument against a') that mirrors the argument that we gave against a). Consequently, the normative inferentialist has no obvious reason to endorse b') over a').

Section II – *The Purpose of Harmony.*

I have argued so far that inferentialists have no obvious reason to adopt a harmony requirement. Does the absence of a harmony requirement mean that harmony is an idle notion? I do not believe so. Rather, the purpose of devising a notion of harmony is to help us determine whether or not the meanings of our ordinary constants (i.e. $\wedge$, $\vee$, $\neg$, $\rightarrow$, $\forall$ and $\exists$) are (partly) given by their elim rules. For if we can devise an account of harmony that lets us extract the (standard) elim rules that govern our ordinary constants from their intro rules then we will have shown that the following view is tenable:

8) The elimination rules that govern our ordinary constants do not contribute to giving their meanings.

The tenability of 8) depends on our ability to develop an adequate account of harmony for the following reason: if 8) is true then we cannot simply identify the justificatory arguments in L with the arguments that are instances of the rules that give the meanings of the constants in L. There must therefore be some other criterion that determines when an argument is justificatory. If we can develop an account of harmony that lets us extract the standard elim rules that govern our

[128] This remark should not be taken to suggest that I believe we have full control over our own dispositions. For this is clearly false. However, while we have at least some control over them, there is no reason to believe that we have any such control over normative facts. This is the difference that I seek to highlight.

ordinary constants from their intro rules we will then have shown that such a criterion is available. This criterion consists of the following:

9) An argument is justificatory iff it is an instance of either a) one of the rules that give the meaning of a logical constant or b) an elim rule that is in harmony with the intro rules that (partly) give the meanings of the logical constants.[129]

On the other hand, if we cannot develop an adequate account of harmony then it is unclear how we can formulate a criterion that determines when an argument in L is justificatory. It is worth highlighting that 9) differs from 7) and does not motivate any kind of harmony requirement. This difference stems from the fact that 8) is weaker than Gentzen's thesis. Specifically, 8) is weaker than Gentzen's thesis because a) it only concerns our ordinary constants and b) it leaves open the possibility that a constant's meaning is given (partly) by rules that are neither intro nor elim rules (e.g. reductio ad absurdum).

I will now propose a constraint on any adequate account of harmony. I will call this constraint the "weak locality requirement":

10) An account of harmony satisfies the "*weak locality requirement"* if it lets us demonstrate that a rule is harmonious without assuming that any particular elim rule governs the constants in our language.[130]

An adequate account of harmony must meet this requirement in order to guarantee that we do not engage in circular reasoning when we use the account to determine whether or not an argument in our language is justificatory.[131] However,

---

[129] One might worry that this criterion is vulnerable to counterexamples such as a) arguments from "O is red" to "O is colored" as well as b) cases of warrant transmission failure of the kind discussed by Davies and Wright (2004). The latter include Moorean arguments that aim to refute skeptical hypotheses. The criterion given in 9), however, applies solely to mathematical arguments presented in a first order language such as that of set theory or arithmetic (i.e. arguments in which the premises and conclusion all consist of first order mathematical statements). Arguments such as that from "O is red" to "O is colored" therefore lie outside the scope of 9). Moreover, the epistemic factors that lead to warrant transmission failure in Wright's and Davies' examples do not arise in this setting.

[130] See the introductory paragraph for the definition of "governs".

[131] The locality requirement leaves open the possibility, though, that we may need to employ an elim rule e in order to demonstrate that e is harmonious. It is not clear, however, that this is problematic. For as Dummett (1974) highlights, there is an important difference between employing a particular rule in a demonstration that the rule is

if we try to devise an account of harmony that satisfies 10) we seem, at first, to encounter an issue when we consider connectives such as ⊃. In particular, the meaning of ⊃ is given (at least partly) by the following introduction rule (henceforth the "⊃ intro" rule):

$$\frac{\begin{array}{c}[A]\\ \vdots \\ B\end{array}}{A \supset B}\ \supset\text{I}$$

Put otherwise, the connective ⊃ has the meaning it does in virtue (partly) of the fact that we are disposed to perform inferences that are instances of the given rule. What does it mean, though, to say that an inference is an "instance" of the rule above? It seems to mean the following: that we can obtain the premises and conclusion of the inference by (uniformly) replacing 'A' and 'B' with particular sentences and replacing the dots with a proof of B from A. This definition of "instance" then raises the following question: namely, what do we mean by "proof"? If we characterize a "proof" as a derivation in which every step is an "instance" of a rule in our system we have then defined "instance" in terms of "proof" and vice versa. This circularity is not problematic if we *already* have a list of all the rules in our system – for we can then define the terms "instance" and "proof" via simultaneous induction. However, in order to define "proof" in this way, we must assume precisely which rules do and do not belong in our proof system. Such an assumption may at first appear to violate the locality requirement described in 10). For in making such an assumption we seem to be assuming that certain elimination rules govern our constants.

This reaction is too quick, however. For there is a difference between a) assuming that a rule governs a constant in the epistemic sense discussed earlier and b) assuming that we may appeal to a rule in a proof (I will call such rules "proof-admissible"). The latter assumption is logical rather than epistemic. It therefore does not violate the locality requirement described in 10). Nevertheless, if explaining the intro rule for ⊃ requires assuming the full range of rules in our proof system, then clearly harmony cannot serve as a criterion of proof-admissibility. This leaves us without any apparent criterion for it. To avoid this outcome, I will suggest

---

legitimate versus assuming the legitimacy of the rule as a premise in the given demonstration.

that we treat the notion of "proof" as conceptually primitive.[132] This will let us characterize the intro rule for $\supset$ without assuming all the rules in our system. Moreover, in order to let harmony provide us with a criterion for proof-admissibility, I will propose the following strengthened locality requirement:

11) An account of harmony satisfies the *"strong locality requirement"* if it lets us demonstrate that a rule is harmonious without assuming either a) that any particular elim rule governs the constants in our language or b) that any particular elim rule is proof-admissible.[133]

If our account of harmony satisfies this requirement then this opens the door to the following criterion for proof-admissibility: specifically, a rule is proof admissible if it either contributes to giving the meanings of our constants or is in harmony with the rules that do.

Section III – *Accounts of Harmony.*

Now that I have established the purpose of harmony, I will discuss existing accounts of the notion itself. Going forward, I will use the phrase "cut formula" to describe any formula in a deduction that is both the conclusion of an intro rule and the major premise of an elim rule. I will begin by examining Dummett's (1991, ch. 11) notion of "intrinsic harmony".[134] Let C be a connective in our language that is governed by a set of intro rules R and let e be an elim rule for C. Then,

12) The elim rule e is "intrinsically harmonious" if there is an effective procedure that lets us remove any cut formula in a deduction that has principal connective C and that is also the major premise of an application of e.

---

[132] The claim that a particular rule may be used in a proof will then become a substantive rather than a definitional claim.

[133] While harmony may help provide us with a criterion of proof-admissibility, it is important that we do not define "proof" in terms of "harmony". This is because any viable account of harmony appeals to the notion of "proof". This gives us further reason to treat the notion of "proof" as primitive.

[134] This account of harmony is inspired by Prawitz' (1965) inversion principle. Prawitz characterizes this principle as follows: "let α be an application of an elimination rule that has B as consequence. Then, deductions that satisfy the sufficient condition for deriving the major premise of α, when combined with deduction of the minor premises α, already "contain" a deduction of B; the deduction of B is thus obtainable directly from the given deductions without the addition of α" (p. 33).

Going forward, I will use the phrase "leveling procedure" as a label for any procedure of the kind just described.[135] We may use the example of the material conditional (i.e. the connective $\supset$) to illustrate the applicability of intrinsic harmony to the elim rules that govern our ordinary connectives. The intro and elim rules for $\supset$ consist of the following:

$$\dfrac{\begin{array}{c}[A]\\ \vdots \\ B\end{array}}{A \supset B}\ \supset\text{I} \qquad \dfrac{A \supset B \quad A}{B}\ \supset\text{E}$$

Now suppose a deduction D contains a cut formula with principal connective $\supset$ that is also the major premise of an application of $\supset$ elim. Then D contains a sub-deduction of the form shown on the left below and we may remove the cut formula in this sub-deduction by performing the re- arrangement shown on the right:

$$\dfrac{\dfrac{\begin{array}{c}A\,{}^{1}\\ \vdots\\ \delta_1\\ \vdots\\ B\end{array}}{A \supset B}\ \supset\text{I} \qquad \begin{array}{c}\vdots\\ \delta_2\\ \vdots\\ A\end{array}}{B}\ \supset\text{E} \qquad \begin{array}{c}\vdots\\ \delta_2\\ \vdots\\ A\\ \vdots\\ \delta_1\\ \vdots\\ B\end{array}$$

We have therefore shown that $\supset$ elim satisfies intrinsic harmony.

There is a problem, however, with using intrinsic harmony as an account of logical harmony. Specifically, as we saw in the case of $\supset$, in order to demonstrate that an elim rule satisfies intrinsic harmony we must sometimes know that it is possible to compose certain deductions. For example, in the case of $\supset$ we must know that if we have a deduction δ_2 of A with a set of undischarged premises Δ and a deduction δ_1 of B with undischarged premises $\Gamma \cup \{A\}$ it is then always possible to

[135] The existence of a leveling procedure for every logical constant in a language L does not guarantee that the proof system for L is normalizable. This is because normalization requires more than just the removal of cut formulae – rather, it also requires the removal of so-called cut segments (see Mancosu et al. (2021, ch. 4) for details). It also does not guarantee that the proof system for L is systematically conservative – where a deductive system for a language L is "systematically conservative" if every connective C in L yields a conservative extension of the fragment L–{C}. See Steinberger (2013) for a useful discussion of the relationships these between different proof-theoretic constraints.

obtain a deduction of B with undischarged premises $\Gamma \cup \Delta$ by grafting δ_2 onto δ_1. However, this is a claim that we cannot simply assume but must rather prove.[136] In order to establish it, we must use an inductive proof that requires us to assume all the rules in our system (see Mancosu et al. (2021, ch. 4) for a presentation of the relevant proof). It therefore follows that we cannot demonstrate that the elim rule that governs $\supset$ is intrinsically harmonious without knowing all the elim rules in our system. The notion of intrinsic harmony therefore does not satisfy the strong locality requirement described in 11). This gives us a reason to not identify intrinsic harmony with the notion of logical harmony.

Read (2010) devises an alternative account of harmony. To explain his account, let S be an arbitrary sentence with principal operator C. Going forward (and in keeping with Read), I will call a collection Γ of premises and/or sub-proofs a "direct ground for S" if we can infer S from Γ via the single application of one of the introduction rules for C. On Read's account, an elimination rule for C is harmonious provided that it lets us infer a sentence from S if and only if the given sentence follows from each direct ground for S. For example, the introduction rule for $\wedge$ consists of the following:

$$\frac{\alpha \qquad \beta}{\alpha \wedge \beta}\ \wedge\mathrm{I}$$

This is the only introduction rule for $\wedge$ and so a sentence of the form '$\alpha \wedge \beta$' has only one direct ground – namely, the collection of premises that includes α and β. According to Read, the (sole) harmonious elimination rule for $\wedge$ therefore consists of the following:

$$\frac{\alpha \wedge \beta \qquad \begin{array}{c}\overbrace{(\alpha)\quad(\beta)}\\ \vdots \\ \gamma\end{array}}{\gamma}\ \wedge\mathrm{E}$$

This rule lets us infer any sentence from $\alpha \wedge \beta$ that follows from the sole direct ground for $\alpha \wedge \beta$. Things become more complicated, though, when we consider the connective $\rightarrow$ (going forward, I will use '$\rightarrow$' and '$\supset$' interchangeably). The only

[136] One might wonder why such a proof is needed. For the claim may seem obvious enough to need no proof. The answer is that a rule that is correctly applied in a proof may cease to be correctly applied if we insert certain additional steps at the beginning of the relevant proof. This phenomenon may arise, for example, on account of the eigenvariable conditions for certain quantifier rules (e.g. the rules for $\forall$and $\exists$).

direct ground for a sentence of the form '$\alpha \rightarrow \beta$' consists of a derivation of β from α. If we wish to apply Read's account of harmony to $\rightarrow$ we must then say that the elim rule for $\rightarrow$ lets us infer a sentence γ from $\alpha \rightarrow \beta$ if and only if γ follows from any derivation of β from α. The obvious problem, however, is that it is not clear what it means to say that a sentence "follows" from such a derivation. Negri/Von Plato (2001) endorse Read's account of harmony and state that "the fact that γ can be derived from the existence of such a derivation can be expressed by – 'If γ follows from β, then it already follows from α'. This is achieved precisely by the elimination rule [below]" (p. 8).[137]

$$\frac{\alpha \rightarrow \beta \qquad \alpha \qquad \begin{array}{c}[\beta] \\ \vdots \\ \gamma\end{array}}{\gamma} \rightarrow\text{-E}$$

However, this characterization clearly requires us to strain the meaning of the word "follow". Read's account of harmony therefore lacks uniformity across different connectives. Moreover, there is another problem with his account. Specifically, the introduction rules for disjunction consist of the following:

$$\frac{\alpha}{\alpha \vee \beta} \vee\text{I} \qquad \frac{\beta}{\alpha \vee \beta} \vee\text{I}$$

A sentence of the form $\alpha \vee \beta$ therefore has two direct grounds – namely α and β. According to Read, the (sole) harmonious elimination rule therefore consists of the following:

$$\frac{\alpha \vee \beta \qquad \begin{array}{c}(\alpha) \\ \vdots \\ \gamma\end{array} \qquad \begin{array}{c}(\beta) \\ \vdots \\ \gamma\end{array}}{\gamma} \vee\text{E}$$

However, as Rumfitt (2017) observes, it is unclear on Read's account whether or not this elimination rule permits the use of collateral premises in the sub-derivations of γ. For if γ must follow solely from the direct grounds for $\alpha \vee \beta$ – as Read claims – then this requirement seems to prohibit the use of any such collateral

[137] Negri and Von Plato motivate their account of harmony by appealing to the following inversion principle: "Whatever follows from the direct grounds for deriving a proposition must follow from that proposition" (p. 6). This inversion principle differs from the one put forward by Prawitz. See Moriconi and Tesconi (2008) for an overview of different inversion principles.

premises. If the rule does not permit the use of such premises, though, then it is too weak.[138] In light of this issue, a different account of harmony is therefore required.

Weir (1986) develops an account of harmony that is designed to accommodate classical logic.[139] In keeping with his terminology, I will call a collection Π of premises and/or sub-proofs a "sufficient condition for S under R" – where S is a sentence and R is an intro rule – if R lets us infer S from Π. On Weir's account, a collection E of elimination rules for a constant C is said to be harmonious if it satisfies two conditions.[140] The first condition consists of the following.

> *Condition 1:* Let R be any one of the intro rules for C, let S be an arbitrary sentence with principal connective C and let Π be a sufficient condition for S under R. Then we can obtain another sufficient condition for S under R that depends on the same assumptions as Π by taking a sequence of proofs each of which is formed by i) deriving S from Π by R, ii) adding the discharged assumptions as minor premises next to S and then iii) applying a member of E.

We can illustrate this condition by considering the standard elim rules for conjunction.[141] Let Π be a sufficient condition for 'A$\wedge$B' under $\wedge$ intro that consists of a derivation of A from premises Δ together with a derivation of B from premises Γ. We can then show that the standard elim rules for conjunction satisfy condition 1 by constructing the following derivations:

$$\dfrac{\begin{array}{cc}\Delta & \Gamma \\ \vdots & \vdots \\ A & B\end{array}}{\dfrac{A \,\&\, B}{A}} \qquad \dfrac{\begin{array}{cc}\Delta & \Gamma \\ \vdots & \vdots \\ A & B\end{array}}{\dfrac{A \,\&\, B}{B}}$$

[138] The resulting rule is the elim rule of quantum logic. I will assume, however, that we do not wish to weaken the classical elim rule for $\vee$.

[139] While standard formulations of classical logic do not satisfy Weir's account of harmony, Weir devises a non-standard formulation of it that does.

[140] On Weir's account, harmony is not a trait of particular elimination rules. Rather, it is a trait that may apply to a *collection* of elimination rules.

[141] These elim rules let us infer φ from $\varphi\wedge\psi$ and ψ from $\varphi\wedge\psi$. The standard intro rule for $\wedge$, meanwhile, lets us infer $\varphi\wedge\psi$ from the combination of premises φ and ψ.

Both these derivations are constructed in the manner described in condition 1. Taken together, they provide us with a sufficient condition for 'A∧B' under ∧ intro and this sufficient condition depends on the same assumptions as Π (i.e. $\Delta \cup \Gamma$). We can similarly show that the standard elim rule for → satisfies condition 1 as well (this rule lets us infer B from A→B and A). In order to do so, let Π' be a sufficient condition for A→B under → intro where Π' consists of a derivation of B from {A}∪Δ.[142] Then we can show that the standard elim rule for → satisfies condition 1 via the following derivation:

$$\dfrac{\dfrac{\Delta \; [A]^{n} \\ \vdots \\ B}{A \rightarrow B}^{n} \quad A}{B}$$

This derivation is constructed in the manner described in condition 1 and is a sufficient condition for A→B under → intro. Moreover, it depends on the same premises as Π' (i.e. {A}∪Δ).

I will now explain Weir's second condition for harmony. This condition consists of the following – where C is a connective and E consists of a collection of elim rules for C.

> *Condition 2:* Let e be an arbitrary member of E, let S be an arbitrary sentence with principal connective C and let Φ be a collection of premises and/or sub-proofs to which we can apply e. Then there is a proof of S that consists of first i) a sequence of occurrences of Φ each of which is followed by e and then ii) the application of an introduction rule for C at which point all the minor premises in Φ are discharged.

We can illustrate this condition by considering Weir's rules for disjunction. His intro rules consist of the following:

[142] The → intro rule lets us infer A→B from a derivation of B from A.

$$\dfrac{\begin{array}{c}[\sim B]^{1}\\ \vdots\\ A\end{array}}{A \vee B}\,1 \qquad \dfrac{\begin{array}{c}[\sim A]^{1}\\ \vdots\\ B\end{array}}{A \vee B}\,1$$

His elim rules, meanwhile, consist of the following:

$$\dfrac{A \vee B \qquad \sim A}{B} \qquad \dfrac{A \vee B \qquad \sim B}{A}$$

Now let e be the elim rule for $\vee$ on the left and let Φ be a collection of premises to which we can apply e. Φ then includes the premises A$\vee$B and $\sim$A and so we can show that e satisfies condition 2 via the following derivation:

$$\dfrac{\dfrac{A \vee B \qquad \sim A^{1}}{B}}{A \vee B}\,1$$

A similar derivation lets us show that Weir's second elim rule for $\vee$ satisfies condition 2 as well.

There is a problem, however, with Weir's account of harmony. Specifically, let $\oplus$ be a connective with the following intro rules:

$$\dfrac{A}{A \oplus B} \qquad \dfrac{B}{A \oplus B}$$

As Murzi (2020) observes, the following collection of elim rules is then harmonious on Weir's account:

$$\dfrac{A \oplus B}{A} \qquad \dfrac{A \oplus B}{B}$$

Taken together, the intro and elim rules for $\oplus$ let us derive any claim from any other and therefore behave like the 'tonk' rules. This outcome is problematic for the following reason. Specifically, suppose we introduce a connective C the meaning of which is given by the intro rules for $\oplus$. If Weir's account of harmony is correct it then follows via 10) that C is governed by all the rules shown above. However, there is no connective that is governed by all these rules and so we must therefore reject Weir's account of harmony. Proponents of Weir's account might try to salvage it by

claiming that it is impossible to introduce a connective whose meaning is given by the intro rules for $\oplus$. Murzi, for example, considers requiring that the meaning of any connective be given by one intro rule alone. If we accept this claim, though, we must then deny that the meaning of $\vee$ is given by its standard intro rules (for these rules are identical to the intro rules for $\oplus$). Is this outcome plausible? If we accept it, it will then follow via 9) that one of the following is true: either, a) certain instances of the standard intro rules for $\vee$ are not justificatory arguments or, alternatively, b) the meaning of $\vee$ is given by rules that are distinct from the standard intro rules for $\vee$ but whose instances nevertheless include every instance of the standard intro rules. I will assume that option a) is implausible – for it seems evident that any inference from A to A$\vee$B or from B to A$\vee$B transmits justification. As for option b), meanwhile, it is not obvious that there are any rules that meet the relevant description that may plausibly be said to give the meaning of $\vee$. I will therefore conclude that we should reject Weir's account of harmony.[143]

Section IV – *Dummett's Account of Validity*.

Let us take stock. I have examined existing accounts of harmony and found them deficient. The deficiencies of each account highlight certain features that any successful account must possess. First, any successful account must satisfy the locality requirement described in 12). Second, any successful account must avoid deeming rules as harmonious that are either too strong or too weak. With these requirements in mind, I will now explain and assess the proof-theoretic account of logical validity that Dummett develops in chs. 11-13 of *The Logical Basis of Metaphysics*.[144] While I will reject this account, it will nevertheless serve as the basis for the account of harmony that I propose in the following section.

Dummett begins by endorsing Gentzen's thesis (i.e. the claim that the introduction rules alone give the meanings of the logical constants). He then develops a proof-theoretic account of validity that makes no use of elimination

[143] See Tennant (2020) for an additional account of harmony and Steinberger (2009) for a demonstration that the account yields implausibly strong rules for the existential quantifier.

[144] This account is inspired by but differs from the account developed by Prawitz (1978, 1985). It may be seen as a competitor to the more familiar model-theoretic account developed by Tarski (1936).

rules.[145] Dummett's definition of validity proceeds via simultaneous induction and is formulated in the setting of Gentzen-style natural deduction trees. For ease of exposition, though, I will instead use the equivalent formulation that Goldfarb (1998) provides in the setting of Fitch-style proofs (this formulation is restricted to propositional logic).[146] On this formulation, the terms 'valid' and 'canonical' are defined in the following way – where the term 'argument' denotes a Fitch-style deduction:[147]

(a) An argument A is *valid* iff there is an effective procedure that transforms any canonical argument C for the premises of A into a canonical argument C' for the conclusion of A – where C' has the same premises as C.

(b) An argument A is *canonical* iff i) all its premises are atomic, ii) it makes use solely of introduction and boundary rules except in its sub-arguments and iii) all of its sub-arguments are valid.

This definition requires some clarification. First, '$\perp$' is a zero-place connective rather than an atomic sentence and therefore cannot serve as a premise in a canonical argument. Second, boundary rules are rules that permit the inference of certain atomic sentences or '$\perp$' from other atomic sentences and/or '$\perp$' (e.g. a rule that lets us infer '$\perp$' from the combination of atomic sentences P and Q). Third, when determining whether or not a sub-argument is valid we should assume that the premises of the main argument are always included among the premises of the relevant sub-argument. Lastly, it is worth clarifying condition a). In order to do so, let A be an argument with premises Γ and conclusion C. Then for A to be valid it must satisfy the following criterion. Suppose that we have a canonical argument for each premise in Γ and let Δ consist of the collection of all the premises that are

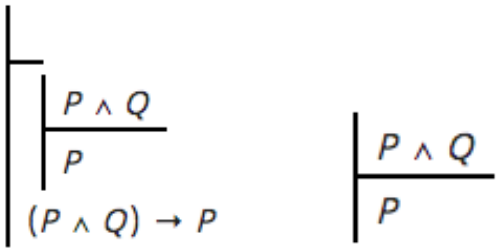
[145] Dummett argues that his account of validity is supported by a claim that he calls the "Fundamental Assumption". According to the Fundamental Assumption, any statement that is verifiable could be verified by obtaining a proof of the statement that ends in the use of an introduction rule for the statement's principal connective. I will not assess the Fundamental Assumption here though or the role that it plays in Dummett's account. See Kurbis (2012) for a discussion of these issues.

[146] Given that my focus in this chapter is not exegetical, I take no stand on whether or not Goldfarb's formulation is a faithful reconstruction of Dummett's account.

[147] As Goldfarb states, "the simultaneous induction works because discharge of premises increases logical complexity. Thus, whether a deduction with conclusion F is a valid canonical argument depends on the validity of arguments whose premises and conclusion are of strictly lesser logical complexity than F."

used in these arguments. Then we must be able to transform these arguments into a canonical argument for C whose premises are all in Δ.

We may illustrate the difference between canonical and non-canonical arguments with the following example. The argument on the left below is a canonical argument while the argument on the right is not:



The argument on the left is canonical because its sub-proof is valid and the argument only makes use of an introduction rule outside of its sub-proof – namely, the introduction rule for '$\rightarrow$'.[148] The argument on the right is *not* canonical because the conclusion does not follow from the argument's premise via introduction and boundary rules alone. It follows from Dummett's definition of validity that all canonical arguments are valid. It is worth highlighting that on Dummett's account the validity of an argument depends solely on its premises and conclusion (i.e. the intervening steps of an argument do not affect its validity). Lastly, Dummett defines $\ulcorner \neg P \urcorner$ as $\ulcorner P \rightarrow \bot \urcorner$. The connective '$\bot$' has no introduction rule but has an elimination rule that lets us infer anything from it.

Dummett's account may be used to demonstrate the validity of various familiar rules of inference (I will say that a rule is "valid" if every instance of the rule is a valid argument).[149] It trivially follows from Dummett's account that all introduction rules are valid. I will now show, for the sake of illustration, how we may use Dummett's account to demonstrate the validity of the $\rightarrow$ elim rule. The $\rightarrow$ intro rule is shown on the left below while the $\rightarrow$ elim rule is shown on the right:

[148] The introduction rule for $\rightarrow$ lets us infer $\ulcorner \varphi \rightarrow \psi \urcorner$ from any valid sub-proof that has premise φ and conclusion ψ.

[149] Certain principles of classical logic such as the rule of double negation elimination are not valid on Dummett's account though.

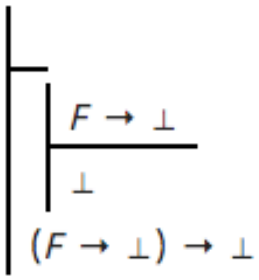


Let A be an instance of the rule on the right. Then A is an argument with premises ⌜ F→G ⌝ and F and conclusion G – where F and G are formulas. Now let C be a canonical argument for ⌜ F→G ⌝ with premises Γ and let C' be a canonical argument for F with premises Δ. In order to show that A is valid we must show that we can turn C and C' into a canonical argument for G with premises $\Gamma \cup \Delta$. It follows from the definition of "canonical" that C ends in the use of $\rightarrow$ intro. This in turn implies that C contains a valid sub-argument S with premises $\Gamma \cup \{F\}$ and conclusion G.[150] Suppose now that we take the collection Ω of canonical arguments for $\Gamma \cup \{F\}$ that consists of the following: the argument C' and an argument for each member of Γ that consists in each case of nothing more than a simple premise introduction. The collection of premises used in Ω then consists of $\Gamma \cup \Delta$. Given that the sub-argument S with premises $\Gamma \cup \{F\}$ and conclusion G is valid it follows that we can turn the arguments in Ω into a canonical argument for G with premises $\Gamma \cup \Delta$. We have therefore shown that we can turn C and C' into a canonical argument for G with premises $\Gamma \cup \Delta$. This implies that A is valid. Given that A was an arbitrary instance of $\rightarrow$ elim it follows that $\rightarrow$ elim is a valid rule.

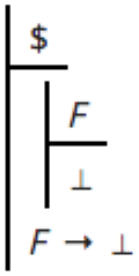
Goldfarb (1998) shows that Dummett's account of validity faces various counterexamples in a language that lacks boundary rules. First, he shows that if there are no boundary rules then any argument with no premises and conclusion ⌜ ¬¬F ⌝ is valid on Dummett's account if F does not contain '⊥'. Second, he shows that if there are no boundary rules then any argument with the premise ⌜ F→G ⌝ and conclusion G is valid on Dummett's account if F does not contain '⊥' *and* G does not have any atomic sentences in common with F. Rather than rehearse the full details of Goldfarb's reasoning I will instead discuss the particular case in which an argument has no premises and conclusion ⌜ ¬¬F ⌝ – where F is an atomic sentence and is therefore distinct from '⊥'. In order to show that this argument is valid we must show that there is a premise-free canonical argument for ⌜ ¬¬F ⌝. It follows from Dummett's account of negation that we ought to interpret ⌜ ¬¬F ⌝ as

[150] It is worth remembering that the premises of C are automatically included among the premises of any of its sub-arguments.

⌜ (F→⊥)→⊥ ⌝. Let us therefore consider the argument for ⌜ ¬¬F ⌝ that has no premises, a sub-proof S with premise ⌜ F→⊥ ⌝ and conclusion '⊥', and final conclusion ⌜ (F→⊥)→⊥ ⌝. This argument is shown below (I will call it 'B'):



This argument only uses an introduction rule outside of its sub-proof. Therefore, if we can show that the sub-proof in the argument is valid we will have shown that the argument is canonical and by extension that the argument with no premises and conclusion ⌜ ¬¬F ⌝ is valid. Given that our language contains no boundary rules we may show that the sub-proof S is valid via the following reasoning. Let C be a canonical argument for ⌜ F→⊥ ⌝. It follows from the definition of "canonical" that C has the following form (where $ is a collection of atomic premises):



Given that our language contains no boundary rules it follows that the sub-argument in C is invalid. This is because there is no canonical proof of '⊥'. We therefore cannot transform every collection of canonical arguments for $∪{F} into a canonical argument for '⊥' with the same exact premises. It follows that there is no canonical proof of ⌜ F→⊥ ⌝ and so it is trivially true that the sub-argument S in B is valid. This in turn implies that B is a premise-free canonical proof of ⌜ (F→⊥)→⊥ ⌝ and therefore of ⌜ ¬¬F ⌝.

The inclusion of boundary rules in our language will not let Dummett avoid every counter-example of the kind just described. This is because our language may clearly contain an atomic sentence F where no boundary rule permits the derivation of '⊥' from F. If this is the case, then there will again be a premise free canonical argument with conclusion ⌜ ¬¬F ⌝. However, Goldfarb observes that Dummett may avoid the counterexamples described so far if he revises his account of validity in the following manner: by claiming that an argument is valid if and only

if it is valid in the original sense (i.e. the sense described in (a) and (b)) given *any* choice of boundary rules. Going forward, I will call this added requirement the "boundary choice" requirement and will call an argument valid* if it is valid in Dummett's original sense (i.e. the sense defined at the start of this section).

However, this revised account of validity still faces a problem that Goldfarb does not mention. Specifically, the elimination rule for '⊥' appears invalid. We may show this as follows. Let P, Q and R be distinct atomic sentences and let us suppose that our language contains a boundary rule that permits the derivation of '⊥' from P and Q but does *not* contain a boundary rule that permits the derivation of R from P and Q. There is then a canonical argument for '⊥' with premises P and Q. Given that the elimination rule for '⊥' permits the inference of any sentence from '⊥' it therefore permits the inference of R from '⊥'. However, given our choice of boundary rules, there is no procedure that lets us transform the canonical argument with premises P and Q and conclusion '⊥' into a canonical argument with the same premises and conclusion R. The '⊥' elimination rule is therefore invalid on the revised account of validity.[151] In order to avoid this problem, Goldfarb must impose the following constraint on any choice of boundary rules – henceforth, the "coherency constraint". Specifically, let B be a choice of boundary rules and let us suppose that some rule in B permits the inference of '⊥' from a collection of atomic sentences Γ. Then for any atomic sentence A, there is a rule in B that permits the inference of A from Γ.

Goldfarb observes that Dummett may instead avoid the given counterexamples by imposing a schematic requirement on his original account of validity. On this revised account, an argument is valid if and only if it remains valid* whenever we uniformly substitute different formulas in place of the sentence letters in the given argument. Going forward, I will call arguments that meet this description "schematically valid*". It is worth highlighting that even if Dummett adopts this approach he must still require that the boundary rules in any language satisfy the coherency constraint. For there will otherwise be languages in which the ⊥ elim rule is invalid for the reasons discussed in the previous paragraph. Can Dummett solve this issue by treating ⊥ elim as a boundary rule instead? The answer is no. For in order to treat ⊥ elim as a boundary rule we must lift the requirement that boundary rules only permit the inference of atomic sentences. However, if we

[151] It is worth highlighting that this remains the case even if we treat ⊥ elim as a boundary rule. For even if ⊥ elim is a boundary rule there will nonetheless be a choice of boundary rules that does not contain '⊥' elim but contains a rule that lets us infer '⊥' from P and Q.

relax this requirement we will then be unable to show that familiar rules such as $\wedge$ elim are valid. This is because any demonstration of the rule's validity relies on the assumption that any canonical argument for '$p \wedge q$' ends in the use of the $\wedge$ intro rule. However, if there are boundary rules that permit the inference of complex sentences (as $\perp$ elim does) then this is no longer an assumption that we can make.[152]

While the modified versions of Dummett's account avoid Goldfarb's counterexamples, there is a problem that afflicts them both. Specifically, both accounts render most arguments invalid that make use of boundary rules. For such arguments will usually not be schematically valid*. Nor will they be valid* given any choice of boundary rules. There is clearly a sense, however, in which arguments that make use of boundary rules are valid (e.g. the argument with premise "O is red" and conclusion "O is colored"). We might call this sense "general validity" and distinguish it from the narrower notion of "logical validity". Arguments that make use of boundary rules will then be valid in the general but not the logical sense. While we might wish to treat one of the modified versions of Dummett's account as an account of logical validity, it is not obvious, though, how we can generalize either account into an account of general validity. For we cannot simply lift the schematic or boundary choice requirement and identify general validity with validity* (i.e. Dummett's original account of validity). This is because doing so would require us to treat Goldfarb's original counterexamples as generally valid in any language that does not contain boundary rules. However, there are clearly arguments with no premises and conclusion $\ulcorner \neg\neg F \urcorner$ – where F is an atomic sentence – that are neither generally nor logically valid.

I will conclude my discussion of Dummett's account with a technical observation by Goldfarb. This observation will be relevant to my discussion of harmony in the following section. Let L be a propositional language that contains no boundary rules. If α is a collection of atomic sentences and φ is a formula in L then let "$\alpha \Vdash \varphi$" stand for the claim that there is a canonical proof of φ with premises α.[153] The following four claims are then true – where α and β are arbitrary collections of atomic sentences in L while F and G are arbitrary sentences in L:

i) $\alpha \Vdash F \wedge G$ if and only if $\alpha \Vdash F$ and $\alpha \Vdash G$.

[152] Dummett proposes, in fact, that we posit boundary rules of this kind (see ch. 12). While I am not against this proposal, it poses a problem for his account of validity for the reason that I just described.

[153] I am using Dummett's definition of "canonical" (i.e. as given in (b)).

ii) $\alpha \Vdash F \vee G$ if and only if $\alpha \Vdash F$ or $\alpha \Vdash G$.

iii) $\alpha \Vdash F \rightarrow G$ if and only if $\forall \beta$(if $\alpha \subseteq \beta$ and $\beta \Vdash F$, then $\beta \Vdash G$).

iv) $\alpha \Vdash \neg F$ if and only if $\forall \beta$(if $\alpha \subseteq \beta$ then not $\beta \Vdash F$).

Claims i) and ii) follow immediately from the definition of "canonical" as well as the fact that L contains no boundary rules. We may establish the left to right direction of iii) via the following reasoning. First, suppose that $\alpha \Vdash F \rightarrow G$. Any canonical proof of $F \rightarrow G$ in a language that contains no boundary rules must end in the use of the introduction rule for $\rightarrow$. It follows that there is a canonical proof of $F \rightarrow G$ that contains a valid* sub-proof with conclusion G and premises $\alpha \cup \{F\}$. Now let Γ be an arbitrary collection of canonical arguments for $\alpha \cup \{F\}$ and let Δ be the collection of all the premises that are used in the arguments in Γ. It follows that we can turn the canonical arguments in Γ into a canonical proof of G whose premises are all in Δ. Now let β be an arbitrary collection of atomic sentences such that $\alpha \subseteq \beta$ and $\beta \Vdash F$. Then there is clearly a collection Σ of canonical arguments for $\alpha \cup \{F\}$ such that β contains all the premises used in Σ. It follows that we can turn the arguments in Σ into a canonical argument for G whose premises are all in β in which case $\beta \Vdash G$. This concludes the demonstration of the left to right direction of iii). To establish the right to left direction, suppose that $\forall \beta$(if $\alpha \subseteq \beta$ and $\beta \Vdash F$, then $\beta \Vdash G$). Let us now consider the argument with premises $\alpha \cup \{F\}$ and conclusion G. Let Γ be an arbitrary non-empty collection of canonical arguments for $\alpha \cup \{F\}$ and let Δ be the collection of all the premises used in Γ. Then Γ contains a canonical argument for F with premises Π. Seeing as $\Pi \subseteq \Delta$ it follows that there exists a canonical proof of F with premises Δ in which case $\Delta \Vdash F$. Given that L contains no boundary rules it is also true that $\alpha \subseteq \Delta$.[154] We may therefore infer via our initial supposition that $\Delta \Vdash G$. We have now shown that the argument with premises $\alpha \cup \{F\}$ and conclusion G is valid*. This in turn implies that there is a canonical proof of $F \rightarrow G$ with premises α and so we have therefore established the right to left direction of iii). Similar reasoning lets us establish iv).[155]

Goldfarb observes that claims i) to iv) may be read as stating the conditions under which a (complex) sentence in L is true at a node in the following Kripke tree: namely, the tree $\mathbb{K}$ in which every node is a collection of atomic sentences in L and

[154] This is because if L contains no boundary rules then any canonical proof of an atomic sentence must contain that same atomic sentence among its premises.

[155] To establish iv), it helps to keep in mind that it is never the case that $\alpha \Vdash \bot$.

the accessibility relation between the nodes is the subset relation.[156] On this construal of i) to iv), the sentence "$\alpha \Vdash \varphi$" may be read as stating that φ is true at the node in K that consists of the set of atomic sentences α. As Goldfarb shows, we may use this observation to demonstrate that if L is a language that contains no boundary rules then any argument in L that is schematically valid* is also classically valid (i.e. valid according to the rules of classical logic). In order to demonstrate this result, we may prove the converse claim. First, let the pair $<\Gamma, \varphi>$ be a classically *in*valid argument – where Γ is a collection of premises and φ is the argument's conclusion. Then there is a two-valued interpretation V of the sentence letters in Γ and φ that makes all the sentences in Γ true but the sentence φ false. Let us substitute the sentence '$p \rightarrow p$' in place of every sentence letter in Γ and φ to which V assigns T and the constant '$\bot$' in place of every sentence letter in Γ and φ to which V assigns F. Let us call the resulting set of premises Γ' and the resulting conclusion φ'. Then in order to evaluate whether or not a member of $\Gamma' \cup \{\varphi'\}$ is true at a node in K we need only compute the truth-value of the given member according to the classical truth-tables.[157] Given that every sentence in Γ' is true and $\varnothing$ is a node in K it follows that $\varnothing \Vdash S$ for every $S \in \Gamma'$. However, φ' is false and so it is *not* the case that $\varnothing \Vdash \varphi'$. The argument with premises Γ' and conclusion φ' is therefore not valid*. This in turn implies that the argument with premises Γ and conclusion φ is not schematically valid*.

[156] A Kripke tree consists of a pair $\langle W, \leq \rangle$ where W is a set of nodes and the accessibility relation $\leq$ is a partial order on W (i.e. a reflexive, transitive and asymmetric relation). A Kripke frame, meanwhile, is a triple $\langle W, \leq, @ \rangle$ where @ is a distinguished member of W (i.e. its "actual" node). Lastly, a Kripke model is a quadruple $\langle W, \leq, @, V \rangle$ where V is a function that assigns a set of atomic sentence letters to each member of W. These models are used to devise a model-theoretic semantics for intuitionistic logic. See MacFarlane (2021, ch. 6) for a presentation of this material.

[157] We may prove this via induction on the complexity of formulas. Specifically, we may prove the following claim. Suppose K is a Kripke tree and φ is a formula. Now suppose that we uniformly substitute '$p \rightarrow p$' and '$\bot$' in place of the sentence letters in φ. Then for any node in K, φ is true at the given node if and only if φ is true according to the classical truth-tables.

Section V – *Harmony Revisited.*

I will now use the foregoing discussion of Dummett's account of validity in order to devise an account of logical harmony. To achieve this aim, I will formulate two requirements that I call "weak harmony" and "stability". I will then characterize an elim rule as harmonious if and only if it satisfies both requirements. The condition of weak harmony is meant to ensure that an elim rule is not too strong while the condition of stability is meant to ensure that it is not too weak.[158] Going forward, I will continue to say that an argument is valid* if it is "valid" in Dummett's original sense (i.e. the sense described at the beginning of the previous section). I will now propose that we adapt Dummett's account of validity into an account of weak harmony in the following way:

13) An elim rule is "weakly harmonious" if all its instances are valid* in a language that contains the relevant intro rules but no boundary rules.

This account restricts the form of weakly harmonious elim rules to those that do not employ sub-proofs as premises. This is because it does not make any sense to claim that an instance of a rule that employs sub-proofs as premises is valid*. This is because Dummett's account of validity states that an argument is valid iff any collection of canonical arguments for the premises can be transformed into a canonical argument for the conclusion. However, if one of the premises consist of a sub-proof then it does not make sense to speak of a canonical argument for it.[159]

We must therefore adopt Weir's elim rules for disjunction in place of the standard elim rule.[160] The qualification in 13) regarding boundary rules serves two

---

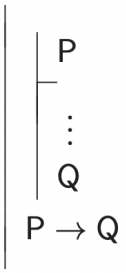
[158] Whether an elim rule is too strong or too weak will obviously depend on the introduction rules that our system contains.

[159] This observation has the following consequence: that on Dummett's account of validity (or at least on Goldfarb's reconstruction) it does not make sense to describe the standard intro rule for $\rightarrow$ as valid given that any instance of this rule has a premise that consists of a sub-proof. However, it is not clear that this outcome poses any problem for Dummett's account of validity. For Dummett's account applies to arguments rather than rules. Moreover, this outcome does not prevent him for claiming that the standard intro rule for $\rightarrow$ gives the meaning of $\rightarrow$. Nor does it prevent him for claiming that the rule preserves truth.

[160] We may show that Weir's elim rules are weakly harmonious via the following reasoning. Let us consider the rule that lets us infer Q from P$\vee$Q and $\neg$P. Suppose that we have a canonical argument C for P$\vee$Q and a canonical argument C' for $\neg$P. Given that our language contains no boundary rules it follows that C ends in the use of $\vee$ intro and C'

purposes. First, it ensures that the question of whether or not an elim rule is harmonious does not depend on the rules that govern our language's non-logical vocabulary. Second, the qualification lets us appeal to the technical result that was described at the end of the previous section. This result ensures that the harmonious rules that govern our ordinary connectives do not let us derive any non-classical theorems. This is because the harmonious rules are all weakly harmonious on my account of harmony. Moreover, any instance of a weakly harmonious rule is schematically valid* and, as we have already seen, is therefore classically valid as well.

The account of weak harmony provided in 13) requires a further qualification. To highlight the necessary qualification, we should first observe that the standard intro rule for $\rightarrow$ consists of the following (I will continue to work in the setting of Fitch style proofs):



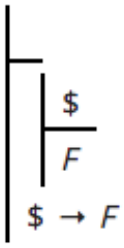
This rule lets us infer P$\rightarrow$Q from a proof of Q from P – where a proof of Q from P consists of a step by step derivation that may or may not employ elim rules. However, Dummett's account of validity makes no mention of elim rules. Consequently, his intro rule for $\rightarrow$ is formulated differently. Rather than letting us infer P$\rightarrow$Q from a proof of Q from P, Dummett's version of the rule lets us infer P$\rightarrow$Q from a valid* sub-argument with premise P and conclusion Q. It is a consequence of Dummett's account, however, that the validity* of any such sub-argument does not depend on the intervening steps between P and Q. There is therefore a difference between the standard intro rule for $\rightarrow$ and the rule that Dummett's account employs. A similar point applies to any intro rule that employs

---

ends in the use of $\neg$ intro. This in turn implies that C contains either a canonical proof of P or a canonical proof of Q. C', meanwhile, contains a valid sub-argument with conclusion '$\perp$'. Suppose now that C contains a canonical proof of P. Given that C' contains a valid sub-argument with conclusion '$\perp$' it follows that we can obtain a canonical proof of '$\perp$'. However, there is no canonical proof of '$\perp$' and we must therefore reject our supposition that C contains a canonical proof of P. It follows that C contains a canonical proof of Q in which case we can easily turn C into a canonical proof of Q with the same premises as C. We have therefore shown that Weir's elim rule is valid*.

sub-proofs as premises (i.e. Dummett's version of any such rule employs valid* sub-arguments in place of sub-proofs). I will therefore call Dummett's version of any such rule its "Dummettian analogue". Importantly, I will assume that the meanings of our connectives are given (partly) by their standard intro rules (i.e. those that employ sub-proofs as premises) rather than their Dummettian analogues. In order to determine whether or not an elim rule is weakly harmonious in the sense described in 13) we must therefore check whether or not every instance of the given rule is valid* in a language that i) contains no boundary rules but ii) contains the Dummettian analogues of the intro rules that (partly) give the meanings of the connectives in the relevant language.

One might wonder why I have assumed that the meanings of our connectives are given partly by their standard intro rules rather than their Dummettian analogues. For my qualification of 13) is only needed if we make this assumption. I will therefore illustrate my reason for making this claim with the following example. Let F consists of Fermat's last theorem and let $ consist of the conjunction of all the ZF Axioms that are needed to prove F. Now let us suppose for the sake of argument that the meaning of $\rightarrow$ is given partly by the Dummettian analogue of its standard intro rule (I will call this rule the "$\rightarrow$ intro$_D$ rule"). It then follows that we can establish the conclusion $\$\rightarrow F$ via the following argument:



In other words, we can establish $\$\rightarrow F$ by simply inferring it from a sub-argument that has premises $, conclusion F and no intervening steps. We can establish $\$\rightarrow F$ in this way because the argument above is an instance of $\rightarrow$ intro$_D$. Moreover, we can then infer F from the combination of $\$\rightarrow F$ and $ via modus ponens. Given that $\rightarrow$ is governed by $\rightarrow$ intro$_D$ as well as modus ponens it follows that we can establish (and therefore justify believing) Fermat's last theorem via the aforementioned reasoning. However, this is clearly false and so we should therefore reject the assumption that the meaning of $\rightarrow$ is partly given by $\rightarrow$ intro$_D$.

One might try to resist this argument by claiming that we can only infer a conclusion from a sub-argument if we know that the sub-argument is valid*. For if this is true, then it will no longer seem outlandish to claim that we can establish Fermat's last theorem via the reasoning described above. There is a problem,

however, with this proposal. Specifically, it is not clear how we can acquire knowledge of the fact that a sub-argument is valid* other than by deducing this result from Dummett's account of validity. However, doing so will require us to use Dummett's version of the $\forall$ intro rule and this rule employs a sub-argument as a premise.[161] It follows, on the current proposal, that in order to use $\forall$ intro as a means of establishing the validity* of the original sub-argument we will need to know that this new sub-argument is valid* as well. This observation gives rise to an infinite regress and so it follows that we can never use $\forall$ intro as a means of establishing new claims. This result is implausible and so we should therefore reject the proposal at hand. Consequently, I will maintain that the meanings of our connectives are partly given by their standard intro rules rather than their Dummettian analogues.

As I mentioned earlier, in order to provide a full account of logical harmony I will supplement the requirement of weak harmony with a requirement called "stability". While the requirement of weak harmony is meant to ensure that an elim rule is not too strong, the requirement of stability is meant to ensure that it is not too weak. I will formulate this requirement as follows:[162]

14) Let C be a connective, let e be an elim rule for C and let S be an arbitrary sentence with principal connective C. Then e is stable if and only if any application of e to S gives us back part or all of a direct ground for S.[163]

The standard elim rules for $\wedge$, for example, are both stable given that the application of either rule to $A \wedge B$ gives us one half of a direct ground for $A \wedge B$. The standard elim rule for $\rightarrow$, meanwhile, is also stable given that the application of $\rightarrow$ elim to $A \rightarrow B$ yields a derivation of B from A with collateral premise $A \rightarrow B$. This derivation is itself a direct ground for $A \rightarrow B$.

I will now explain the ways in which my account of harmony avoids the pitfalls faced by previous accounts. First, unlike the notion of intrinsic harmony, my account satisfies the strong locality requirement described in 11). This is because

---

[161] We will need to use the $\forall$ intro rule in order to show that *every* canonical argument for the premise(s) of the sub-argument can be transformed into a canonical argument for the sub-argument's conclusion.

[162] It is worth recalling that a direct ground for a sentence with principal connective C is a collection of premises and/or sub-proofs from which we can infer S via one of the intro rules for C.

[163] This requirement resembles but is not as demanding as Weir's first condition for harmony. Unlike Weir's condition, it applies to particular rules rather than collections of rules.

in order to show that a rule is weakly harmonious we need only show that it is valid* in a language that contains the relevant intro rules and no boundary rules. Importantly, we can prove such results without making any assumptions about the elim rules in our system. This is because Dummett's account of validity makes no use of elimination rules. Therefore, while we may still need to devise an inductive proof in order to show that certain derivations can be grafted onto one another, devising this proof will only require us to assume the intro rules in our system. Unlike Read's account, my account of harmony does not treat the quantum elim rule for $\vee$ as harmonious. It does, however, treat Weir's elim rules for $\vee$ as harmonious and these rules are strong enough to let us derive the standard elim rule for $\vee$.[164] Lastly, unlike on Weir's account, the elim rules for $\oplus$ are not harmonious. This is because they do not satisfy the weak harmony requirement.

Conclusion

I have devised an account of harmony that lets us extract the elim rules that govern our ordinary connectives from their intro rules. This gives us a reason to believe that the elim rules that govern our ordinary connectives do not contribute to giving their meanings. While certain classical rules such as double negation elimination are not harmonious on my account, this observation gives us no reason to revise classical logic unless we adopt a harmony requirement. This is something, however, that I have argued we have no reason to do. It remains to be seen whether or not my account of harmony can be extended to handle the rules for our familiar quantifiers. If so, this will give us reason to believe that their elimination rules do not contribute to giving their meanings either.

[164] The standard rule lets us infer a sentence S from the combination of P$\vee$Q together with a sub-derivation of S from P and a sub-derivation of S from Q. These sub-derivations may employ collateral premises.

Concluding Remarks

In this dissertation, I have sought to accomplish several things. First, I have defended a view of metaphysical explanation that treats some such explanations as derivational and gives center stage to the notion of ground. I then used this view to develop an epistemology for non-transparent grounding claims and connected this epistemology to the logic of ground. Specifically, I argued that certain rules in the logic of ground play an important role in enabling knowledge of non-transparent grounding claims.

I then applied the framework developed in the first chapter to the inferentialist program in the philosophy of logic. Specifically, I highlighted an explanatory challenge that inferentialists have overlooked – namely, that of explaining why our connectives obey their familiar truth tables. I called this challenge the "truth table challenge". I then developed a solution to this challenge in which I derived the truth tables from premises concerning truth as well as the referents of the logical constants. Some of these premises consist of grounding claims and so my solution has the explanatory structure discussed in the previous chapter. Moreover, I argued for my solution via the kind of inference to the best explanation that I defended in chapter one. Finally, I used my solution to motivate a reference based constraint on the rules that may give the meaning of a connective. This constraint rules out 'tonk' and thereby gives us a principled response to Prior's original challenge. It also gives us reason to reject the unrestricted inferentialism recently endorsed by Warren. However, like Warren, I treat the meanings of our familiar connectives as given by all their classical rules.

In the third chapter, I examined the proof-theoretic constraint of logical harmony. While formulated in different ways, this constraint is often used by inferentialists to support the revision of classical logic in favor of intuitionistic logic (an outcome that would pose problems for the view defended in chapter two). I argued that while we have no convincing reason to impose a harmony constraint, the notion of harmony nevertheless retains a valuable role. Specifically, it helps us determine whether or not a connective's meaning is partly given by its elimination rule(s). I then developed an account of harmony that is designed to serve this purpose. The account builds on Dummett's conception of logical validity and avoids the pitfalls that I identified in existing accounts.

This dissertation is unified by the explanatory lens through which it approaches a range of logical questions. In emphasizing this lens, the dissertation highlights the inadequacy of purely formalistic approaches to questions in the

philosophy of logic. This outcome should not be surprising given that explanatory considerations play an important role in guiding theory choice across almost every other philosophical area. To name but a few examples, such considerations are invoked when defending scientific realism in the philosophy of science (on the basis that realism best explains scientific success), mind-brain identity theories in the philosophy of mind and forms of rationalism in different areas of epistemology. There is no reason to think that the philosophy of logic should be any different. Moreover, as we have seen in the case of Prior's challenge, purely formalistic approaches often lead to significant under-determination. Specifically, it is unclear from a purely formalistic vantage point whether we should address the puzzle of 'tonk' via a conservativity constraint, a harmony constraint, a referential constraint or something else entirely. Explanatory considerations therefore provide an important means by which to narrow the space of competitive theories. As in science, so in the philosophy of logic.

This dissertation also highlights two directions for future research. First, while I have argued that certain metaphysical explanations have a derivational structure, I have not provided a systematic account of the features that determine the comparative strength of such explanations. Pursuing this task will require attention to the details of metaphysical practice as well as to the precedents set by other domains (e.g. science and mathematics). Second, this dissertation applies an explanatory approach to a domain in which this approach has rarely been taken (i.e. the philosophy of logic). In doing so, it raises the following question: in which other domains are explanatory concerns generally overlooked? Exploring this question may lead us to reevaluate the way in which we approach familiar philosophical issues.

References

Audi, P. (2012). “Grounding: Toward a Theory of the “In-Virtue-Of” Relation”, *The Journal of Philosophy*, 109(12), pp. 685–711.

Bates, J. (2009), “A Defense of the Explanatory Argument for Physicalism”, *The Philosophical Quarterly*, 59(235), pp. 315-324.

Baron, Sam, and James Norton (2019), “Metaphysical Explanation: The Kitcher Picture”, *Erkenntnis*, 86(1), pp. 187-207.

Belnap, N. (1962). “Tonk, Plonk and Plink.” *Analysis,* 22(6), pp. 130-134.

Belnap, N., and Massey, G. (1990). “Semantic Holism”, *Studia Logica: An International Journal for Symbolic Logic*, vol. 49, no. 1, pp. 67–82.

Bliss, Ricki and Kelly Trogdon (2021), “Metaphysical Grounding”, *The Stanford Encyclopedia of Philosophy.*

Boghossian, P. (1996). “Analyticity Reconsidered”, *Nous*, 30(3), pp. 360-391.

Boghossian, P. (2003). “Blind Reasoning”, *Supplement to the Proceedings of the Aristotelian Society*, 77(1), pp. 225–248.

Bonnay, Denis & Westerståhl, Dag. (2016). “Compositionality Solves Carnap’s Problem”, *Erkenntnis*, 81(4), pp. 721-739.

Burgess, J. (1984). “Dummett's Case for Intuitionism”, *History and Philosophy of Logic*, 5(2), pp. 177-194

Carnap, R. (1937). *The Logical Syntax of Language*. Translated by Amethe Smeaton, Open Court Publishing Company, 2002.

Carnap, R. (1943). *Formalization of Logic*. Harvard University Press, 1968.

Cook, R. (2005). “What’s Wrong with Tonk(?)”, *Journal of Philosophical Logic*, 34(2), pp. 217-226.

Correia, F. (2010), “Grounding and Truth-Functions”, *Logique et Analyse*, 53(211), pp. 251–279.

Dasgupta, S. (2014), “The Possibility of Physicalism”, *Journal of Philosophy*, 111(9), pp. 557–592.

Dasgupta, S. (2016). “Metaphysical Rationalism”, *Noûs*, 50(2), pp. 379-418.

Davies, Martin and Crispin Wright (2004), “On Epistemic Entitlement”, *Aristotelian Society Supplementary Volume*, 78, pp. 167-245.

Dummett, M. (1968). “The Reality of the Past”, *Proceedings of the Aristotelian Society*, vol. 69, pp. 239–58.

Dummett, M. (1974). “The Justification of Deduction”, reprinted in *Truth and Other Enigmas*. Cambridge, MA: Harvard University Press, 1978.

Dummett, M. (1975). “The Philosophical Basis of Intuitionistic Logic.” Chapter 14. In *Truth and Other Enigmas*. Cambridge, MA: Harvard University Press.

Dummett, M., (1978), “What do I Know when I Know a Language,” reprinted in *The Seas of Language*. Oxford University Press, 1993.

Dummett, M. (1982). “Realism”, *Synthese*, 52(1), pp. 55–112.

Dummett, M., *The Logical Basis of Metaphysics*, Harvard University Press, Cambridge, Mass., 1991.

Dummett, M. “What is a Theory of Meaning? (II).” *The Seas of Language*, Oxford University Press, 1993.

Dummett, M. (2002). “‘Yes’, ‘No’ and ‘Can’t Say’.” *Mind*, 111(442), pp. 289-295.

Epstein, Peter and Umrao Sethi (2025), “An Empirical Stalemate: Why Science Fails to Settle a Central Philosophical Debate about Perception”, *Philosophy and Phenomenological Research*, 112(1), pp. 276-289.

Ferreira, F. (2008). “The Co-ordination Principles: A Problem for Bilateralism”, *Mind*, vol. 117, no. 468, pp. 1051–1057.

Field, H. (1994). “Deflationist Views of Meaning and Content”, *Mind*, 103(411), pp. 249–285.

Field, H. (2006). “Compositional Principles vs Schematic Reasoning”, *The Monist*, 89(1), pp. 9–27.

Fine, K. (2010), “Some Puzzles of Ground”, *Notre Dame Journal of Formal Logic*, 51(1), pp. 97-118.

Fine, K. (2012a), “Guide to Ground”, in Fabrice Correia & Benjamin Schnieder, *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge University Press, pp. 37-80.

Fine, K. (2012b), “The Pure Logic of Ground”, *The Review of Symbolic Logic*, 5(1), pp. 1–25.

Fitch, F. B. (1963). “A Logical Analysis of Some Value Concepts”, *The Journal of Symbolic Logic*, 28(2), pp. 135–142.

Frege, G. (1892). "On Sense and Reference." *Translations from the Philosophical Writings of Gottlob Frege*. Trans. Max Black. Eds. Peter Geach and Max Black. Oxford: Basil Blackwell, 1952. 56–78.

Gentzen, G., (1934). “Untersuchungen über das logische Schließen. I & II”, *Mathematische Zeitschrift*, 39: 176–210 and 405–431. English translation “Investigations into Logical Deduction”, *American Philosophical Quarterly*, 1: 288–306 (1964), and 2: 204–218 (1965).

Goldfarb, W. (1998). “On Dummett's ‘Proof-Theoretic Justifications of Logical Laws.’” *Advances in Proof-Theoretic Semantics*, edited by Thomas Piecha and Peter Schroeder-Heister, vol. 43, Springer, 2016, pp. 195–211.

Hand, M. (1999). “Antirealism and Falsity.” *What Is Negation?*, edited by Dov M. Gabbay and Heinrich Wansing, vol. 13, pp. 185–198.

Harris, J. H. (1982). “What’s so Logical about the “Logical” Axioms?” *Studia Logica*, 41(2), pp. 159-171.

Hempel, Carl G. and Paul Oppenheim (1948), “Studies in the Logic of Explanation”, *Philosophy of Science*, 15(2), pp. 135–175.

Hjortland, O. T. (2014). ”Speech Acts, Categoricity, and the Meanings of Logical Connectives”, *Notre Dame Journal of Formal Logic*, 55(4), pp. 445-467.

Horwich, P. (1982). “Three Forms of Realism”, *Synthese*, 51(2), pp. 181–201.
Kim, J. (2005), *Physicalism, or Something Near Enough*, Princeton University Press.

Kitcher, P. (1981), “Explanatory Unification”, *Philosophy of Science*, 48(4), pp. 507-531.

Kripke, S. (1975). “Outline of a Theory of Truth”, *Journal of Philosophy*, 72(19), pp. 690-716.

Kripke, S. (1980). *Naming and Necessity: Lectures Given to the Princeton University Philosophy Colloquium*. Cambridge, MA: Harvard University Press.

Kurbis, N. (2012). “How Fundamental is the Fundamental Assumption?” *Teorema: Revista Internacional de Filosofa*, 31(2), pp. 5-19.

Kürbis, N. (2016). “Some Comments on Ian Rumfitt's Bilateralism”, *Journal of Philosophical Logic*, vol. 45, no. 6, pp. 623–644.

Kürbis, N. (2017). “Bilateralism: Negations, Implications and some Observations and Problems about Hypotheses.” In J. Fichot & T. Piecha (eds.), Beyond Logic. Proceedings of the Conference held in Cerisy-la-Salle, 22-27 May 2017.

Kürbis, N. (2023). “Supposition: A Problem for Bilateralism”, *Bulletin of the Section of Logic*, 53(3), pp. 301-327.

Lange, M. (2014), “Aspects of Mathematical Explanation”, *Philosophical Review*, 123(4), pp. 485-531.

Lange, M. (2015), “Depth and Explanation in Mathematics”, *Philosophia Mathematica*, 23(2), pp. 196-214.

Lewis, D. (1986), “Causal Explanation”, in *Philosophical Papers, Volume II*. Oxford University Press. pp. 214-240.

Lipton, P. (2004), *Inference to the Best Explanation*, (2nd ed.), Routledge.

Litland, J. (2017), “Grounding Ground”, in *Oxford Studies in Metaphysics*, *Volume 10*, Karen Bennett and Dean W. Zimmerman (eds.), Oxford University Press, pp. 270-316.

Loux, M. J. “Realism and Anti-Realism: Dummett’s Challenge.” *The Oxford Handbook of Metaphysics*, edited by Dean Zimmerman and Michael J. Loux, Oxford University Press, 2003, pp. 633–655.

MacFarlane, J. *Philosophical Logic: A Contemporary Introduction*. Routledge, 2021.

Maddy, P. (1988), “Believing the Axioms I”, *The Journal of Symbolic Logic*, 53(2), pp. 481-511.

Mancosu, P. (2001), “Mathematical Explanation: Problems and Prospects”, *Topoi*, 20(1), pp. 97-117.

Mancosu, P. (2008), “Mathematical Explanation: Why it Matters”, in P. Mancosu (ed.), *The Philosophy of Mathematical Practice*, Oxford University Press, pp. 134-150.

Mancosu, Paolo, et al. *An Introduction to Proof Theory: Normalization, Cut-Elimination, and Consistency Proofs*. Oxford University Press, 2021.

Maurin, A.-S. (2019), “Grounding and Metaphysical Explanation: It’s Complicated”, *Philosophical Studies*, 176(6), pp. 1573-1594.

McLaughlin, B. (2010), “Consciousness, Type Physicalism, and Inference to the Best Explanation”, *Philosophical Issues*, 20(1), pp. 266-304.

Miller, A. (2002). “What is the Manifestation Argument?” *Pacific Philosophical Quarterly*, 83(4), pp. 352-383.

Milne, P., (2015). “Inversion Principles and Introduction Rules.” In: Wansing H. (ed.) *Dag Prawitz on Proofs and Meaning*. first ed. Outstanding Contributions to Logic, 7. Cham, Heidelberg, New York, Dordrecht, London: Springer, pp. 189-224.

Moriconi, E. and Tesconi, L. (2008). “On Inversion Principles”, *History and Philosophy of Logic*, 29(2), 103-113.

Murzi, J., & Hjortland, O. T. (2009). “Inferentialism and the Categoricity Problem: Reply to Raatikainen”, *Analysis*, 69(3), pp. 480–488.

Murzi, J., & Steinberger, F. (2015). “Inferentialism.” Blackwell Companion to Philosophy of Language. Ed. B. Hale, C. Wright, and A. Miller. Wiley Blackwell, 2017. 197-224.

Murzi, J. (2020). “Classical Harmony and Separability”, *Erkenntnis*, 85, pp. 391-415.

Murzi, J., & Topey, B. (2021). “Categoricity by Convention”, *Philosophical Studies*, 178(10), pp. 3391-3420.

Negri, S. and von Plato, J. (2001). *Structural Proof Theory*. Cambridge University.

Pautz, A. (2013), “The Real Trouble for Phenomenal Externalists: New Empirical Evidence.” In R. Brown (ed.), *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*, Springer, pp. 237-289.

Peacocke, C. (1987). “Understanding Logical Constants: A Realist’s Account”, *Proceedings of the British Academy*, 73: 153–200.

Peacocke, C. (1993). “How are A Priori Truths Possible?” *European Journal of Philosophy*, 1(2), pp. 175–199.

Prawitz, D., *Natural Deduction: A Proof-Theoretical Study*, Dover, Mineola, N. Y., 1965.

Prawitz, D. (1978). “Proofs and the Meaning and Completeness of the Logical Constants.” *Essays on Mathematical and Philosophical Logic*, edited by J. Hintikka et al., vol. 122, Springer. Synthese Library (Studies in Epistemology, Logic, Methodology and Philosophy of Science).

Prawitz, D. (1985) “Remarks on Some Approaches to the Concept of Logical Consequence”, *Synthese*, 62(2), pp. 153-71.

Priest, G. (2005). *Doubt Truth to be a Liar*. Oxford University Press.

Prior, A. N. (1960). “The Runabout Inference-Ticket”, *Analysis*, 21(2), pp. 38–39.

Raatikainen, P. (2004). “Conceptions of Truth in Intuitionism”, *History and Philosophy of Logic*, 25(2), pp. 131-145.

Raatikainen, P. (2008). “On Rules of Inference and the Meanings of Logical Constants”, *Analysis*, 68(4), pp. 282–287.

Raven, M. J. (2015), “Ground”, *Philosophy Compass*, 10(5), pp. 322–333.

Read, S. (2010). “General-Elimination Harmony and the Meaning of the Logical Constants”, *Journal of Philosophical Logic*, 39:557–76.

Rosen, G. (2010), “Metaphysical Dependence: Grounding and Reduction”, in Modality: Metaphysics, Logic, and Epistemology, Bob Hale and Aviv Hoffman (eds.), Oxford University Press, pp. 109–136.

Rubenstein, E. (2024). “Two Approaches to Metaphysical Explanation”, *Noûs*, 58(4), pp. 1107-1136.

Rumfitt, I. (2000). “‘Yes’ and ‘No.’”, *Mind,* 109(436), pp. 781–824.

Rumfitt, I. (2002). “Unilateralism Disarmed: A Reply to Dummett and Gibbard”, *Mind*, 111(442), pp. 305-321.

Rumfitt, I. (2017). “Against Harmony”, chapter 10 of volume I of *A Companion to the Philosophy of Language*, edited by B. Hale, C. Wright, and A. Miller, Wiley Blackwell, pp. 225-249.

Smiley, T. (1996). “Rejection”, *Analysis,* 56(1), pp. 1–9.

Russell, B. (1907), “The Regressive Method of Discovering the Premises of Mathematics”, in Bertrand Russell, *Essays in Analysis*, London: Allen and Unwin, 1973, pp. 272-283.

Salmon, W. C. (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton University Press.

Schaffer, J. (2009), “On What Grounds What”, in *Meta-metaphysics: New Essays on the Foundations of Ontology*, Chalmers, David, David Manley, and Ryan Wasserman (eds.), Oxford University Press, pp. 347–383.

Schaffer, J. (2010), “Monism: The Priority of the Whole”, *Philosophical Review*, 119(1), pp. 31–76.

Schaffer, J. (2016), “Grounding in the Image of Causation”, *Philosophical Studies*, 173(1), pp. 49–100.

Sellars, W. (1954). “Some Reflections on Language Games”, *Philosophy of Science,* 21(3), pp. 204-228.

Sider, T. (2020), “Ground Grounded”, *Philosophical Studies*, 177(3), pp. 747–767.

Siscoe, R. (2021), “Grounding and A Priori Epistemology: Challenges for Conceptualism”, *Synthese*, 199(3), pp. 11445-11463.

Skiles, Alexander and Kelly Trogdon (2021), “Should Explanation Be a Guide to Ground?”, *Philosophical Studies*, 178(12), pp. 4083–4098.

Soames, S. (1998). *Understanding Truth*. Oxford University Press.

Steinberger, F. (2009). “Not So Stable”, *Analysis*, 69(4), pp. 655-661.

Steinberger, F. (2011). “What Harmony Could and Could Not Be,” *Australasian Journal of Philosophy*, 89(4), pp. 617–639.

Steinberger, F. (2011). “Why Conclusions Should Remain Single”, *Journal of Philosophical Logic*, 40(3), pp. 333–55.

Steinberger, F. (2013). “On the Equivalence Conjecture for Proof-Theoretic Harmony”, *Notre Dame Journal of Formal Logic*, 54(1), pp. 79-86.

Steiner, M. (1978), “Mathematical Explanation”, *Philosophical Studies*, 34, pp. 135–151.

Strawson, P. F. (1950). “On Referring”, *Mind*, 59(235), pp. 320–344.

Tarski, A. (1936). “On the Concept of Following Logically,” *History and Philosophy of Logic,* 23(3), pp. 155-196.

Tennant, N. (1999). “Negation, Absurdity and Contrariety,” in *Negation*, ed. Dov M. Gabbay and Heinrich Wansing, Amsterdam: Kluwer: 199-222.

Tennant, N. (2020). “Inferentialism, Logicism, Harmony, and a Counterpoint,” in *Logic, Language, and Mathematics: Themes from the Philosophy of Crispin Wright*, edited by A. Miller, Oxford University Press, pp. 223-248.

Thompson, N. (2016), “Grounding and Metaphysical Explanation”, *Proceedings of the Aristotelian Society*, 116(3), pp. 395-402.

Valle-Inclan, Pedro del. (2024). “Carnap’s Problem, Definability and Compositionality.” *Journal of Philosophical Logic*, 53(5), pp. 1321-1346.

Van Fraassen, B. C. (1980), *The Scientific Image*, Oxford University Press.

Van Inwagen, P. (1981). “Why I Don’t Understand Substitutional Quantification”, *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 39(3), pp. 281–285.

Warren, J. (2015a). “Truth by Convention”, *Philosophical Quarterly*, 65(258), pp. 84-93.

Warren, J. (2015b). “Talking with Tonkers”, *Philosophers’ Imprint*, 15(24), pp. 1-24.

Warren, J. (2020). *Shadows of Syntax: Revitalizing Logical and Mathematical Conventionalism*. Oxford University Press.

Weir, A. (1986). “Classical Harmony”, *Notre Dame Journal of Formal Logic*, 27(4), 459–482.

Williamson, T. (2007), *The Philosophy of Philosophy*, Blackwell Publishing.

Wilsch, T. (2015), “The Nomological Account of Ground”, *Philosophical Studies*, 172(12), pp. 3293-3312.

Wilson, J. (2014), “No Work for a Theory of Grounding”, *Inquiry*, 57(5), pp. 535–579.

Wittgenstein, L., (1922). *Tractatus Logico-Philosophicus*, Trans. C. K. Ogden. London: Routledge and Kegan Paul.

Woodward, James and Christopher Hitchcock (2003), “Explanatory Generalizations, Part I: A Counterfactual Account”, *Noûs*, 37(1), pp. 1–24.