Taming the Impossible

by

Matthias Jenny

B.A., University of Zurich (2010)
M.A., Brandeis University (2012)

Submitted to the Department of Linguistics and Philosophy
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2017

Author . . . .
Department of Linguistics and Philosophy
September 8, 2017

Certified by . . . .
Vann McGee
Professor of Philosophy
Thesis Supervisor

Accepted by . . . .
Roger White
Professor of Philosophy
Chair of the Committee on Graduate Students

Taming the Impossible

by

Matthias Jenny

Submitted to the Department of Linguistics and Philosophy
on September 8, 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Philosophy

Abstract

The semantic paradoxes and other statements about impossibilities have proved to be obstacles to a satisfactory theory of conditionals. In my dissertation, which consists of two parts, I propose a new approach to the impossible that yields an improved theory of conditionals.

A prominent response to the semantic paradoxes is *glut theory*. Glut theorists avoid paradox by giving up material modus ponens. But they argue that they can help themselves to this rule in areas where no paradoxes loom. In chapter 1, I argue that this does not work and that giving up modus ponens in paradoxical domains leaves glut theorists with a weak logic everywhere.

One option that's available to glut theorists involves pragmatic innovation. In chapter 2, I explore the consequences of giving glut theorists the pragmatic resources that are already available the proponents of *gap theory*, the dual of glut theory. The resulting hybrid theory, which makes use of two distinct speech acts of assertion, is *glap theory*. Surprisingly, the logic of glap theory is a quite strong logic that adds to the logics of glut and gap theory two hybrid forms of modus ponens.

Turning to counterfactual conditionals, the second half of my dissertation concerns the *vacuity thesis*, which says that all counterpossible conditionals are vacuously true. In chapter 3, I argue that the strongest case against the vacuity thesis comes from counterpossibles as they appear in relative computability theory. I show that relative computability theorists crucially invoke counterpossibles when they define the central notions of their theory. I also provide a model theory for a quantified language that can express such counterpossibles.

The logical properties of counterfactuals about relative computability deserve closer attention. In chapter 4, I provide an axiomatization of a propositional fragment of the model theory developed in chapter 3 and prove that the axiomatization is complete and that the resulting conditional logic is decidable. This logic display some surprising features. While validating modus ponens, it also contains a restricted form of the import-export law.

Thesis Supervisor: Vann McGee
Title: Professor of Philosophy

Acknowledgments

I am incredibly lucky to have had Vann McGee as my thesis advisor. I got a first glimpse of his generosity and dedication when I submitted a draft of my final paper for the first class I took with him: to my amazement, Vann's comments were longer than the paper itself! This set the tone for our working relationship for the years to come. Between regular meetings that spanned whole afternoons, long written comments on numerous drafts, and countless excellent lectures and seminar meetings, Vann has been my greatest intellectual influence. Without him, this thesis truly wouldn't have been possible—trust me, I'm an expert on the impossible. Aside from the personal debt I owe him, Vann's research is a constant source of inspiration. I will always try to emulate the combination of philosophical depth and technical skill displayed in his papers and books.

I am also truly grateful to the other members of my thesis committee, all of whom have made me a better philosopher through their mentorship, teaching, and writing. Agustín Rayo has been so encouraging during the past five years and he intervened in crucial moments to prevent me from going astray. Bob Stalnaker's theories were the initial spark for much of my dissertation and his mentorship helped me guide along the way. Steve Yablo often drew my attention to ideas that always changed the direction of my thinking for the better.

My departmental colleagues have taught me so much and they listened patiently to my departmental talks and read my drafts. Thanks to Marion Boulicault, Sylvain Bromberger, Alex Byrne, Jennifer Carr, Nilanjan Das, Kevin Dorst, Lyndal Grant, Sally Haslanger, Sam Hesni, Richard Holton, Sophie Horowitz, Abby Jaques, Brendan de Kenessey, Justin Khoo, Rae Langton, Julia Markovits, Rebecca Millsop, Daniel Munoz, Sofia Ortiz-Hinojosa, Kevin Richardson, Damien Rochford, Said Saillant, Bernhard Salow, Ginger Schultheis, Melissa Schumacher, Brad Skow, Jack Spencer, Judy Thomson, Ian Wells, Quinn White, and Roger White. And special thanks to my office mates David Boylan, Cosmo Grant, Ari Koslow, Rose Lenehan, Matt Mandelkern, and Milo Phillips-Brown for making my daily life more entertaining. Philosophy at MIT is more than the sum of its parts, but these people have made my five years here particularly enjoyable.

In addition to my committee and my departmental colleagues, I am grateful to the philosophers and logicans at the University of Connecticut. UConn always felt like an intellectual home away from my MIT home. Thanks to Jc Beall, Jared Henderson, Nathan Kellen, Suzy Killmister, Andrew Parisi, David Ripley, Marcus Rossberg, Lionel Shapiro, Stewart Shapiro, Andrew Tedder, and Sam Wheeler. And thanks to my teachers at Brandeis, especially Eli Hirsch, Sophia Malamud, Berislav Marušić, and Palle Yourgrau, for putting me on the right path and for their continued mentorship.

Thanks also to Danielle Bromwich, Teresa Bruno, Noel Dominguez, Nykki Dular, Amy Ferrer, Mary Grenham, Sally Haslanger, Abby Jaques, Ari Koslow, Dana Miranda, Milo Phillips-Brown, Kevin Richardson, Lisa Rivera, Hayley Webster, and everyone else involved in the first two installments of PIKSI-Boston, especially our amazing Alain Locke Fellows, for giving me hope that our

discipline can be better than it has been in the past.

The members of the administrative staff at MIT's Department of Linguistics and Philosophy do an amazing job keeping everything running smoothly. Thanks to Chrissy Graham, Mary Grenham, Chris Naylor, Jen Purdy, and Matt Sikorski for immediately jumping into action whenever I had a problem. Thanks also to the staff at MIT's International Students Office, especially David Elwell.

Many thanks to my friends and family for their love, friendship, and unwavering support over the years: Abby, Blake, Can, Diego, Dustin, Kevin, Lukas, Marco, Martha, Matt, Migge, Milo, Nykki, and my parents, Ursula and Ulrich, and my sister, Mirjam. Finally, thanks to Vann for bringing his canine friend Cherry with him to his office and to class almost every day. Cherry, you've made Vann's lectures on logic less stressful and my hours spent in his office even more enjoyable. Please don't tell Vann you're the real reason why I asked him to be my advisor.

Contents

Part II: Taming the Undecidable   **74**

Introduction

Here are some claims about conditionals that, from a pretheoretical perspective, are hard to deny:

Identity is a law of logic. This is the claim that every sentence of the form

$$\phi \to \phi$$

is a law of logic. If it's raining, then it's raining. If that's not a law of logic, then it's unclear what would be.

Modus ponens is a rule of logic. This is the claim that every instance of

$$\frac{\phi \to \psi \qquad \phi}{\psi}$$

is logically valid. If it's raining, then the streets are wet. It is raining. So the streets are wet. Again, if that's not a rule of logic, then it's unclear what would be.

Not all counterpossibles have the same truth value. Counterpossibles are counterfactual conditionals with impossible antecedents. Here's a counterpossible that seems to be true:

If water had been an element, then water splitting would have been impossible.

But if that is true, then the following is surely false:

If water had been an element, then water splitting would have been *possible*.

So there seem to be some counterpossibles that are true and others that are false.

Despite the status of these claims as truisms, there are popular logics that deny them. Kleene's Strong 3-valued Logic ($K3$), which is an ingredient in the *gap theorist's* response to the liar paradox, doesn't contain identiy as a law of logic. The Logic of Paradox (LP), which is the dual of $K3$ and is invoked in the *glut theorist's* response to the liar, doesn't validate modus ponens. And the Stalnaker-Lewis logic of counterfactuals (C) treats all counterpossibles as vacuously true.

All of these logics attempt to tame the impossible in one way or another. The liar sentence is the sentence that says of itself that it is not true. It seems to describe an impossible scenario. $K3$ tames the liar sentence by giving up the *law of excluded middle*:

$$\phi \vee \neg\phi$$

But because the material conditional $\rightarrow$ is defined so that $\phi \rightarrow \psi$ abbreviates $\neg\phi \vee \psi$, giving up excluded middle means giving up identity as well.

LP tames the liar sentence by giving up the *rule of explosion*:

$$\frac{\phi \qquad \neg\phi}{\psi}$$

But again, given the definition of the material conditional and an uncontroversial background theory, this means that they also give up (material) modus ponens.

The logic C of counterfactuals defines a semantics for counterfactuals that involves a comparative similarity relation among possible worlds. Unable to make sense of comparative similarity among possible and *im*possible worlds, C indiscriminately declares all counterpossibles as true.

$K3$, LP, and C have greately enhanced our understanding of the liar paradox and of counterfactuals. But the ways in which they tame the impossible can be improved. This dissertation, which consists of two parts, proposes two new logics, one designed to respond to the liar paradox and one designed to accommodate a class of non-vacuous counterpossibles.

In response to the fact that $K3$ and LP need to give up the law of identity and the rule of modus ponens, resprctively, gap and glut theorists argue that they can nevertheless help themselves to these principles in areas where no paradoxes loom. This is undoubtedly true for gap theorists, who can simply assert those instances of the law of identity that they regard as true. But it is less obviously true for glut theorists, since we've learned from Lewis Carroll that rules aren't the kinds of things we can simply assert. Glut theorists have offered ways around this. In chapter 1, the first half of which appeared in *Thought* 6(1):43–53 (2017), I investigate the most prominent proposals, such as those involving shriek rules, conversational implicatures, and a supposed use of 'not' as a force indicator. I argue that none of these proposals work and that glut theorists are thus unable to be

selective in their use of modus ponens in public demonstrations of proofs. Thus, giving up this rule in paradoxical domains leaves them with a weak logic everywhere.

One option that's available to glut theorists involves pragmatic innovation. In chapter 2, I explore the consequences of giving glut theorists the pragmatic resources that are already available the proponents of *gap theory*, the dual of glut theory. The resulting hybrid theory, which makes use of two distinct speech acts of assertionm, one taken from gap theory and one taken from glut theory, is *glap theory*. Surprisingly, the logic of these two types of assertion isn't simply the logic of gap theory and the logic of glut theory taken together. Rather, it's a stronger logic that adds to the other two logics two hybrid forms of modus ponens. Moreover, the logic of glap theory allows for an adequate, fully structural treatment of the semantic paradoxes and the paradoxes of vagueness. I argue that glap theory strikes the right balance between strength, simplicity, and adequacy in the face of paradox.

Turning to counterfactuals, in chapter 3, a version of which is forthcoming in *Noûs*, I argue that the strongest case against the vacuity thesis comes from counterpossibles as they appear in a mathematical sub-discipline called relative computability theory. These are counterfactuals such as 'If the halting problem were algorithmically decidable, then the validity problem of the predicate calculus would also be algorithmically decidable.' Such counterfactuals are often found in introductory remarks about relative computability. This invites the suggestion that they don't need to be taken fully literally. In response to this, I show that relative computability theorists crucially invoke such counterfactuals when they define the central notions of their theory, and I argue that alternative ways of defining these notions fail. But we can't just rest content with rejecting the vacuity thesis without offering something in its place. I patch up the orthodox account of counterfactuals by providing a model theory for a quantified language that can express the above counterpossibles and many more. The result is a language that can express informative facts about what would have held if certain impossible things had held.

The logical properties of counterfactuals about relative computability deserve closer attention. In chapter 4, I provide an axiomatization of a propositional fragment of the model theory developed in chapter 3 and prove that the axiomatization is complete. This means that claims about relative

11

computability that can be expressed in this propositional language are true just in case they are consequences of my axioms. I also show that the resulting *Conditional Logic of Turing Reducibility* is decidable. This means that we have a fully general procedure for determining the truth or falsity of a wide class of counterpossible claims about relative computability. This conditional logic displays some surprising features. While validating modus ponens, it also contains a restricted form of the import-export law $\big(\phi \mathbin{\square\!\!\rightarrow} (\psi \mathbin{\square\!\!\rightarrow} \chi)\big) \leftrightarrow \big((\phi \wedge \psi) \mathbin{\square\!\!\rightarrow} \chi\big)$, which, in its unrestricted form, is famously incompatible with modus ponens.

Although chapter 2 builds on chapter 1 and chapter 4 on chapter 3, all chapters are designed to be read on their own. This results in some unavoidable redundancies.

Part I

Taming the Paradoxes

Chapter 1

The Flight from Gluts

1.1 Introduction

Gottlob Frege famously held that "nothing is added to [a] thought by ... ascribing to it the property of truth" (1956, 293). This idea is commonly expressed with the slogan that truth is *transparent*: ϕ and $Tr(\ulcorner\phi\urcorner)$[1]—the sentence that says that ϕ is true—are fully intersubstitutable in extensional contexts. Unfortunately, in classical logic, the law of excluded middle, i.e. $\vdash \phi \vee \neg\phi$, and the rule of explosion, i.e. $\phi, \neg\phi \vdash \psi$, allow us to derive any sentence from the liar sentence if we have transparency. It's tempting to put the blame on transparency here. However, it isn't entirely obvious what to replace transparency with.[2] That is why a number of authors have instead blamed classical logic. Saul Kripke (1975), Robert Martin and Peter Woodruff (1975), Peter Woodruff (1984), and Bradley Dowden (1984) have shown how we can preserve transparency if we dispense with either excluded middle or explosion. The most conservative deviation from classical logic that gives up excluded middle is Kleene's Strong 3-valued Logic ($K3$), and the most conservative non-explosive logic is the Logic of Paradox (LP).[3] Model theoretically, both of these logics are *three-valued* logics: they introduce a third truth value in addition to the classical truth values of truth and falsity. Proponents of $K3$ are called *gap theorists* because they interpret the third truth value as a truth

---

[1] '$\ulcorner\phi\urcorner$' is a term for ϕ in the object language. Note that the corner quotes here are Gödel quotes, not Quine quotes. In section 1.4, I will occasionally use corner quotes as Quine quotes. I will let context distinguish between the two uses.

[2] See McGee (1990) and Halbach (2011) surveys of some of the options.

[3] See Kleene (1950) for $K3$ and Asenjo (1966) and Priest (1979) for LP.

value gap—neither true nor false—while proponents of LP are called *glut theorists*, because they interpret the third truth value as a truth value glut—both true *and* false.

However, the move to such subclassical logics has serious drawbacks. In giving up modes of reasoning that are central to classical logic, we are left with logics where "nothing like sustained ordinary reasoning can be carried on," to quote Solomon Feferman (1984, 95).[4] As a partial remedy for this, Hartry Field (2008) and Jc Beall (2009) have recently devised stronger logics that build on $K3$ and LP, respectively, but that are still weak enough to be compatible with transparency.

Despite their virtues, Field's and Beall's logics have significant drawbacks of their own. Most notably, they are vastly more complex than classical logic, $K3$, and LP. For the latter three logics, we have algorithms for deciding whether a sentence of propositional logic is classically valid. By contrast, the task of deciding whether a sentence is valid in Field's logic, which has been more thoroughly studied than Beall's, is vastly more complex even than deciding whether a sentence of the language of arithmetic is true, the latter of which is already far beyond what's humanly possible.[5] So while Field's logic may come close to classical logic in terms of strength, it is doubtful whether it meets Feferman's challenge of being able to allow for sustained ordinary reasoning if it is humanly impossible to determine whether a given sentence is a theorem of the logic or not.[6]

More may be said on behalf of Field (2008), and also on behalf of Beall (2009), but the foregoing suggests that attempting to find a logic that's stronger than $K3$ or LP that meets both transparency and Feferman's challenge remains elusive. It may therefore be worth taking another look at $K3$ and LP to see how much we can get out of them. This is the topic of the present chapter, where I set aside the more sophisticated subclassical logics.

Recently, Beall (2011, 2013b, 2015a) has taken some first steps towards determining how much we can get out of $K3$ and LP. Following Gilbert Harman (1986, ch. 2), Beall distinguishes between *logic* and *reasoning*. Certain inferences may be justified by the standards of reasoning, even though they don't involve moving from premises to a conclusion that logically follows from the

---

[4]At the time, Feferman was talking about $K3$, but he (2012, 190) has since observed that the same holds for LP.

[5]As McGee (2010, 430-31) shows by building on work due to Welch (2008), the set of valid sentences in Field's logic is complete Π^1_2.

[6]It may be worried that we're being too demanding here. After all, classical first-order logic is also undecidable. However, validity in classical first-order logic, unlike Field's logic, is at least semi-decidable, and thus axiomatizable, since it is complete Σ^0_1.

premises. And, in fact, Beall argues, in most domains of inquiry, applications of excluded middle and explosion are perfectly legitimate, even though they are *logically* invalid according to either $K3$ or LP. Therefore, goes Beall, if we're reasoning about a classical domain, we may be confident that reliance on classical modes of reasoning won't lead us astray.

In this chapter, I argue that glut theorists have a much harder time recapturing classical reasoning in select domains than gap theorists. I do so by investigating the feasibility of three prominent proposals of how glut theorists can recapture classicality, namely by way of shriek rules (subsection 1.3.1), conversational implicatures (subsection 1.3.2), and metalinguistic negation (section 1.4). What's attractive about these proposals is that they only appeal to ordinary, well-understood resources. If glut theorists are able to recapture classical reasoning using only such resources, then they may come close to meeting Feferman's challenge. Of course, there will remain areas where classical reasoning is inappropriate, but such is the price of transparency. Unfortunately, as I will argue, the glut theorist's attempts at recapturing classical reasoning using any of these resources is ill-fated when it comes to public reasoning.

1.2 Classical recapture for gap and glut theorists

> *Give a person a fish and you feed them for a day; teach a person to fish and you feed*
> *them for a lifetime.*

Suppose you and I are subclassical logicians, and suppose you want to teach me classical mathematics. One way to teach me would be to simply assert any mathematical statement you've proven using classical reasoning. However, a more sustainable method would be to teach me how to prove theorems myself. One way to do so would be to carry out some derivations and hope that I'll catch on. But upon inspecting your proofs, I am baffled. Your proofs are grossly fallacious; almost none of them are underwritten by the subclassical logic we both adhere to.

If we're both paracomplete proponents of $K3$, this problem is easily overcome. When you present a proof of some theorem of, say, classical set theory you simply add as an additional premise the claim $\forall x \forall y (x \in y \vee \neg x \in y)$, i.e. the claim that set theory is *complete* in the sense that any two sets are such that either the first is a member of the second or it isn't. Since $K3$ becomes classical

logic when we add the law of excluded middle as an axiom, this guarantees that your set-theoretic proof, which is classically valid, becomes valid according to $K3$. Thus, by asserting the relevant instances of the law of excluded middle, $K3$ logicians can recapture classical reasoning in select domains.

Unfortunately, things aren't so simple if instead we're para*consistent* proponents of LP. As a glut theorist, you need to fill the gaps in your proof not by communicating to me that mathematics is complete but that mathematics is *consistent*. Since glut theory is of course compatible with the non-classicality of mathematics, and indeed there are glut theorists who claim that mathematics is inconsistent,[7] there will be no way for me to infer that you take mathematics to be consistent from the fact that you're a glut theorist. It's tempting to think that you can communicate to me that mathematics is consistent by asserting what might be thought to be the dual of the above, i.e. $\neg\exists x\exists y(x \in y \wedge \neg x \in y)$. However, $\neg\exists x\exists y(x \in y \wedge \neg x \in y)$ is logically equivalent to $\forall x\forall y(x \in y \vee \neg x \in y)$ in both LP and $K3$ (as well as in classical logic), and so this attempt at expressing consistency just amounts to an assertion of excluded middle again. And unlike in $K3$, excluded middle is logically valid in LP, so I wouldn't learn anything from your assertion of $\neg\exists x\exists y(x \in y \wedge \neg x \in y)$ that I couldn't already figure out on my own.

More generally, the problem is this: in order for Beall's strategy of recapturing classical reasoning in select domains to be fully effective, gap and glut theorists need to coordinate among themselves which domains they take to be classical. After all, just like any other conversation, public reasoning operates against a background of shared assumptions, the so-called *common ground*.[8] Gap theorist can add to the common ground and establish that a certain domain is classical by simply asserting the right instance of the law of excluded middle. How might glut theorists achieve the same thing?[9] In the next two sections, I critically discuss two answers to this question that have been offered in the literature.

<hr>

[7]See Priest (2006, esp. ch. 17), and Mortensen (2013) for an overview.

[8]See Stalnaker (2014).

[910]This issue is related to the so-called "just true" problem, which is the problem of how glut theorists can express that a sentence is just true and not also false. See Parsons (1984), Batens (1990, sect. 4), Parsons (1990, sect. 6), Simmons (1993, §4.4), Everett (1994, 1996), Bromand (2002), Olin (2003, ch. 2), Shapiro (2004), Littmann and Simmons (2004), Priest (2008a, §6.3), Priest (2006, §20.4), Beall (2009, 2013a), Rossberg (2013), Berto (2014), Hughes (2015), and Murzi and Carrara (2015b).

1.3 Perlocutionary recapture: shrieking and whistling

We can distinguish attempts at recapturing classicality into two initial groups: those involving per-locutionary tools and those involving illocutionary tools. This distinction is of course familiar from J. L. Austin (1962), who uses 'illocution' to designate the *force* of a speech act and 'perlocution' to designate the causally downstream effects of a speech act. In this section, we'll look at two perlocutionary attempts at recapturing classicality, one involving so-called shriek rules and one involving conversational implicatures.

1.3.1 Shrieking

Beall (2013a) develops a device based on Graham Priest's (2006, §8.5) notion of "shrieking."[11] He proposes that glut theorists may adopt an extra-logical rule of inference, called a *shriek rule*, that allows them to infer anything from a contradiction in classical domains such as mathematics. This rule says that if you have a sentence of mathematics ϕ_M, then you may infer anything from $\phi_M \wedge \neg\phi_M$.

It's tempting to express the idea behind shrieking in terms of an extra-logical *axiom* instead of a rule. It's tempting, that is, to think that a glut theorist can get the same effect by asserting $(\phi_M \wedge \neg\phi_M) \rightarrow \bot$, where $\bot$ is some sentence that entails every other sentence. However, since *modus ponens* for the material conditional $\rightarrow$ isn't valid in LP, we may not conclude from a glut theorist's assertion of ϕ_M and $(\phi_M \wedge \neg\phi_M) \rightarrow \bot$ that she regards ϕ_M as classical. In fact, in LP just as in classical logic, $(\phi \wedge \neg\phi) \rightarrow \bot$ is a logical truth, for any ϕ, and so its assertion by a glut theorist doesn't tell us anything we didn't already know.[12]

The difference between shriek rules and shriek axioms points to the more general fact that in LP, unlike in classical logic, there is a big difference between a rule that allows us to infer ψ from

---

[11] See also Field (2008, 388). See Murzi and Carrara (2015a) for some worries about shrieking in addition to the ones discussed in this section.

[12] Priest (2006) as well as Beall (2009) have devised logics based on LP that contain a conditional that does validate modus ponens. So in those logics, a shriek axiom may do just as well as a shriek rule. (Though see Beall (2013a, §2.1) for worries about Priest's version of shriek axioms.) But as mentioned in the introduction, the present focus is on determining how much we can get out of the simple logics $K3$ and LP.

ϕ and the corresponding axiom $\phi \rightarrow \psi$.[13] This is unfortunate. If $(\phi_M \wedge \neg\phi_M) \rightarrow \bot$ did indeed achieve its desired effect, then glut theorists would only need to assert it along with ϕ_M in order to express that they take ϕ_M to be classical. With a shriek *rule*, things aren't as straightforward. Before we get to that, note that shriek rules may well help glut theorists selectively recapture classicality *in thought*. But we've been discussing the issue of how two or more glut theorists can *coordinate* on the classicality of a certain domain.

You could attempt to express to me that you take mathematics to be classical, by asserting that you've adopted the relevant rule. But absent a way for us to coordinate on the classicality of rules and their adoption, I won't be able to rule out that it's also *false* that you've adopted the rule. You can't rule out, that is, that it's both true and false that you're committed to ψ_M. So while you can tell me that you're committed to the classicality of mathematics, you can't tell me that *that* is a classical truth, a truth that's not also false, that you are thus committed. Compare this again with how simple it is for gap theorists to unequivocally commit themselves to the classicality of mathematics: they simply need to assert the relevant instances of the law of excluded middle.

Might I be able to read off of your *behavior* that you take mathematics to be classical? In general, people's inferential behavior is not a good guide to what they accept as consequences of their beliefs.[14] But perhaps the present case is different. Note that adding material modus ponens to LP suffices to recapture classical reasoning. So, perhaps you can infer ψ_M from $\phi_M \rightarrow \psi_M$ and ϕ_M often enough to indicate to me that you accept a version of modus ponens restricted to sentences of mathematics. Whether this may work can't be determined from the armchair. But note that this route to recapturing classicality would be much more circuitous compared to the ease with which gap theorists can commit themselves to the classicality of mathematics.

1.3.2 Whistling

It has also been suggested that glut theorists can express classicality by way of conversational implicatures. Perhaps I can compute the implicature that you take mathematics to be classical if you, the

---

[13]For this reason, it also wouldn't help to add $\big(\phi_M \wedge (\phi_M \rightarrow \psi_M)\big) \rightarrow \psi_M$ as an axiom, because this sentence is already a logical truth of LP. So perhaps Lewis Carroll's (1895) Tortoise was a proponent of LP?

[14]See Harman (1986, ch. 2) and Harman (2009). See also Scharp (2013, 82) for a similar observation.

glut theorist, never assert both a mathematical sentence and its negation.[15] My pragmatic reasoning might go like this: it is a maxim of conversation, called the *maxim of quantity*, that one ought to say just enough of what's relevant for the purposes of a conversation, not more and not less. It follows from this maxim that if you thought that the sentence you asserted was both true and false, you would have asserted it *and* its negation. Since you didn't assert the negation, you must think that the sentence is classical. Do this often enough and I may be able to infer that you take mathematics to be classical.

However, in the present case, the maxim of quantity conflicts with another conversational maxim, the maxim that says to not assert anything that's already in the common ground, even if it's relevant to the topic under discussion. Since Robert Stalnaker (1978, 49) endorses this maxim, we'll call it *Stalnaker's maxim*. A result of Stalnaker's maxim is that speakers will often not assert things that they think the audience already believes. This creates trouble for the maxim of quantity, since in many conversations glut theorists will believe that their audience already believes the *negation* of a certain sentence, and what they want to establish is that the sentence itself is true *in addition* to being false. So, they will only assert the sentence, without its negation. Indeed, glut theory's core thesis, the thesis that there are sentences that are both true and false, is a case in point. According to glut theorists' own lights, this is a non-classical thesis in that it is both true and false. But glut theorists don't often go around asserting the falsity of their core thesis.[16] That's plausibly because at least in conversations with classical logicians, the falsity of this claim is already common ground. As a result, your pragmatic reasoning would lead you astray if you inferred that I believe that the claim that there are sentences that are both true and false is classical from the fact that I don't go around asserting its negation.

Another problem with this pragmatic strategy is that the purported conversational implicatures, even if they were generated, wouldn't display the same behavior displayed by implicatures in general.[17] There are two general features of conversational implicatures that are relevant here; purported implicatures of classicality exemplify the first one but not the second. The first feature, stressed by

---

[15]See Shapiro (2004, 339), Priest (2006, 291), Beall (2009, 51–2), and Armour-Garb and Priest (2005, 168). See Grice (1967) for the pragmatic theory that's in the background here.

[16]For an honorable exception, see Priest (1979, 239).

[17]Thanks to Ben Burgis for urging me to think more about this point.

Grice, is that we can cancel implicatures. Here the supposed implicatures of classicality behave like ordinary implicatures: to cancel the supposed implicature that what you asserted is classical, you simply add an assertion of its negation. But there's another feature of implicatures that naturally goes with their cancelability: implicatures can usually be made explicit. For example, if you say of some people that they moved in together and adopted a puppy, thereby generating the implicature that the moving in occurred before the adoption, you can make that implicature explicit by saying that they moved in together *and then* adopted a puppy. If there's any doubt on the part of your audience about whether your chosen sentence order corresponds to the temporal order of the events your assertion is about, you can dispel that doubt by making the temporal order explicit. And the same holds for most conversational implicatures. The purposes of most implicatures isn't to express the inexpressible, but to cut corners.[18] The same isn't the case for the supposed conversational implicatures of classicality. On the current proposal, glut theorists can only ever *suggest*, but never actually *say*, that something is classical. That a domain is classical therefore becomes one of those mysterious Tractarian truths that we can only allude to but never actually put into words, at least not unequivocally. What we thought was a mundane feature of many domains, namely that they are classical, takes on almost mystical qualities.[19]

It may be wondered whether the above clash between the maxim of quantity and Stalnaker's maxim points to a general flaw with the Gricean approach to pragmatics and its reliance on multiple principles. Perhaps, that is, the problem lies not with the present proposal of how to express classicality, but rather with the theoretical framework within which it is proposed. An alternative approach to pragmatics that only relies on one principle of communication is the *relevance-theoretic* approach due to Dan Sperber and Deirdre Wilson (1995; 2012). Robyn Carston puts the *presumption of optimal relevance*, which is the principle at the core of this approach, as follows: "Speakers should not be, and are expected not to be, as explicit as possible. They should encode only what

---

[18]It's difficult to find anyone explicitly endorsing this feature of implicatures in the literature. Perhaps that's because this feature is taken to be so deeply entrenched in the common ground that it's usually not worth emphasizing. However, Fox (2007, 78) endorses the more limited claim that all *scalar* implicatures can be made explicit; and the purported implicature of classicality would be at least a distant relative of scalar implicatures.

[19]See Wittgenstein (1922): "There is indeed the inexpressible. This *shows* itself; it is the mystical" (§6.522; emphasis in the original). Of course, Priest (2002, 2014a,b) thinks that there are ineffable things that we can nonetheless talk about. But according to him, these things are only found at the periphery of language.

they cannot rely on their addressees to infer easily" (2002, 289). However, it's hard to see how the presence of this principle in communicative situations would lead to anything like an implicature of classicality. In fact, in our present situation, the principle seems to be entirely idle. After all, we've already seen that classicality can't be encoded; that's why there was a need for a pragmatic strategy to begin with. So, within the relevance-theoretic approach to communication, classicality can only ever be inferred by the addressee. But the whole problem of coordinating on classical domains rests on the premise that classicality can't be inferred easily.

At this point it may be protested that surely things can't be *that* bad. After all, glut theorists do say all the time that certain domains are classical while others aren't.[20] Shouldn't we take them at their word, or at what their word implicates? Indeed, perhaps there is some pragmatic difference between an assertion of ϕ and an assertion $\phi \wedge \neg Tr(\neg \ulcorner \phi \urcorner)$, even if, as per transparency, there is no logical difference. I don't take myself to have refuted this possibility once and for all. But glut theorists owe us a story about what exactly this pragmatic difference amounts to, and ideally this story would steer clear of mysticism. Note also that for Beall's brand of glut theory, the resources available to tell a story about the pragmatic difference between ϕ and $\phi \wedge \neg Tr(\neg \ulcorner \phi \urcorner)$ are particularly limited. For, Beall's brand of glut theory comes with a strong form of deflationism about truth that says that the nature of truth is exhausted by the function played by the truth predicate as a device of generalization.[21] It's not obvious that this attitude towards truth leaves room for a pragmatic difference between ϕ and $\phi \wedge \neg Tr(\neg \ulcorner \phi \urcorner)$.

Let's take stock. We've looked at two perlocutionary ways of establishing that a certain domain is classical, i.e. two types of speech act which aim to produce the effect of establishing that a domain is classical, and we've found them wanting. In the remainder of the chapter, we'll look at whether there is an *illocutionary* act that glut theorists can rely on.

---

[20]See for example Priest (2008a, ch. 3) and Beall (2009, §1.5).

[21]See Beall (2009, §1.1) and Beall (2015b, §9.2). See also Armour-Garb and Beall (2005b) for a more general overview.

1.4 Illocutionary recapture: metalinguistic negation

Following Priest (2006, 290–5), it may be thought that glut theorists can use the speech act of *denial* to establish that a given domain is classical.[22] Contrary to Frege (1918) and many others, denial is said to be a *sui generis* speech act, not to be reduced to the assertion of a negation. Whereas the assertion of a negation is governed by the norm that one ought to assert a negation only if the negatum is either just false or both true and false, it is said that denial is governed by the norm that one ought to deny something only if it is untrue. If such a speech act exists, then glut theorists can establish that a domain is classical by denying certain claims. For example, to communicate that set theory is classical, a glut theorist could deny the claim $\exists x \exists y (x \in y \land \neg x \in y)$, i.e. the claim that set theory is inconsistent. This is how gap theory and glut theory are exact duals of each other: gap theorists establish the classicality of set theory by asserting that set theory is complete; glut theorists might be able to do so by denying that it is inconsistent.

They might be able to do so, but only if the speech act of denial exist. Even fellow glut theorist David Ripley doubts that it does: "I don't know of any phenomenon studied outside the realm of philosophical logic that could fill the theoretical role occupied by denial in our philosopher's theories" (Ripley, 2015, 292).[23] Some authors, including Priest (2008a, §4.3), disagree. They think that the 'not' of natural language is ambiguous between a truth function and a force indicator.[24] They think, that is, that sometimes when we utter $\neg\phi$, we deny ϕ.

What is the evidence for this ambiguity claim? To start, note that in ordinary speech situations, a sentence may be unassertable for reasons other than because it's false. When an assertion of a sentence suggests that the speaker presupposes something that isn't presupposed by all parties in the conversation, then its negation is generally also unassertable. For example, if Ahmed has never smoked, then 'Ahmed hasn't stopped smoking' is usually unassertable, because its assertion would presuppose that Ahmed has smoked in the past. Similarly, if an assertion of a certain sentence would

---

[22] See also Restall (2005), Priest (2008a, ch. 6), and Estrada-González and Olmedo-García (2013, 96).

[23] See also Murzi and Carrara (2015a) for a criticism of Priest's appeal to denial.

[24] See Price (1990), Richard (2008, ch. 2), and Richard (2009). Horn (1989, §6.2.1), and Tappenden (1999, 279) following him, speak of a "pragmatic ambiguity" here, but Carston (1996, 311) reads Horn as essentially appealing to semantic ambiguity. See Foolen (1991) and Pitts (2011) for further discussion. Note that none of these people are glut theorists.

give rise to false conversational implicatures, then the sentence is unassertable even if it is literally true. For example, if the room was sweltering, then 'The room was warm' is unassertable, since its assertion would give rise to the scalar implicature that the room wasn't sweltering. But now note that with the right sort of stress, the following sentences are perfectly assertable (adapted from Geurts, 1998):

Ahmed hasn't *stopped* smoking—he's never smoked.

The room wasn't *warm*, it was *sweltering*.

These kinds of cases were originally discussed by Laurence Horn (1985, 1989) under the label *metalinguistic negation*. In the above, the speaker doesn't assert that Ahmed hasn't stopped smoking or that the room wasn't warm. For, that would make the respective follow-up clauses contradict what they follow up on, which is of course not what's intended. Rather, what seems to be going on here is that the speaker denies that Ahmed has ever smoked and that the room was *just* warm and not sweltering. So the 'not' in these examples doesn't seem to be truth-functional negation; rather, it's tempting to think that 'not' acts as a force indicator.

This appeal to metalinguistic negation offered on behalf of the glut theorist is of course a bit peculiar. The case of 'Ahmed has stopped smoking' and 'The room was warm,' and cases of presupposition failures and false implicatures more generally, suggest that there are unassertable sentences whose negations are also unassertable if the negation is read in the ordinary truth-functional way. But these same sentences are *deniable*, i.e. their negations are correctly uttered if the negation is read as a force indicator. The above are thus cases of sentences that are deniable but whose truth-functional negations aren't assertable. That's not what we expect to find in a glut-theoretic context. As we saw, according to the glut-theoretic understanding of denial, a sentence is deniable only if it is untrue, whereas its truth-functional negation is assertable if its negatum is either just false or both true and false. Since, according to glut theory, every sentence that's untrue is false, we should thus get that every sentence that's deniable is such that its truth-functional negation is assertable. That's not what we find in the above cases, and so the phenomenon of metalinguistic negation doesn't provide us with direct evidence for the kind of phenomenon that glut theorists need in order to recapture classicality.

But perhaps glut theorists appeal to metalinguistic negation with a more modest goal in mind. Perhaps all that this appeal is supposed to show is that there is a real pragmatic phenomenon involving 'not' that isn't to be analyzed in terms of truth-functional negation. For now, I will assess the appeal to metalinguistic negation as such. I will investigate, that is, the claim that metalinguistic negation provides glut theorists with an illocutionary tool to recapture classicality, while ignoring the fact that metalinguistic negation seems to have the wrong pragmatic profile vis-à-vis truth-functional negation. I will argue that even when judged by this very weak standard, metalinguistic negation falls short of providing glut theorists with what they need. I will then return to the desired pragmatic profile of denial and argue that if we were to modify metalinguistic negation in such a way as to give it that profile, then it would invite the revenge phenomenon that glut theory is supposed to be immune against.

1.4.1   Expressive limitations

A serious worry about metalinguistic negation is that it can't seem to be used to settle questions and it can only be carried out in reaction to a previous utterance. Thus, even though an appeal to metalinguistic negation is supposed to help glut theorists overcome the expressive limitations that stunt their ability to recapture classicality, some of these expressive limitations persist. Meanwhile, gap theorists may happily use assertion for their purposes, since assertion is of course the paradigmatic speech act used to settle questions and to initiate a conversation.

What evidence is there for thinking that metalinguistic negation is limited in these ways? Regarding its inability to settle questions, Rob van der Sandt and Emar Maier write:

> Just as the primary function of assertion is to convey new information, the primary function of a [metalinguistic negation][25] is to object to information which has been entered before and to remove it from the discourse record. (van der Sandt and Maier, 2003, 2)[26]

Using Robert Stalnaker's (2014) ideas, we can gloss this quote as saying that rather than proposing

---

[25] van der Sand and Maier use 'denial' here; I use 'metalinguistic negation' to indicate that the present proposal is just one possible way to implement what glut theorists call 'denial.'

[26] See Ladusaw (1980, 143), Price (1983, 169), and van der Wouden (1997, 237), and van der Sandt (2003, §2) for similar observations.

to remove some possibilities from the context set, a denial proposes to *restore* certain possibilities that were previously removed. And regarding metalinguistic negation's purely reactive nature, van der Sandt writes:

> Cases of [metalinguistic negation][27] only occur as a reaction to an utterance of a previous speaker. It clearly makes no sense to enter a room and [utter] ['T]he King of France isn't bald, since he does not exist.['] Such an utterance does, however, make sense to [deny] a previous claim that the King of France is bald. (van der Sandt, 1988, 93)[28]

However, in a somewhat different context, Timothy Smiley (1996, 1) makes a proposal that is essentially an attempt to turn metalinguistic negation into a speech act that can be used to settle questions and to initiate a conversation. Smiley suggests that we have a mechanism of metalinguistic negation by way of polar questions that we can pose and immediately respond to in the negative. A polar question is a question that only allows for two responses, 'yes' and 'no.' So, to metalinguistically negate ϕ, Smiley proposes that we utter $\ulcorner\phi$? No.$\urcorner$ For example, to metalinguistically negate that I'm hungry, I would utter 'Am I hungry? No.'[29] Call this device at attempting to effect metalinguistic negation *Smiley's device*.

After introducing his device, Smiley immediately grants that it "is not adequate to deal with every case" (1996, 1). Consider an utterance of the following:

> Has Ahmed stopped smoking? No.

As Smiley seems to admit with respect to a related example, such an utterance would not in general result in a denial of both the claim that Ahmed doesn't currently smoke and the question's presupposition that Ahmed used to smoke. Rather, an isolated utterance of the above seems to express that Ahmed used to smoke and that he still does. Of course, things change if we follow up the above with an utterance of 'Ahmed never smoked.' But such an utterance would be an *assertion* of the proposition that Ahmed never smoked. And once we add that assertion, the preceding 'Has Ahmed stopped smoking? No' is wholly redundant. Note also that while it's tempting to try to effect a metalinguistic

---

[27] See footnote 25.

[28] See also Horn (1989, 74–5). Davis (2016, 30) denies this property of metalinguistic negation. However, the example he uses involves a non-referring name. A negative free logic would treat his example as involving ordinary truth-functional negation.

[29] Rumfitt (2000, 799) also endorses the use of this device. See also Khoo (2015, §III) for a discussion of 'No' as a device of denial, though Khoo only discusses it in the context of a dialogue, not a monologue.

negation in the above example by emphasizing the 'stopped' in 'Has Ahmed stopped smoking?,' it would be exceedingly odd to start a conversation that way. Rather, such emphasis only seems to be appropriate as part of a dialogue, viz.:

Asra: "Has Ahmed stopped smoking?" Briannah: "Has he *stopped* smoking? No."

It's not mysterious why metalinguistic negation should only work in reaction to a previous utterance. When we object to someone else's utterance, we may object to it for a multitude of reasons. We may object to a presupposition made by the speaker, object to an implicature generated by the speaker's utterance, and much more. But in uttering $\ulcorner\phi$? No,$\urcorner$ it would be a pragmatic misstep to object to a presupposition made by an utterance of ϕ or to an implicature generated by it. For, if that's what we object to, we should have just phrased our question in a different way so as to avoid the presupposition or the objectionable implicature. So, what's being objected to with the 'No' in an utterance of $\ulcorner\phi$? No$\urcorner$ can only be the assertoric content of ϕ.[30]

1.4.2 Embedding metalinguistic negation

But things get worse for glut theorists. Standard tests developed in semantics to determine whether something is a force indicator suggest that metalinguistic 'not' is *not* a force indicator. There's a long tradition dating back to Frege (1918) and Geach (1965) that assumes that force indicators don't embed in complex sentence constructions. The famous Frege-Geach problem assumes that the simple sentences appearing in complex sentences don't have any force themselves but rather what's embedded is just the content of those simple sentences. So, if we can find occurrences of 'not' within larger sentence structures that seem to have the same effect as in the examples above, then that suggests not only that 'not' doesn't act as a force indicator in those complex sentences, but also that it doesn't do so in the simple sentences above. And indeed, this is exactly what we find.[31]

---

[30] By the assertoric content of ϕ I mean the immediate contribution that an assertion of ϕ would make to a conversation if all of the parties agree about any presuppositions that an assertion of ϕ may rest on. I say that the assertive content of ϕ is the *immediate* contribution that an assertion of ϕ would make to a conversation because the assertive content of ϕ does not include any conversational implicatures that the conversational parties may compute on the basis of the fact that ϕ has been asserted. See Dummett (1991, 47–50), Stanley (1997), and Ninan (2010) for further discussion.

[31] See Geurts (1998, 283) and Ripley (2015, 292).

Consider the following examples:

> If Ahmed hasn't *stopped* smoking but has never smoked, then his lungs will be very healthy.

> If the room wasn't *warm* but *sweltering*, then Nicole was uncomfortable.

In these examples, the 'not' that we previously had reason to believe is a force indicator appears in the antecedent of a conditional. This suggests that the 'not' isn't a force indicator—neither here nor in our previous unembedded examples.[32]

In addition to these well-known examples, we can also show that metalinguistic negation embeds inside of quantifiers.[33] Suppose that Ahmed and Lupe are in a room. Lupe used to smoke but has now stopped, whereas Ahmed has never smoked. We can then respond as follows to the question whether everyone in the room has stopped smoking:

(1) (a) Well, Ahmed is in the room and he hasn't *stopped* since he's never smoked.

  (b) So, somebody in the room hasn't *stopped* smoking.

It may be worried that 'not' in (1–b) takes wide scope and so that this isn't a genuine case of quantification into the scope of negation. However, note that if we explicitly pull the 'not' out of the scope of 'somebody,' then the result becomes decidedly odd or even false:

(2) It's not the case that somebody in the room has *stopped* smoking.

The emphasis on 'stopped' doesn't seem to be doing anything here and so it's tempting to judge this sentence false, since Lupe *has* stopped smoking. What's more, unlike in the case of (1–b), it would be quite odd to infer (2) from (1–a). So, if the 'not' in (1–b) indicates a speech act of denial, then what is being denied isn't 'Somebody in the room has *stopped* smoking.'

But then what *is* being denied in (1–b)? The scope of the quantifier is an open sentence, roughly 'x is in the room and x hasn't *stopped* smoking,' and so the scope of the 'not' is an open sentence as well, namely 'x has *stopped* smoking.' So, on the assumption that what is involved in this example

---

[32]It is these kinds of considerations that lead Carston (1996, 2002), Iwata (1998), and Geurts (1998) to develop theories of 'not' according to which 'not' is unambiguous.

[33]The following is inspired by Swanson's (2011) treatment of the idea that epistemic modals are force modifiers.

is a speech act of denial, what is being denied is an open sentence. But presumably, the objects of speech acts are only ever (the contents of) closed sentences.[34]

It may be worried that the above example of quantifying-in only shows that 'not' isn't ambiguous when it's used to deny a presupposition. But perhaps there's still reason to think that it's ambiguous when used to deny a conversational implicature. After all, some authors, such as Mahrad Almotahari (2017), argue that there are important differences between these two uses of 'not.' For example, in the example about smoking, the follow-up clause 'he's never smoked' provides an explanation for 'Ahmed hasn't *stopped* smoking,' whereas in the example about the room, 'it was *sweltering*' provides an alternative to the problematic 'The room was warm.' However, we can also find examples involving conversational implicatures where we can quantify into the scope of the relevant 'not.' Suppose it was warm but not sweltering on 360 days of last year, whereas it was sweltering on 5 days. Then we can respond as follows to the question whether it was warm on all days last year:

>Some days last year weren't *warm*, they were *sweltering*.

In sum, metalinguistic negation seems to behave in a perfectly compositional way. Why is this significant? Recall that we temporarily suspended the belief that metalinguistic negation has the wrong pragmatic profile, viz. that it gives rise to cases where it is correct to metalinguistically negate a sentence whose truth-functional negation isn't assertable. This contrasts with the glut-theoretic understanding of denial according to which whenever a sentence is deniable, its negation is assertable. Let's now suppose, counterfactually, that metalinguistic negation has the pragmatic profile of denial as understood by glut theorists. Glut theorists commonly assume that denial is consistent and complete in the following way:

Denial-consistency. There is no sentence that is both assertable and deniable.

Denial-completeness. Given full information, every sentence is either assertable or deniable.

Now, let's assume that, instead of a force indicator, we have an operator $\mathcal{D}$ that expresses denial. In such a case, denial-consistency and denial-completeness give rise to the following two principles:

---

[34]As Stainton (2006) discusses, there is evidence that we sometimes make assertions using things other than complete sentences. But even on Stainton's theory, such assertions are ultimately interpreted as having full propositional contents.

$\mathcal{D}$-**consistency.** $\phi, \mathcal{D}\phi \vdash \bot$

$\mathcal{D}$-**completeness.** $\vdash \phi \vee \mathcal{D}\phi$

Given that $\mathcal{D}$ is an operator, it embeds in any sentence environment in a fully compositional way. This means that we'll have a "revenge" liar sentence λ that "says" $\mathcal{D}Tr(\ulcorner\lambda\urcorner)$. Thus, if we read $\mathcal{D}\phi$ as saying that ϕ is just false and not also true, then λ says of itself that it is just false and not also true. As Beall (2009, §3.1) shows, if we add a transparent truth predicate to LP, then an operator like $\mathcal{D}$ that obeys $\mathcal{D}$-consistency and $\mathcal{D}$-completeness allows us to derive $\bot$ in much the same way that truth-functional negation allows us to derive $\bot$ in classical logic in the presence of a transparent truth predicate. The upshot of this is that glut theorists *can't* have an operator that expresses denial; rather, they need denial to be expressed by a force indicator.[35] But since it looks like metalinguistic negation isn't a force indicator but rather a device that's fully compositional, glut theorists can't rely on a version of metalinguistic negation that fits the pragmatic profile of denial as described by denial-consistency and denial-completeness.

1.5 Conclusion

Let's retrace our steps. We started with the observation that glut theorists have trouble coordinating on the classicality of a given domain, while gap theorists have no such troubles. Then we found that perlocutionary ways of resolving this problem, either by way of shriek rules or by way of conversational implicatures, are inadequate. Then, looking to illocution, we found that metalinguistic negation lacks many of the features that would be required for a fully adequate solution to the glut theorists' expressive troubles.

One perhaps surprising upshot of our discussion is that when comparing the relative merits of different logics, it matters whether those logics are seen to be logics that govern thought or logics that govern speech. As we've seen, the method of shrieking plausibly allows a solitary glut theorist to recapture classicality in thought. Likewise, nothing that we've said rules out that there is a mental

---

[35]Ripley (2015) disagrees on this point. Rather than giving up on denial-as-operator, he gives up on denial-consistency. But as he admits in §10.5, this amounts to identifying $\mathcal{D}$ with paraconsistent truth-functional negation, and thus to giving up the game.

attitude of rejection that has all of the features that we found to be lacking in the case of metalinguistic negation. What we have seen, however, is that when it comes to recapturing classicality *in speech*, the glut theorists' prospects are bleak, at least if they restrict themselves to the currently available pragmatic tools. This is particularly significant in light of the fact that some authors such as Catarina Dutilh Novaes (2015) have recently argued that when asking about the normative role of logic, we should focus on dialogical interactions rather than on solitary thought.

Chapter 2

Glap Theory

2.1   Introduction

Consider the liar sentence:

(λ) The sentence λ is not true.

The liar sentence λ says of itself that it is not true. Is it true or not? Well, it either is or it isn't. If it is, then the world is as λ says it is, and λ says that λ is not true. And if λ it isn't true, then the world is as λ says it is, and so λ is true. So either way, λ is both true and not true.[1] But can that really be? Perhaps we can say this: λ is *kind of* both true and not true. But then again, is λ *really* either true or not true?

Or consider the following sorites series:

Isaiah Thomas	$5'9''$	John Wall	$6'4''$	DeMarcus Cousins	$6'11''$
Michael Adams	$5'10''$	Bill Bradley	$6'5''$	Pau Gasol	$7'0''$
Terrell Brandon	$5'11''$	Michael Jordan	$6'6''$	Wilt Chamberlain	$7'1''$
Chris Paul	$6'0''$	Kawhi Leonard	$6'7''$	Kareem Abdul-Jabbar	$7'2''$
John Stockton	$6'1''$	LeBron James	$6'8''$	Arvydas Sabonis	$7'3''$
Tony Parker	$6'2''$	Kevin Durant	$6'9''$	Ralph Sampson	$7'4''$
Russell Westbrook	$6'3''$	Kevin Love	$6'10''$	Yao Ming	$7'5''$

---

[1]This presentation is adapted from Rayo (2013).

35

The average height for NBA players is roughly $6'7''$. Is Tony Parker short (for a basketball player)? He kind of is and isn't.[2] But is he *really* both? How could that be? Perhaps we shouldn't have assumed that he either is or isn't short for a basketball player.

To say that λ is both true and not true and that Tony Parker is both short and not short is to be a *glut theorist*.[3] To deny that λ is either true or not true or that Tony Parker is either short or not short is to be a *gap theorist*.

I believe that glut and gap theorists each get half the story right. I am a *glap theorist*: λ is kind of both true and not true; Tony Parker is kind of both short and not short. But, really, we shouldn't say that λ is either true or not true, Tony Parker is either short or not short. λ is a truth-value glap; Tony Parker falls into the glap between NBA players that are short and those that aren't.

What are glaps? They're kind of both gluts and gaps. But, really, we shouldn't say that they are either gluts or gaps.

By now you should have noticed that I am using the words 'kind of' and 'really' in a distinct way. I'm using them as indicators of assertoric strength: 'kind of' indicates what I call *permissive assertion*, 'really' what I call *restrictive assertion*.[4] The distinction between permissive and restrictive assertion allows us to reconstrue the disagreement between glut and gap theorists and it points to a diagnosis of what glut theorists get right and what gap theorists get right. It also points the way to a novel non-classical response to the liar paradox and the sorites paradox that overcomes many of the shortcomings of glut and gap theory. In particular, there is a percise sense in which the logic of glaps is stronger than the logic of gluts and the logic of gaps taken together.

I proceed as follows. I give a gentle introduction to glut theory and gap theory in section 2.2 and section 2.3. Section 2.4 contains the core of the philosophical motivation for glap theory. Section

---

[2]See Ripley (2011) and Alxatib and Pelletier (2011) for empirical evidence that ordinary speakers find this answer natural.

[3]Note that not everyone who's a glut theorist about the liar is also a glut theorist about the sorites; Beall (2009) defends glut theory about the liar but Beall (2014) criticizes glutty approaches to the sorites. For the purposes of this chapter, I assume for simplicity that glut theory about one domain goes hand in hand with glut theory about the other.

[4]Two quick notes on this terminology. First, my notion of restrictive assertion is the same as Ripley's (2013a) notion of *strict* assertion, and my notion of permissive assertion is the same as Ripley's notion of *tolerant* assertion. However, my logic of restrictive and permissive assertion is quite different from Ripley's logic ST. I thus use this slightly different terminology to ward off possible confusion. I discuss a close relative of ST and it's relation to my logic in subsection 2.10.4. Second, perhaps contrary to Yablo (2014, 33), the 'permissive' in 'permissive assertion' isn't supposed to suggest that, when ϕ and $\neg\phi$ are both permissively assertable, neither ϕ nor $\neg\phi$ is forced on us and that either is permitted. Rather, permissive assertion is permissive in the sense that *both* ϕ are $\neg\phi$ are permitted. It doesn't generally follow from the fact that it's permissible to do Φ and permissible to do $\neg\Phi$ that it is permissible to do both Φ and $\neg\Phi$.

2.5 discusses the revenue phenomenon. Then, after taking stock in section 2.6, I begin the formal development of glap theory in section 2.7. Readers who are mainly interested in the philosophical picture may skip many of those details, although they may wish to take a look at section 2.9 and subsection 2.10.4, which contain further philosophical discussions.

2.2 Thesis: gluts

Let's look at gluts in a bit more detail.

The simplest logic of gluts is Priest's (1979) Logic of Paradox LP. Its model theory adds to the truth values 1, representing classical truth, and 0, representing classical falsity, a third truth value $\frac{1}{2}$, representing truth *and* falsity. The connectives conjunction $\wedge$, disjunction $\vee$, negation $\neg$, and the material conditional $\rightarrow$ are interpreted by way of a generalization of the classical clauses: the truth value of $\phi \wedge \psi$ is the lower of the truth values of ϕ and ψ, that of $\phi \vee \psi$ the higher of the two, that of $\neg\phi$ is 1 minus that of ϕ, and that of $\phi \rightarrow \psi$ is the higher of that of $\neg\phi$ and ψ. For example, if ϕ is 0 and ψ is $\frac{1}{2}$, then $\neg\phi \wedge \psi$ is the lower of the truth values $1-0$ and $\frac{1}{2}$, that is $\frac{1}{2}$. Interpreting the connectives in this way ensures that we have that $\phi \rightarrow \psi$ and $\neg\phi \vee \psi$ always have the same truth value, and so we also have contraposition: $\phi \rightarrow \psi$ and $\neg\psi \rightarrow \neg\phi$ always have the same truth value. Furthermore, we have the De Morgan laws: $\neg(\phi \wedge \psi)$ and $\neg\phi \vee \neg\psi$ always have the same truth value, as do $\neg(\phi \vee \psi)$ and $\neg\phi \wedge \neg\psi$.

The addition of the third truth value $\frac{1}{2}$ gives us the ability to classify the liar sentence λ and the sentence 'Tony Parker is short' as being neither classically true nor classically false. This in turn allows us to preserve the thought that truth is *transparent* and vague predicates are *tolerant*.

Where 'Tr' is the truth predicate, transparency is the thought that a sentence ϕ and the sentence $Tr(\ulcorner\phi\urcorner)$ that says that ϕ is true are intersubstitutable in extensional contexts—that they always have the same truth value.[5] A transparent truth predicate is desirable because it can function as a device for expressing generalizations. For example, the sentence

Everything the pope says is true

---

[5]See Field (2008) and Beall (2009).

amounts to the infinite conjunction[6]

   If the pope says that ϕ_1, then ϕ_1; and if the pope says that ϕ_2, then ϕ_2; . . . .

But in classical logic, transparency leads to disaster, because contradictions are *explosive*: we can infer anything from a sentence and its negation, such as λ and $\neg\lambda$. But the logic of gluts LP isn't explosive. That's because logical consequence is defined as preservation of truth values 1 or $\frac{1}{2}$: an argument is valid just in case, if each premise has either truth value 1 or $\frac{1}{2}$, then the conclusion has either truth value 1 or $\frac{1}{2}$ as well. If the premises are λ and $\neg\lambda$ and the conclusion is some arbitrary sentence ϕ, then the argument may be invalid, because λ and $\neg\lambda$ both have truth value $\frac{1}{2}$ while ϕ may have truth value 0.[7]

   Tolerance is the thought that if a vague predicate applies to one element in a sorites series, then it also applies to the next one.[8] For example, if Tony Parker is short, then so is Russell Westbrook. In classical logic, this again spells disaster: Isaiah Thomas is clearly short, and Yao Ming clearly isn't short. But if 'short' is tolerant, then the fact that Isaiah Thomas is short tells us that Michael Adams is short as well. And that tells us that Terrell Brandon is short, and so on. A few more applications of tolerance tells us that Yao Ming is short as well. But it's not, and so the whole series explodes. In LP, we can hold on to the thought that 'short' is tolerant without being committed to the claim that Yao Ming is short. That's because, in addition to explosion, material modus ponens is invalid in LP as well: we may have cases where ϕ and $\phi \rightarrow \psi$ are both true but ψ is not true—as in the case where ϕ has truth value $\frac{1}{2}$ and ψ has value 0, and so $\phi \rightarrow \psi$ has value $\frac{1}{2}$.

2.3   Antithesis: gaps

But can truth and falsity really overlap? Isn't it part of their nature that they don't? That's the thought that motivates gap theory. The simplest logic of gaps is Kleene's (1950, §54) Strong 3-valued Logic $K3$. Its model theory also uses the three truth values 1, $\frac{1}{2}$, and 0, and it interprets the connectives just

---

[6]As we're about to see, modus ponens is invalid in LP. This means that we technically need to characterize this thought slightly differently in the context of glut theory: a commitment to 'Everything the pope says is true' amounts to a commitment to infinitely many rules of inference of the form 'The pope says that ϕ_1' $\vdash \phi_1$, 'The pope says that ϕ_2' $\vdash \phi_2$, ...

[7]See Dowden (1984) for how to achieve transparency in LP.

[8]See Colyvan (2009), Weber (2010), Priest (2010), and Weber et al. (2014)

like LP does. But instead of having the truth value $\frac{1}{2}$ represent both truth and falsity, $\frac{1}{2}$ represents the absence of truth or falsity. As a result, logical consequence is defined as preservation of just truth value 1: an argument is valid just in case, if each premise has truth value 1, then the conclusion has truth value 1 as well.

It's sometimes said that when a sentence ϕ has truth value $\frac{1}{2}$, then gap theorists think that ϕ is neither true nor false. But that's not quite accurate, at least not if we assume transparency. Gap theorists won't *assert* that ϕ is neither true nor false. Truth being a norm of assertion, gap theorists will only assert sentences that have truth value 1. But if ϕ has truth value $\frac{1}{2}$ and we assume transparency, then $\neg Tr(\ulcorner\phi\urcorner) \wedge \neg Tr(\ulcorner\neg\phi\urcorner)$, the claim that ϕ is neither true nor false, will be equivalent to $\neg\phi \wedge \neg\neg\phi$, which will have truth value $\frac{1}{2}$ as well. So rather than asserting that ϕ is neither true nor false, gap theorists will refrain from asserting that it is either.[9]

Gap theorists refrain from asserting that the liar sentence is either true or not true, that Tony Parker is either short or not short, and that it's either true or false that if Tony Parker is short, then so is Russell Westbrook. This still gives them transparent truth,[10] and it allows them to assert that Isaiah Thomas is short but Yao Ming isn't.

Here's one way of thinking about the difference between glut and gap theorists. Both camps agree with the classical camp that there are sentences, such as "2+2=4' is true' or 'Isaiah Thomas is short,' that are assertable and whose negations aren't assertable. And they also agree with the classical camp that there are sentences, such as "2+2=5' is true' or 'Yao Ming is short,' that aren't assertable but whose negations are. Furthermore, glut and gap theorists agree with each other that there's a third kind of sentence, such as the liar sentence λ and 'Tony Parker is short,' that have a quite different status: their assertability and the assertability of their negations stand and falls together. Glut theorists are permissive—they assert both λ and $\neg\lambda$ and both 'Tony Parker is short' and 'Tony Parker isn't short.' Gap theorists are restrictive—they refuse to assert any of them.

The way I understand glut and gap theory, glut and glap theorists don't disagree about "the

---

[9]There is thus a sense in which gap theorists can't say as much as they would like, because they can't say that the liar sentence is neither true nor false. At the same time, there's a sense in which glut theorists have to say more than they would like: given the above equivalence, they have to say not just that the liar sentence is both true and false but also that it's neither.

[10]See Kripke (1975).

facts" concerning the liar or Tony Parker, and they also don't disagree about the meaning of 'not.' To the extent that they disagree about anything, they disagree about whether to use permissive or restrictive assertion.

Permissive assertion. A model represents ϕ as permissively assertable iff ϕ has either truth value 1 or $\frac{1}{2}$. LP-consequence preserves permissive assertability.

Restrictive assertion. A model presents ϕ as restrictively assertable iff ϕ has truth value 1. $K3$-consequence preserves restrictive assertability

Thus, whenever the gap theorists restrictively asserts ϕ, the glut theorist may permissively assert ϕ; but sometimes the glut theorist may permissively assert ϕ even when the gap theorist refuses to restrictively assert ϕ. Any disagreement between glut and gap theorists can then be dissolved if glut theorists use the assertoric strength indicator 'kind of' for permissive assertion and gap theorists use 'really' for restrictive assertion. When the glut theorist says,

The liar sentence λ is **kind of** either true or not true; and in fact it's **kind of** both

and the gap theorist refuses to say,

λ is **really** either true or not true

they don't disagree. Compare: if you *conjecture* that there are infinitely many twin primes and I refuse to *swear* that there are, we don't disagree.

I'm not proposing that conceiving of the debate between glut and gap theorists as one involving two speech acts captures the intent of every glut or gap theorist. But I am proposing that it is fruitful to do so. It also naturally suggests a hybrid theory that combines both kinds of assertion. The resulting *glap theory* overcomes many of the shortcomings of glut and gap theory.

Before we move on to a presentation of glap theory, a quick word on how much pragmatic innovation it requires. Just like we can distinguish between restrictive and permissive assertion, we can distinguish between restrictive and permissive *denial*:

Permissive denial. A model represents ϕ as permissively deniable iff ϕ has either truth value 0 or $\frac{1}{2}$.

Restrictive denial. A model presents ϕ as restrictively deniable iff ϕ has truth value 0. $K3$-consequence preserves restrictive assertability

Adding restrictive denial to restrictive assertion wouldn't help the gap theorist, since she can already restrictively deny ϕ by restrictively asserting $\neg\phi$. Likewise for adding permissive denial to the glut theorist's permissive assertion. However, adding permissive denial to restrictive assertion, and restrictive denial to permissive assertion, does make a difference. Because a paradoxical sentence receives the non-classical truth value, our gap theorist can assert neither it nor its negation. But equipped with permissive denial, she can deny both it and its negation, thereby expressing that this sentence falls into the assertibility gap.[11] Likewise, because our glut theorist asserts both classically true and paradoxical sentences, she cannot use assertion to express that a sentence is classically true. But equipped with restrictive denial, she can do so, by denying its negation.

It is for these reasons that glut and gap theorists have made use of permissive and restrictive denial, respectively.[12] But now note that if we already have permissive or restrictive denial, then we can get permissive or restrictive assertion with the help of negation: we can simply identify permissive assertion with the permissive denial of a negation, and likewise for restrictive assertion and denial. In fact, there are no deep considerations that would speak in favor of developing glap theory with the help of two primitive speech acts of assertion over developing it with permissive/restrictive assertion and restrictive/permissive denial. I will develop it with the help of two primitive speech acts of assertion purely out of convenience.

2.4   Synthesis: glaps

A model for glap theory still uses 1 and $\frac{1}{2}$ to represent permissive assertability and 1 to represent restrictive assertability. But our logic now has two premise sets: one set for the sentences that are assumed to be restrictively assertable and one for those assumed to be permissively assertable. And we have two distinct, but interdefinable, consequence relations. If Γ_P is a set of permissively assertable premises and Γ_R a set of restrictively assertable ones, then ϕ is a permissively assertable conclusion iff whenever every member of Γ_P has truth value 1 or $\frac{1}{2}$ and every member of Γ_R has

---

[11]We can now put the characterization of gap theory offered in the introduciton more precisely: there we said that a gap theorist denies that the liar sentence is either true or not true or that Tony Parker is either short or not short. What I had in mind was *permissive* denial.

[12]See Priest (2006, §20.4) and Priest (2008a, §6.3) for the case of gluts and Tappenden (1999) and Richard (2008, ch. 2) for the case of gaps.

truth value 1, then ϕ has truth value 1 or $\frac{1}{2}$. And ϕ is a restrictively assertable conclusion iff, in those same circumstances, ϕ has truth value 1. (These notions of restrictive and permissive consequence are developed in more detail starting in section 2.7.)

Of course, glap theory is only an improvement over either glut or gap theory if permissive and restrictive assertion served distinct but important purposes. It becomes clear that they do once we realize how each helps mitigate some of the drawbacks of the other. Permissive assertion is very useful when talking about the liar or Tony Parker. Let λ and $\neg\lambda$ again be the liar sentence and its negation, and let τ and $\neg\tau$ be the sentence 'Tony Parker is short' and 'Tony Parker isn't short.' When someone permissively asserts λ as well as $\neg\lambda$ or τ as well as $\neg\tau$, we immediately know how, according to the speaker, things stand with the liar or Tony Parker. Similarly when the speaker asserts "2+2=4' is true' or 'Isaiah Thomas is short.' We learn less from a permissive assertion of "2+2=4' is true' or 'Isaiah Thomas is short' or a refusal to restrictively assert any of λ, $\neg\lambda$, τ, and $\neg\tau$. Since a permissive assertion of, say, 'Isaiah Thomas is short' is compatible with a permissive assertion of 'Isaiah Thomas isn't short,' we can't infer from a speaker's permissive assertion of 'Isaiah Thomas is short' that the speaker doesn't also believe that Isaiah Thomas isn't short.[13] Likewise, if a speaker refuses to assert either τ or $\neg\tau$, we won't immediately know whether that's because the speaker doesn't have enough evidence regarding the Tony Parker's height or whether she knows all there is to know about it.

Equipped with both kinds of assertion, we can express ourselves more completely: to state what we think about Isaiah Thomas, we restrictively assert 'Isaiah Thomas is short,' which immediately commits us to the restrictive unassertability of 'Isaiah Thomas isn't short.' And to state what we think about Tony Parker, we permissively assert both τ and $\neg\tau$, which tells our audience that we take ourselves to have learned everything there is to learn about whether Tony Parker is short.

But the improvements aren't just in expressive power. Solomon Feferman famously complained that LP and $K3$ are so weak that "nothing like sustained ordinary reasoning can be carried on" in them (1984, 95).[14] We already saw that explosion isn't valid in LP. That's what allows glut theorists to accept that λ and $\neg\lambda$ are both true without being committed to the truth of every sentence

---

[13]This is sometimes called the "just true" problem. See chapter 1.

[14]At the time, Feferman was talking about $K3$, but he (2012, 190) has since observed that the same holds for LP.

whatsoever. And we saw that the law of excluded middle $\phi \vee \neg\phi$ isn't valid in $K3$, which is what allows gap theorist to block the derivation of both λ and $\neg\lambda$ by refusing to assert $\lambda \vee \neg\lambda$. But we also already saw that the reason why glut theorists can hold on to the claim that 'short' is tolerant without being committed to the claim that Yao Ming is short is that their logic doesn't validate modus ponens for the material conditional. And since the material conditional $\rightarrow$ is defined so that $\phi \rightarrow \psi$ abbreviates $\neg\phi \vee \psi$, the failure of the law of excluded middle in $K3$ leads to a failure of the law of identity $\phi \rightarrow \phi$ in $K3$. Reasoning by modus ponens and the law of identity are central to most applications of classical logic. To rectify these shortcomings of LP and $K3$, Field (2008), a gap theorist, and Beall (2009), a glut theorist, have proposed adding new conditionals to $K3$ or LP, respectively, that aren't defined in terms of negation and disjunction. But the resulting logics are exceedingly complex.[15]

By making use of both permissive and restrictive assertion, glap theorists can improve on the weaknesses of LP and $K3$ while preserving the simplicity of those logics. Although material modus ponens remains invalid for permissive assertion and the law of identity remains invalid for restrictive assertion, they can use modus ponens when reasoning with restrictive assertion and the law of identity when reasoning with permissive assertion.

And that's not all. The logic of permissive and restrictive assertion isn't just the logic of permissive assertion and the logic of restrictive assertion combined—the two kinds of assertion interact to yield a stronger logic. In the context of the paradoxes, a stronger logic is generally a good thing. For, what the paradoxes show is that classical logic is too strong to accommodate transparency and tolerance, but what Feferman's complaint shows is that weakening classical logic risks giving up too much for the sake of preserving transparency and tolerance. It is therefore an immediate improvement that the logic of glap theory is stronger than the logic of glut theory and the logic of gap theory combined while allowing for transparency and tolerance.

Here is the precise sense in which the logic of glap theory is stronger than LP and $K3$ combined. In addition to material modus ponens for permissive assertion, we also have two hybrid forms of material modus ponens: if $\phi \rightarrow \psi$ is permissively assertable and ϕ is restrictively assertable, then ψ

---

[15]See McGee (2010) for the case of Field's logic.

is permissively assertable; and if $\phi \to \psi$ is restrictively assertable and ϕ is permissively assertable, then ψ is restrictively assertable.[16]

Here is how the first rule works in practice; a similar example could be given for the second one. Suppose you are a glut theorist and you know that Reggie has made it his life's work to solve the liar paradox. Your glut theorist friend then tells you that *if* Reggie has made it his life's work to solve the liar paradox, then the sentence β written on the blackboard in his office is true. Unbeknownst to you, β reads as follows:

The sentence written on the blackboard in Reggie's office is not true

Thus, unbeknownst to you, β is a liar-like sentence. If you knew what it said, you would recognize that your glut theory commits you to its truth (and also to its untruth). But you don't know what it says; all you know is that if Reggie has made it his life's work to solve the liar paradox, then β is true, and that Reggie has in fact made it his life's work to solve the liar paradox. Because modus ponens isn't valid in your logic LP, you can't infer from your knowledge that β is true. But now suppose you're a *glap* theorist. You know that Reggie is far from a borderline case of someone who's made it their life's work to solve the liar paradox. If anyone has made this their life's work, it's Reggie. Thus, for a glap theorist, the claim that Reggie has made it his life's work to solve the liar paradox isn't just permissively assertable, it is also restrictively assertable. Reggie hasn't just *kind of* made it his life's work to solve the liar paradox, he *really* has. In contrast, given that β is a liar-like sentence, the conditional that *if* Reggie has made it his life's work to solve the liar paradox, then he β is true, is only permissively assertable, not restrictively. But this isn't a probem. Using the above first hybrid form of modus ponens, it nevertheless follows that β is permissively assertable. In other words, you can reason as follows:

It's kind of true that if Reggie has made it his life's work to solve the liar paradox, then β is true. And Reggie really has made it his life's work to solve the liar paradox. So β must be kind of true.

This reasoning isn't available to glut theorists, but it is available to glap theorists. (I provide a more

---

[16]As shown in section 2.8, adding these two hybrid forms of material modus ponens to LP and $K3$ yields a complete logic of permissive and restrictive assertion. They thus capture the whole extent to which the logic of permissive and restrictive assertion goes beyond the combination of LP and $K3$.

detailed discussion of the ways in which the logic of glap theory is stronger than LP and $K3$, as well as the ways in which it isn't, in section 2.9.)

Aside from having a stronger logic at their disposal, glap theorists also have a distinctive advantage over glut theorists who insist that they use the exact same speech act as gap theorists (and classical logicians, for that matter) when they assert $\lambda \wedge \neg\lambda$. A typical response to such an assertion is an incredulous stare, followed by the question whether glut theorists *really* mean what they say. Hard-line glut theorists must—and in fact do—respond by insisting that they *really* do. In contrast, when glap theorists (permissively) assert $\lambda \wedge \neg\lambda$ and are asked if they *really* mean what they say, they may respond that they only *kind of* mean it. This makes glap theory less susceptible to incredulous stares than hard-line glut theory.

2.5 Revenge?

The "revenge" phenomenon is a phenomenon where whenever a diagnosis of the liar paradox is offered, we seem to be able to use the tools employed in the diagnosis to construct a close cousin of the liar sentence that the diagnosis can't handle.[17] For example, suppose you said that what's wrong with the liar sentence is that it is meaningless. We could then construct the following sentence:

(μ) The sentence μ is either not true or else meaningless

Now, if μ is either not true or meaningless, then it is true. But if it's true, then it must be meaningful. So suppose μ is meaningful and true. Then it's either not true or meaningless. So it seems that μ is not true or meaningless iff it's both true and meaningful. So it must be that either μ is meaningless but true or meaningful but not true. The first option is a nonstarter. So μ must be meaningful. But that reduces μ to the original liar sentence λ that says of itself that it is not true.

Just like it's not immediately obvious how to respond to the original liar paradox, it's not immediately obvious how to respond to the liar's revenge. But given the ubiquity of the revenge phenomenon, we should expect that glap theorists have to contend with it as well.

And indeed they do, although because glap theory borrows its model theory from LP and $K3$,

---

[17]See Beall (2007b).

the revenge sentence is the same for the former as it is for the latter.[18] Consider the following sentence:

(γ) The sentence γ does not have truth value 1

Reasoning classically, we have that γ either has truth value 1 or it doesn't. If it does, then we immediately have a contradiction. So what if γ does not have truth value 1? Then what it says is true. Does that mean that it has truth value 1 or that it has truth value $\frac{1}{2}$? If it has truth value $\frac{1}{2}$, then we're in the clear. But that's implausible. After all, the model theory for glap theory, just like that for LP and $K3$, is developed in a classical metatheory, which doesn't allow for sentences that have non-classical truth values. Since γ is a claim about model theory, γ should have truth value 1. Thus, it seems that γ has truth value 1 iff it doesn't have truth value 1. But that's a contradiction *in the classical metatheory*. The only way to avoid this contradiction is to insist that γ isn't a sentence that can be expressed in the language of glap theory; or rather that the concept—call it C—of having truth value 1 isn't a concept that can be expressed in that language.

Glap theory seems to be in trouble now. C seems to be a perfectly intelligble concept. In fact, it seems that the model theory for glap, glut, and gap theory wouldn't be intelligble if C weren't intelligible. But that means that glap theory only avoids the explosion that the liar sentence causes in classical logic by placing restrictions on what concepts can be expressed in the language of glap theory. How is that an improvement over the old Tarskian (1933) response to the liar paradox that restricts its language so that it can't contain a transparent truth predicate?

Whether the glap theorist's response to the liar paradox is an improvement over the Tarskian one depends on whether it is just as important for a language to be able to express C as it is for a language to contain a transparent truth predicate. That it is important for a language to contain a transparent truth predicate was already discussed in section 2.2: a transparent truth predicate yields significant expressive advantages. What about C, the concept of having truth value 1?[19]

Two paragraphs ago I suggested that the model theory for glap, glut, and gap theory wouldn't

---

[18]This problem is closely connected to the "just true" problem discussed in section 2.4. The present discussion shows that there can't be a predicate for classical truth in glap theory, even though we can express classical truth using restrictive assertion.

[19]The following three paragraphs are heavily indebted to Beall's (2007a, §1.4.1) discussion of what he calls "too easy revenge."

be intelligble if C weren't intelligible. That's not quite accurate. All that this model theory needs is a concept of a valuation function that takes sentences of the language to the objects 1, $\frac{1}{2}$, and 0. The choice of the objects 1, $\frac{1}{2}$, and 0 is purely arbitrary. Any three objects would do. It's just that we informally gloss 1 as classical truth, 0 as classical falsity, and $\frac{1}{2}$ as "other." But that gloss is misleading. More accurately, 1, $\frac{1}{2}$, and 0 should be glossed as (classical) truth *in a model*, (classical) falsity *in a model*, and "otherness" *in a model*.

The relationship between truth and truth in a model is a controversial issue, one that's tied to the question how model theory relates to the theory of meaning.[20] On one view, the spirit of which may trace back to Wittgenstein (2009), model theory and the theory of meaning are entirely distinct enterprises. Model theory is simply a mathematical theory that is useful in establishing things like the consistency (or non-triviality) of a theory. Semantics, whatever it is, is to be built on entirely different foundations. Thus, to the extent that truth plays any role in the theory of meaning—and it needn't play any[21]—it is an entirely distinct concept from truth in a model.

On this understanding of the relationship between model theory and the theory of meaning, and between truth and truth in a model, it isn't at all worrisome that the glap theorist's language can't express the concept C, *because that concept doesn't make sense to begin with*. The concept of having truth value 1 is a confused algamam of the notion of having value 1 in a model and the concept of truth. Transparent truth, on the other hand, is highly useful, as we've seen. It is thus a significant point in favor of glap theory that it can accommodate transparent truth. And that glap theorist can't accommodate the concept of having truth value 1 isn't a strike against it at all.

Here is a different answer to the question how model theory relates to the theory of meaning, one that roughly coincides with how leading semanticists such as Heim and Kratzer (1998) conceive of their enterprise. Developing a theory of meaning is a bit like doing model theory. What semanticists do is describe the *intended* model of our language. Thus, in an extensional setting, there are many valuation functions, but one valuation function describes the actual meanings of our sentences. It is because 'Grass is green' means that grass is green and because grass is in fact green that the intended valuation function assigns 1 to 'Grass is green.' If that's right, then the concept C makes a

---

[20]See Etchemendy (1990) and Yalcin (2017) for helpful discussions.
[21]See Brandom (1994).

lot of sense: it amounts to the concept of having truth value 1 in the intended model. That the glap theorist's language can't express C would thus seem to be a serious problem. Glap theory purports to contain a theory of truth. But in order to develop a theory of meaning for glap theory, we need to invoke concepts such as C that glap theory can't make sense of. The situation thus seems to be much like the situation we found in the case of the Tarskian response to the liar paradox.

The most promising way forward for the glap theorist who thinks that the theory of meaning is a bit like model theory is to insist that there are two concepts of truth.[22] There is one, thin, notion, which is just the transparent truth predicate that we've been exploring. And then there is a thicker, explanatory notion, which is the one that's used in semantic theorizing, where we distinguish between an object language that is the object of study and the metalanguage of the semanticist. The glap theorist can accommodate this thicker notion of truth, but even in glap theory this thicker notion of truth cannot be self-contained—that's the lesson from Tarski. Consequently, the classical model theory for glap theory and the corresponding theory of meaning for its language can be carried out within glap theory, but glap theorists mustn't forget that when they are developing the theory of meaning, they aren't using the transparent truth predicate that partially motivates their logic.[23]

2.6   Taking stock

We've seen that there is much to be gained by abandoning gluts and gaps in favor of glaps. Adopting both permissive and restrictive assertion, indicated by 'kind of' and 'really' respectively, delivers a logic that's more powerful than the glut theorist's logic and the gap theorist's logic combined, all the while allowing for a transparent truth predicate and a tolerant attitude towards borderline ascriptions of vague predicates. Just like everyone else, glap theorists need to contend with the revenge phenomenon, but there are two promising avenues available to them, depending on their view of the relationship between model theory and the theory of meaning.

In the remainder, I develop the formal details of the logic of restrictive and permissive asser-

---

[22]See Field (1994) and McGee (2005).

[23]In order to do classical model theory, glap theorists need to "recapture" classical reasoning along the lines described by Beall (2015a). Note that classical reasoning can be recaptured very easily in gap theory, but not as easily in glut theory, as Jenny (2017) discusses. Glap theorists can recapture classicality with the same ease as gap theorists can.

tion. However, futher philosophical discussions of the formal theory can be found in section 2.9 and subsection 2.10.4. section 2.11 develops *subervaluationism*, which does for glaps what super- and subvaluationism do for gaps and gluts. Finally, section 2.12 compares glap theory and subervaluationism.

2.7   Restrictive and permissive consequence

The claim that ϕ is a consequence of a set[24] Γ can usually be glossed as saying that if every member of Γ is assertable, then ϕ is assertable. Once two kinds of assertion enter the picture, we need to complicate this a little.

Instead of assuming that the consequence relation has a single argument on its left-hand side, namely the set of sentences that are supposed to be assertable, we will assume that it has two arguments on its left-hand side: first the set of sentences Γ_R that are assumed to be **r**estrictively assertable and second the set of sentences Γ_P that are assumed to be **p**ermissively assertable. We may then ask, if the members of Γ_R are restrictively assertable and the members of Γ_P are permissively assertable, what else is restrictively and permissively assertable?

To make this question more manageable, we split it up into two questions. First, if the members of Γ_R are restrictively assertable and the members of Γ_P are permissively assertable, what else is restrictively assertable? For this purpose, we define the relation $\vDash_R$: $\langle \Gamma_R, \Gamma_P \rangle \vDash_R \phi$ iff every three-valued $K3/LP$-valuation that assigns 1 to every member of Γ_R and 1 or $\frac{1}{2}$ to every member of Γ_P assigns 1 to ϕ.

Second, we ask, if the members of Γ_R are restrictively assertable and the members of Γ_P are permissively assertable, what else is permissivley assertable? For this purpose, we define the relation $\vDash_P$: $\langle \Gamma_R, \Gamma_P \rangle \vDash_P \phi$ iff every three-valued $K3/LP$-valuation that assigns 1 to every member of Γ_R and 1 or $\frac{1}{2}$ to every member of Γ_P assigns 1 or $\frac{1}{2}$ to ϕ.

In this section and in section 2.8 and section 2.9, we suppose that Γ_R and Γ_P are finite so that we can form their conjunctions $\bigwedge \Gamma_R$ and $\bigwedge \Gamma_P$. It is easily verified that all of the relations studied

---

[24]I assume that the antecedents of the consequence and derivability relations, and, starting in subsection 2.10.1, their succedents, are sets, not multisets. This immediately gives us the structural rules of contraction.

in these sections are compact, and so all propositions proved in this section and the next continue to hold if we give up this finiteness assumption. I officially relax the assumption in section 2.10, where it starts to matter. Note also that here and throughout, $\vdash_{K3}$ is $K3$-derivability and $\vdash_{LP}$ is LP-derivability.[25]

Proposition 2.7.1. Suppose $\Gamma_P \neq \emptyset$. Then $\langle \Gamma_R, \Gamma_P \rangle \vDash_R \phi$ iff $\Gamma_R \vDash_{K3} \bigwedge \Gamma_P \to \phi$.

Proof. The following statements are equivalent:

- $\langle \Gamma_R, \Gamma_P \rangle \vDash_R \phi$.

- For every three-valued $K3/LP$-valuation v, if $v(\gamma_R) = 1$, for all $\gamma_R \in \Gamma_R$, and $v(\bigwedge \Gamma_P) \neq 0$, then $v(\phi) = 1$.

- For every three-valued $K3/LP$-valuation v, if $v(\gamma_R) = 1$, for all $\gamma_R \in \Gamma_R$, then $v(\bigwedge \Gamma_P) = 0$ or $v(\phi) = 1$.

- For every three-valued $K3/LP$-valuation v, if $v(\gamma_R) = 1$, for all $\gamma_R \in \Gamma_R$, then $v(\neg \bigwedge \Gamma_P \vee \phi) = 1$.

- $\Gamma_R \vDash_{K3} \bigwedge \Gamma_P \to \phi$.

$$\boxtimes$$

Proposition 2.7.2. Suppose $\Gamma_R \neq \emptyset$. Then $\langle \Gamma_R, \Gamma_P \rangle \vDash_P \phi$ iff $\Gamma_P \vDash_{LP} \bigwedge \Gamma_R \to \phi$.

Proof. The following statements are equivalent:

- $\langle \Gamma_R, \Gamma_P \rangle \vDash_P \phi$.

- For every three-valued $K3/LP$-valuation v, if $v(\bigwedge \Gamma_R) = 1$ and $v(\bigwedge \gamma_P) \neq 0$, for all $\gamma_P \in \Gamma_P$, then $v(\phi) \neq 0$.

- For every three-valued $K3/LP$-valuation v, if $v(\bigwedge \gamma_P) \neq 0$, for all $\gamma_P \in \Gamma_P$, then $v(\bigwedge \Gamma_R) \neq 1$ or $v(\phi) \neq 0$.

- For every three-valued $K3/LP$-valuation v, if $v(\bigwedge \gamma_P) \neq 0$, for all $\gamma_P \in \Gamma_P$, then $v(\neg \bigwedge \Gamma_R \vee \phi) \neq 1$.

- $\Gamma_P \vDash_{LP} \bigwedge \Gamma_R \to \phi$.

$$\boxtimes$$

---

[25]See Priest (2008b, ch. 8) for tableaux systems for $\vdash_{K3}$ and $\vdash_{LP}$.

Proposition 2.7.3.　(a)　$\langle \Gamma_R, \emptyset \rangle \vDash_R \phi$ iff $\Gamma_R \vDash_{K3} \phi$.

(b)　$\langle \emptyset, \Gamma_P \rangle \vDash_P \phi$ iff $\Gamma_P \vDash_{LP} \phi$.

Proof.　Immediate.　　　　　　　　　　　　　　　　　　　　　　⊠

2.8　Restrictive and permissive derivability

$\vdash_R$ and $\vdash_P$ are the smallest relations that are closed under the following axioms and rules:

$$(A1)\ \frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_R \phi}\ \text{whenever } \Gamma_R \vdash_{K3} \phi \qquad (A2)\ \frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_P \phi}\ \text{whenever } \Gamma_R \vdash_{LP} \phi$$

$$(R1)\ \frac{\langle \Gamma_R, \Gamma_P \rangle \vdash_R \phi \to \psi \qquad \langle \Gamma_R, \Gamma_P \rangle \vdash_P \phi}{\langle \Gamma_R, \Gamma_P \rangle \vdash_R \psi} \qquad (R2)\ \frac{\langle \Gamma_R, \Gamma_P \rangle \vdash_P \phi \to \psi \qquad \langle \Gamma_R, \Gamma_P \rangle \vdash_R \phi}{\langle \Gamma_R, \Gamma_P \rangle \vdash_P \psi}$$

Proposition 2.8.1.　$\langle \Gamma_R, \Gamma_P \rangle \vDash_R \phi$ iff $\langle \Gamma_R, \Gamma_P \rangle \vdash_R \phi$.

Proof.　($\Rightarrow$). Suppose $\langle \Gamma_R, \Gamma_P \rangle \vDash_R \phi$. Suppose first that $\Gamma_P \neq \emptyset$. Then $\Gamma_R \vDash_{K3} \bigwedge \Gamma_P \to \phi$, by Theorem 2.7.1. Then $\Gamma_R \vdash_{K3} \bigwedge \Gamma_P \to \phi$. Then:

$$\begin{array}{l}(A1)\\[2pt](R1)\end{array}\ \frac{\dfrac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_R \bigwedge \Gamma_P \to \phi} \qquad (A2)\ \dfrac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_P \bigwedge \Gamma_P}}{\langle \Gamma_R, \Gamma_P \rangle \vdash_R \phi}$$

Next, suppose that $\Gamma_P = \emptyset$. Then $\Gamma_P \vDash_{K3} \phi$, by Theorem 2.7.3.(a). Then $\Gamma_P \vdash_{K3} \phi$, and so:

$$(A1)\ \frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_R \phi}$$

($\Leftarrow$). (A1) is immediate. For (R1), suppose $\langle \Gamma_R, \Gamma_P \rangle \vDash_R \phi \to \psi$ and $\langle \Gamma_R, \Gamma_P \rangle \vDash_P \phi$ and suppose that some $K3/LP$-valuation v is such that $v(\gamma_R) = 1$, for all $\gamma_R \in \Gamma_R$, and $v(\gamma_P) \neq 0$, for all $\gamma_P \in \Gamma_P$. Then $v(\phi \to \psi) = 1$ and $v(\phi) \neq 0$. Then $v(\psi) = 1$, as desired.　⊠

Proposition 2.8.2.　$\langle \Gamma_R, \Gamma_P \rangle \vDash_P \phi$ iff $\langle \Gamma_R, \Gamma_P \rangle \vdash_P \phi$.

Proof.　($\Rightarrow$). Suppose $\langle \Gamma_R, \Gamma_P \rangle \vDash_P \phi$. Suppose first that $\Gamma_R \neq \emptyset$. Then $\Gamma_R \vDash_{LP} \bigwedge \Gamma_R \to \phi$, by Theorem 2.7.2. Then $\Gamma_R \vdash_{LP} \bigwedge \Gamma_R \to \phi$. Then:

$$\begin{array}{l}(A2)\\[2pt](R2)\end{array}\ \frac{\dfrac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_P \bigwedge \Gamma_R \to \phi} \qquad (A1)\ \dfrac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_R \bigwedge \Gamma_R}}{\langle \Gamma_R, \Gamma_P \rangle \vdash_P \phi}$$

Next, suppose that $\Gamma_R = \emptyset$. Then $\Gamma_R \vDash_{LP} \phi$, by Theorem 2.7.3.(b). Then $\Gamma_R \vdash_{LP} \phi$, and so:

$$(A2) \; \frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_P \phi}$$

($\Leftarrow$). (A2) is immediate. For (R2), suppose $\langle \Gamma_R, \Gamma_P \rangle \vDash_P \phi \to \psi$ and $\langle \Gamma_R, \Gamma_P \rangle \vDash_R \phi$ and suppose that some $K3/LP$-valuation v is such that $v(\gamma_R) = 1$, for all $\gamma_R \in \Gamma_R$, and $v(\gamma_P) \neq 0$, for all $\gamma_P \in \Gamma_P$. Then $v(\phi \to \psi) \neq 0$ and $v(\phi) = 1$. Then $v(\psi) \neq 0$, as desired. $\boxtimes$

Note that these completeness proofs go through without requiring any structural rules for our relations $\vdash_R$ and $\vdash_P$. Of course, if we want to generate the axioms proof theoretically, we'll often have to invoke the structural rules for $\vdash_{K3}$ and $\vdash_{LP}$.

Lemma 2.8.3. If $\langle \Gamma_R, \Gamma_P \rangle \vdash_R \phi$, then $\langle \Gamma_R, \Gamma_P \rangle \vdash_P \phi$.

Proof.

$$(R2) \; \frac{\langle \Gamma_R, \Gamma_P \rangle \vdash_R \phi \quad (A2) \dfrac{}{\langle \bigwedge \Gamma_R, \bigwedge \Gamma_P \rangle \vdash_P \phi \to (\bigwedge \Gamma_R \to \phi)}}{(R1) \; \dfrac{\langle \bigwedge \Gamma_R, \bigwedge \Gamma_P \rangle \vdash_P \bigwedge \Gamma_R \to \phi \quad (A1) \dfrac{}{\langle \bigwedge \Gamma_R, \bigwedge \Gamma_P \rangle \vdash_R \bigwedge \Gamma_R}}{\langle \Gamma_R, \Gamma_P \rangle \vdash_P \phi}}$$

(Note that the converse can't be proved like this because we don't have $\vdash_{K3} \phi \to (\bigwedge \Gamma_R \to \phi)$.) $\boxtimes$

Corollary 2.8.4. (a) Suppose $\Gamma_P \neq \emptyset$. $\langle \Gamma_R, \Gamma_P \rangle \vdash_R \phi$ iff $\Gamma_R \vdash_{K3} \bigwedge \Gamma_P \to \phi$.

(b) Suppose $\Gamma_R \neq \emptyset$. $\langle \Gamma_R, \Gamma_P \rangle \vdash_P \phi$ iff $\Gamma_P \vdash_{LP} \bigwedge \Gamma_R \to \phi$.

Proof. (a) follows from Theorem 2.7.1 and Theorem 2.8.1, (b) from Theorem 2.7.2 and Theorem 2.8.2. $\boxtimes$

Corollary 2.8.5. (a) $\langle \Gamma_R, \emptyset \rangle \vdash_R \phi$ iff $\Gamma_R \vdash_{K3} \phi$.

(b) $\langle \emptyset, \Gamma_P \rangle \vdash_P \phi$ iff $\Gamma_P \vdash_{LP} \phi$.

Proof. (c) follows from Theorem 2.7.3.(a) and Theorem 2.8.1, and (d) from Theorem 2.7.3.(b) and Theorem 2.8.2. $\boxtimes$

Corollary 2.8.6. We have the following normal-form theorems: for $I, J \in \{P, R\}$, $I \neq J$, and $i, j \in \{1, 2\}$, $i \neq j$, if $\langle \Gamma_R, \Gamma_P \rangle \vdash_I \phi$, then there is a canonical proof of this of the form

$$(\text{A}i) \; \frac{\langle \Gamma_R, \Gamma_P \rangle \vdash_I \bigwedge \Gamma_J \to \phi \qquad (\text{A}j) \; \frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_J \bigwedge \Gamma_J}}{\langle \Gamma_R, \Gamma_P \rangle \vdash_I \phi} (\text{R}i)$$

Proof. Follows immediately from the proofs of Theorem 2.8.1 and Theorem 2.8.2. ⊠

2.9 Discussion

Theorem 2.8.4 suggests that the logic P is stronger than LP and that the logic R is stronger than $K3$. But Theorem 2.8.5 also shows that P and R are to some extent conservative over LP and $K3$, respectively. I will now discuss how to interpret these somewhat conflicting verdicts and what it all means for glut and gap theorists in turn.

2.9.1 Glut vs. glap theory

As discussed in section 2.4, adding restrictive assertion to the glut theorist's permissive assertion improves the glut theorist's expressive resources. But Theorem 2.8.4 means that there is a precise sense in which adding a second speech act also gives the glut theorist a logic that is stronger than what would be available through the simple combination of LP and $K3$. In glap theory, the whole is greater than the sum of its parts.

It might be worried at the outset that to claim that P is stronger than LP and $K3$ combined is a bit like saying that the modal logic K is stronger than classical extensional logic CL because the former but not the latter allows use to prove $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$. It's not so much that K is *stronger* than CL but rather that K has *expressive recourses* that CL lacks. Similarly, it might be worried that it is only due to the fact that permissive derivability has a richer structure than LP-derivability and $K3$-derivability that we can prove certain sequents in P that we can't prove in either LP or $K3$. Thus, to the extent that Theorem 2.8.4 shows that more things are permissively assertable in P than in LP or $K3$, that's because, using P, we can assert things in two different premise sets, which we can't do in LP or $K3$. And in general, it shouldn't come as a surprise that if we assert more things, then more things become assertable.

Nevertheless, I propose that there is one sense of logical strength in which P is stronger than LP and $K3$ combined. To see this, we need to reflect on the normative role of logic. There are many

views on what this role is.[26] But surely, one thing that logic does is determine what sentences are assertable on the basis of one's evidence. With this in mind, suppose that E is our theorist's body of evidence and that $K3$ determines that Γ_R is the set of sentences that are (restrictively) assertable on the basis of E and that LP determines that Γ_P is the set of sentences that are (permissively) assertable on the basis of E. If Γ_R is empty so that nothing is restrictively assertable, then Theorem 2.8.5 tells us that giving a glut theorist the ability to restrictively assert things doesn't make anything permissively assertable for her that wasn't permissively assertable before. That's of course wholly unsurprising.

But now suppose that Γ_R is non-empty. It follows from Theorem 2.8.3 that $\Gamma_R \subseteq \Gamma_P$. So giving our glut theorist the ability to restrictively assert things won't make anything restrictively assertable that wasn't already permissively assertable before. But it may make things *permissively* assertable that weren't permissively assertable before. For suppose that that $\Gamma_P \vdash_{LP} \bigwedge \Gamma_R \to \phi$. By Theorem 2.8.3, $\bigwedge \Gamma_R$ is already permissively assertable. But because the material conditional doesn't detach in LP, ϕ may not be permissively assertable prior to the introduction of restrictive assertion. Yet, once equipped with restrictive assertion, *and without changing our theorist's evidence*, our theorist's new logic P will determine that ϕ is permissively assertable. This is surprising: introducing a new speech act doesn't just improve our theorist's expressive resources, it also expands what follows from her evidence.

2.9.2 Gap vs. glap theory

The situation for gap theorists is similar, but with a few more complications. Note first again that adding permissive assertion to the gap theorist's restrictive assertion improves the gap theorist's expressive resources. And Theorem 2.8.4 again means that there is a precise sense in which adding a second speech act also gives the gap theorist a logic that is stronger than what would be available through the simple combination of LP and $K3$. But because the converse of Theorem 2.8.3 doesn't hold, the way in which the logic is strengthened is a bit more subtle.

Suppose again that E is our theorist's body of evidence and that $K3$ and LP determine the

---

[26]See Harman (1986), MacFarlane (2004), and Dutilh Novaes (2015) for discussions.

assertability of Γ_R and Γ_P, respectively, on the basis of E. Γ_P cannot be empty, because LP, unlike $K3$, has logical truths. What's more, as mentioned above, Theorem 2.8.3 tells us that Γ_P contains everything that's contained in Γ_R. However, suppose that the only way in which Γ_P goes beyond Γ_R is that Γ_P contains all LP-consequences of Γ_R. Ripley's (2012, §2) theorem then tells us that the set of sentences that are restrictively assertable from Γ_P is just the set of classical consequences of Γ_P. This means that Γ_P and $\Gamma_R \cup \{\alpha \vee \neg\alpha : \alpha$ is an atomic sentence$\}$ are logically equivalent in $K3$. Thus, if we have

$$\Gamma_R \vdash_{K3} \bigwedge\Gamma_P \to \phi$$

then we have

$$\Gamma_R \vdash_{K3} \left(\bigwedge\Gamma_R \wedge (\bigwedge\{\alpha \vee \neg\alpha : \alpha \text{ appears in } \bigwedge\Gamma_R \text{ or } \phi\}) \right) \to \phi.$$

Next, note that

$$\bigwedge\Gamma_R \to \phi \;{}_{K3}\!\dashv\vdash_{K3} \left(\bigwedge\Gamma_R \wedge (\bigwedge\{\alpha \vee \neg\alpha : \alpha \text{ appears in } \bigwedge\Gamma_R \text{ or } \phi\}) \right) \to \phi.$$

The easiest way to see that is by first contraposing both conditionals and pushing through the negations, which gives us

$$\neg\phi \to \bigvee\neg\Gamma_R \;{}_{K3}\!\dashv\vdash_{K3} \neg\phi \to \left(\bigvee\neg\Gamma_R \vee (\bigvee\{\alpha \wedge \neg\alpha : \alpha \text{ appears in } \bigwedge\Gamma_R \text{ or } \phi\}) \right).$$

($\Rightarrow$) then simply follows by disjunction introduction. For ($\Leftarrow$), supposing $\neg\phi \to \neg\Gamma_R$ immediately gives us $\neg\phi \to \neg\Gamma_R$; and supposing $\neg\phi \to \bigvee\{\alpha \wedge \neg\alpha : \alpha$ is atomic and appears in $\bigwedge\Gamma_R$ or $\phi\}$ gives us $\neg\phi \to \neg\Gamma_R$ by explosion. Putting all of this together gives us $\Gamma_R \vdash_{K3} \bigwedge\Gamma_R \to \phi$, and so $\Gamma_R \vdash_{K3} \phi$ by modus ponens. Thus if our theorist's evidence doesn't determine that anything is permissively assertable that isn't restrictively assertable aside from the LP-consequences of what's restrictively assertable, then giving our theorist the ability to permissively assert things doesn't change what she can restrictively assert.

Suppose now instead that Γ_P goes beyond the LP-consequences of Γ_R. In this case, we may have a situation where $\Gamma_R \vdash_{K3} \bigwedge\Gamma_P \to \phi$ but ϕ is not restrictively assertable if our theorist doesn't have the ability to permissively assert things. But once our theorist has this ability, her new logic will determine that ϕ is restrictively assertable, by Theorem 2.8.4, again without changing our theorist's evidence. This may seem a little bit less surprising than the situation we found in the case of glut theory. For, now we don't get that adding the new speech act increases what's restrictively

assertable unless we assume that LP determines that there's some non-logical truth that's permissively assertable on the basis of E. But it's surprising nevertheless. A tempting, though potentially misleading, way to describe the gap theorist's disposition towards non-classical sentences such as the liar is that they can "see" that they are gappy, but they can't express this. Giving the gap theorist the ability to permissively assert things allows them to express what they could already "see" before. But thanks to the resulting new logic, they can now also restrictively assert things that they couldn't restrictively assert before.

In sum, adding a second type of assertion allows our theorists to define a logic according to which they can say more using their original type of assertion based on their evidence than according to the combination of their original logics.

2.9.3 One glap theory or two?

It's time to adress a potential confusion that the way in which we've been talking may generate. On the one hand, we've been talking about the *two* separate relations $\vdash_R$ and $\vDash_P$, but on the other, we've been talking in terms of *one* glap theory, not two. Officially, glap theory (singular!) is the theory of assertability that we get when we have two speech acts of assertion, a restrictive one and a permissive one. But in bringing out the results of adding one of these speech acts to the other, it makes sense to focus on either the relation $\vdash_R$ or the relation $\vdash_P$. It's important to keep in mind, however, that these are not independent relations. To see that, note that they are interdefinable when $\Gamma_P \neq \emptyset$ and $\Gamma_R \neq \emptyset$, as evidenced by the following fact:

Corollary 2.9.1. (a) Suppose $\Gamma_P \neq \emptyset$. Then $\langle \Gamma_R, \Gamma_P \rangle \vdash_P \phi$ iff $\langle \Gamma_R, \neg\phi \rangle \vdash_R \neg \bigwedge \Gamma_P$.

(b) Suppose $\Gamma_R \neq \emptyset$. Then $\langle \Gamma_R, \Gamma_P \rangle \vdash_R \phi$ iff $\langle \neg\phi, \Gamma_P \rangle \vdash_P \neg \bigwedge \Gamma_R$.

Proof. (a). Follows from Theorem 2.8.1 and Theorem 2.8.2 and the fact that the following statements are equivalent:

- $\langle \Gamma_R, \Gamma_P \rangle \vDash_P \phi$.
- For every three-valued $K3/LP$-valuation v, if $v(\gamma_R) = 1$, for all $\gamma_R \in \Gamma_R$, and $v(\bigwedge \Gamma_P) \neq 0$, then $v(\phi) = 1$.

- For every three-valued $K3/LP$-valuation v, if $v(\gamma_R) = 1$, for all $\gamma_R \in \Gamma_R$, and $v(\phi) \neq 1$, then $v(\bigwedge\Gamma_P) = 0$.

- For every three-valued $K3/LP$-valuation v, if $v(\gamma_R) = 1$, for all $\gamma_R \in \Gamma_R$, and $v(\neg\phi) \neq 0$, then $v(\neg\bigwedge\Gamma_P) = 1$.

- $\langle\Gamma_R, \neg\phi\rangle \vDash_R \neg\bigwedge\Gamma_P$.

(b). Follows from Theorem 2.8.1 and Theorem 2.8.2 and the fact that the following statements are equivalent:

- $\langle\Gamma_R, \Gamma_P\rangle \vDash_R \phi$.

- For every three-valued $K3/LP$-valuation v, if $v(\bigwedge\Gamma_R) = 1$ and $v(\gamma_P) \neq 0$, for all $\gamma_P \in \Gamma_P$, then $v(\phi) \neq 0$.

- For every three-valued $K3/LP$-valuation v, if $v(\gamma_P) \neq 1$, for all $\gamma_P \in \Gamma_P$, and $v(\phi) = 0$, then $v(\bigwedge\Gamma_R) \neq 1$.

- For every three-valued $K3/LP$-valuation v, if $v(\gamma_P) \neq 1$, for all $\gamma_P \in \Gamma_P$, and $v(\neg\phi) = 1$, then $v(\neg\bigwedge\Gamma_R) \neq 0$.

- $\langle\Gamma_R, \neg\phi\rangle \vDash_R \neg\bigwedge\Gamma_P$.

$\boxtimes$

2.10   Transparent truth

As discussed in section 2.2, one reason to be interested in non-classical logics is that they may accommodate a transparent truth predicate. Recall that truth predicate Tr is transparent if we can replace $Tr(\ulcorner\phi\urcorner)$ with ϕ in all extensional contexts and vice versa. Glap theory can handle transparency.

In developing the logic that can contain a transparent truth predicate, we will eventually develop a model theory that relies on *acceptable structures* in the sense of Moschovakis (1974, ch. 5) so that we can have standard syntax. The consequence relation defined using acceptable structures won't be compact, and so we won't be able to assume anymore that Γ_R and Γ_P are finite. As a result, Theorem 2.7.1 and Theorem 2.7.2 won't hold anymore. However, using multiple-conclusion logics, we'll be able to obtain close analogues of these propositions. Thus, for the remainder of the chapter,

57

we'll be working in a multiple-conclusion setting.

2.10.1 Multiple-conclusion R and P

We define $\vDash_{R_m}$, the multiple-conclusion analogue of $\vDash_R$: $\langle \Gamma_R, \Gamma_P \rangle \vDash_{P_m} \Delta$ iff every three-valued $K3/LP$-valuation that assigns 1 to every member of Γ_R and 1 or $\frac{1}{2}$ to every member of Γ_P assigns 1 to at least one member of Δ.

And we define $\vDash_{P_m}$, the multiple-conclusion analogue of $\vDash_R$: $\langle \Gamma_R, \Gamma_P \rangle \vDash_{P_m} \Delta$ iff every three-valued $K3/LP$-valuation that assigns 1 to every member of Γ_R and 1 or $\frac{1}{2}$ to every member of Γ_P assigns 1 or $\frac{1}{2}$ to at least one member of Δ.

We have the following analogues of Theorem 2.7.1 and Theorem 2.7.2 (where for Γ a set of sentences, $\neg\Gamma = \{\neg\gamma : \gamma \in \Gamma\}$, and $\vdash_{K3_m}$ is multiple-conclusion $K3$-derivability and $\vdash_{LP_m}$ is multiple-conclusion LP-derivability[27]):

Proposition 2.10.1. $\langle \Gamma_R, \Gamma_P \rangle \vDash_{R_m} \Delta$ iff $\Gamma_R \vDash_{K3_m} \neg\Gamma_P \cup \Delta$.

Proof. The following statements are equivalent:

- $\langle \Gamma_R, \Gamma_P \rangle \vDash_{R_m} \Delta$.
- For every three-valued $K3/LP$-valuation v, if $v(\gamma_{R_i}) = 1$, for all $\gamma_{R_i} \in \Gamma_R$, and $v(\gamma_{P_i}) \neq 0$, for all $\gamma_{P_i} \in \Gamma_P$, then $v(\delta) = 1$, for some $\delta \in \Delta$.
- For every three-valued $K3/LP$-valuation v, if $v(\gamma_{R_i}) = 1$, for all $\gamma_{R_i} \in \Gamma_R$, then $v(\gamma_{P_i}) = 0$, for some $\gamma_{P_i} \in \Gamma_P$, or $v(\delta) = 1$, for some $\delta \in \Delta$.
- For every three-valued $K3/LP$-valuation v, if $v(\gamma_{R_i}) = 1$, for all $\gamma_{R_i} \in \Gamma_R$, then $v(\neg\gamma_{P_i}) = 1$, for some $\neg\gamma_{P_i} \in \neg\Gamma_P$, or $v(\delta) = 1$, for some $\delta \in \Delta$.
- $\Gamma_R \vDash_{K3_m} \neg\Gamma_P \cup \Delta$.

$\boxtimes$

Proposition 2.10.2. $\langle \Gamma_R, \Gamma_P \rangle \vDash_{P_m} \Delta$ iff $\Gamma_P \vDash_{LP_m} \neg\Gamma_R \cup \Delta$.

Proof. The following statements are equivalent:

- $\langle \Gamma_R, \Gamma_P \rangle \vDash_{P_m} \Delta$.

---

[27] See Avron (1991) and Beall (2011, Appendix) for sequent systems for $\vdash_{K3_m}$ and $\vdash_{LP_m}$, respectively.

- For every three-valued $K3/LP$-valuation v, if $v(\gamma_{R_i}) = 1$, for all $\gamma_{R_i} \in \Gamma_R$, and $v(\gamma_{P_i}) \neq 0$, for all $\gamma_{P_i} \in \Gamma_P$, then $v(\delta) \neq 0$, for some $\delta \in \Delta$.

- For every three-valued $K3/LP$-valuation v, if $v(\gamma_{P_i}) \neq 0$, for all $\gamma_{P_i} \in \Gamma_P$, then $v(\gamma_{R_i}) \neq 1$, for some $\gamma_{R_i} \in \Gamma_R$, or $v(\delta) \neq 0$, for some $\delta \in \Delta$.

- For every three-valued $K3/LP$-valuation v, if $v(\gamma_{P_i}) \neq 0$, for all $\gamma_{P_i} \in \Gamma_P$, then $v(\neg\gamma_{R_i}) \neq 0$, for some $\neg\gamma_{R_i} \in \neg\Gamma_R$, or $v(\delta) \neq 0$, for some $\delta \in \Delta$.

- $\Gamma_P \vDash_{LP_m} \neg\Gamma_R \cup \Delta$.

$\boxtimes$

$\vdash_{R_m}$ and $\vdash_{P_m}$ are the smallest relations that are closed under the following axioms and rules (where $\Delta_i = \{\delta_{i_1}, \delta_{i_2}, \dots\}$):

$$(A1_m) \ \frac{}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{R_m} \Delta} \ \text{whenever } \Gamma_R \vdash_{K3_m} \Delta$$

$$(A2_m) \ \frac{}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{P_m} \Delta} \ \text{whenever } \Gamma_R \vdash_{LP_m} \Delta$$

$$(R1_m) \ \frac{\langle\Gamma_R, \Gamma_P\rangle \vdash_{R_m} \neg\Gamma_P \cup \Delta}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{R_m} \Delta} \qquad (R2_m) \ \frac{\langle\Gamma_R, \Gamma_P\rangle \vdash_{P_m} \neg\Gamma_R \cup \Delta}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{P_m} \Delta}$$

Proposition 2.10.3. $\langle\Gamma_R, \Gamma_P\rangle \vDash_{R_m} \Delta$ iff $\langle\Gamma_R, \Gamma_P\rangle \vdash_{R_m} \Delta$.

Proof. ($\Rightarrow$). Suppose $\langle\Gamma_R, \Gamma_P\rangle \vDash_{R_m} \Delta$. Then $\Gamma_R \vDash_{K3_m} \neg\Gamma_P \cup \Delta$, by Theorem 2.10.1, and so $\Gamma_R \vdash_{K3_m} \neg\Gamma_P \cup \Delta$. Then:

$$\begin{array}{l}(A1_m) \\ (R1_m)\end{array} \frac{\dfrac{}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{R_m} \neg\Gamma_P \cup \Delta}}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{R_m} \Delta}$$

($\Leftarrow$). $(A1_m)$ is immediate. For $(R1_m)$, suppose $\langle\Gamma_R, \Gamma_P\rangle \vDash_R \neg\Gamma_P \cup \Delta$ and suppose that some $K3/LP$-valuation v is such that $v(\gamma_R) = 1$, for all $\gamma_R \in \Gamma_R$, and $v(\gamma_P) \neq 0$, for all $\gamma_P \in \Gamma_P$. Then $v(\phi) = 1$, for some $\phi \in \neg\Gamma_P \cup \Delta$. But also $v(\neg\gamma_P) \neq 1$, for all $\neg\gamma_P \in \neg\Gamma_P$. So $v(\delta) = 1$, for some $\delta \in \Delta$, as desired. $\boxtimes$

Proposition 2.10.4. $\langle\Gamma_R, \Gamma_P\rangle \vDash_{P_m} \Delta$ iff $\langle\Gamma_R, \Gamma_P\rangle \vdash_{P_m} \Delta$.

Proof. ($\Rightarrow$). Suppose $\langle\Gamma_R, \Gamma_P\rangle \vDash_{P_m} \Delta$. Then $\Gamma_P \vDash_{LP_m} \neg\Gamma_R \cup \Delta$, by Theorem 2.10.2, and so $\Gamma_P \vdash_{LP_m} \neg\Gamma_R \cup \Delta$. Then:

59

$$(\text{A2}_m) \;\frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{P_m} \neg\Gamma_R \cup \Delta}$$
$$(\text{R2}_m) \;\frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{P_m} \Delta}$$

($\Leftarrow$). (A2$_m$) is immediate. For (R2$_m$), suppose $\langle \Gamma_R, \Gamma_P \rangle \vDash_P \neg\Gamma_R \cup \Delta$ and suppose that some $K3/LP$-valuation v is such that $v(\gamma_R) = 1$, for all $\gamma_R \in \Gamma_R$, and $v(\gamma_P) \neq 0$, for all $\gamma_P \in \Gamma_P$. Then $v(\phi) \neq 0$, for some $\phi \in \neg\Gamma_R \cup \Delta$. But also $v(\neg\gamma_R) = 0$, for all $\neg\gamma_R \in \neg\Gamma_R$. So $v(\delta) \neq 0$, for some $\delta \in \Delta$, as desired. $\boxtimes$

Corollary 2.10.5.   (a) $\langle \Gamma_R, \Gamma_P \rangle \vdash_{R_m} \Delta$ iff $\Gamma_R \vdash_{K3_m} \neg\Gamma_P \cup \Delta$.

  (b) $\langle \Gamma_R, \Gamma_P \rangle \vDash_{P_m} \Delta$ iff $\Gamma_P \vDash_{LP_m} \neg\Gamma_R \cup \Delta$.

Proof. (a) follows from Theorem 2.10.1 and Theorem 2.10.3, (b) follows from Theorem 2.10.2 and Theorem 2.10.4. $\boxtimes$

Corollary 2.10.6. We have the following normal-form theorems: for $I, J \in \{P, R\}$, $I \neq J$, and $i \in \{1, 2\}$, if $\langle \Gamma_R, \Gamma_P \rangle \vdash_{I_m} \Delta$, then there is a canonical proof of this of the form

$$(\text{A}i_m) \;\frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{I_m} \neg\Gamma_J \cup \Delta}$$
$$(\text{R}i_m) \;\frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{I_m} \Delta}$$

Proof. Follows immediately from the proofs of Theorem 2.10.3 and Theorem 2.10.4. $\boxtimes$

2.10.2   Adding a truth predicate

We are finally in a position to add a transparent truth predicate. To do this, we need to move from a propositional language to a quantified language with a truth predicate Tr.

An R/P-model is an acceptable structure $\langle D, I \rangle$, where D is non-empty and where I is subject to the usual constraints for quantified $K3/LP$-models as well as to the constraint that $I(Tr(\ulcorner \phi \urcorner)) = I(\phi)$. Kripke's (1975) fixed-point construction guarantees that there are such models.

$\vDash_{RT}$ is defined so that $\langle \Gamma_R, \Gamma_P \rangle \vDash_{RT} \Delta$ iff every R/P-model $\langle D, I \rangle$ is such that if I assigns 1 to every member of Γ_R and 1 or $\frac{1}{2}$ to every member of Γ_P, then I assigns 1 to at least one member of Δ.

$\models_{PT}$ is defined so that $\langle \Gamma_R, \Gamma_P \rangle \models_{PT} \Delta$ iff every R/P-model $\langle D, I \rangle$ is such that if I assigns 1 to every member of Γ_R and 1 or $\frac{1}{2}$ to every member of Γ_P, then I assigns 1 or $\frac{1}{2}$ to at least one member of Δ.

To define derivability in RT and PT, we also need to define $\models_{K3T}$ and $\models_{LPT}$, i.e. $K3$- and LP-consequence for a language that contains a transparent truth predicate: $\Gamma \models_{K3T} \Delta$ iff every R/P-model $\langle D, I \rangle$ is such that if I assigns 1 to every member of Γ, then I assigns 1 to at least one member of Δ. And $\Gamma \models_{LPT} \Delta$ iff every R/P-model $\langle D, I \rangle$ is such that if I assigns 1 or $\frac{1}{2}$ to every member of Γ, then I assigns 1 or $\frac{1}{2}$ to at least one member of Δ.

Now, $\vdash_{RT}$ and $\vdash_{PT}$ are the smallest relations that are closed under the following axioms and rules:

$$(\text{A1}_T) \; \frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{RT} \Delta} \quad \text{whenever } \Gamma_R \models_{K3T} \Delta$$

$$(\text{A2}_T) \; \frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{PT} \Delta} \quad \text{whenever } \Gamma_R \models_{LPT} \Delta$$

$$(\text{R1}_T) \; \frac{\langle \Gamma_R, \Gamma_P \rangle \vdash_{RT} \neg \Gamma_P \cup \Delta}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{RT} \Delta} \qquad (\text{R2}_T) \; \frac{\langle \Gamma_R, \Gamma_P \rangle \vdash_{PT} \neg \Gamma_R \cup \Delta}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{PT} \Delta}$$

We then have the following:

Proposition 2.10.7. (a) $\langle \Gamma_R, \Gamma_P \rangle \models_{RT} \Delta$ iff $\langle \Gamma_R, \Gamma_P \rangle \vdash_{RT} \Delta$.

(b) $\langle \Gamma_R, \Gamma_P \rangle \models_{PT} \Delta$ iff $\langle \Gamma_R, \Gamma_P \rangle \vdash_{PT} \Delta$.

Proof. Follows immediately from Theorem 2.10.3 and Theorem 2.10.4. ⊠

Of course, even though the axiomatizations of RT and PT are finitely stateable, RT and PT aren't recursively enumerable, because $\models_{K3T}$ and $\models_{LPT}$ aren't recursively enumerable.

2.10.3 Deriving the structural rules

Note that although we don't need any structural rules to define $\vdash_{RT}$ and $\vdash_{PT}$, we can derive the following versions of the usual structural rules of Cut, Right Weakening, and Left Weakening, as well as Identity:[28]

---

[28] As already noted in footnote 24, the choice of sets rather than multisets for the antecedents and succedents automatically yields that $\vdash_{RT}$ and $\vdash_{PT}$ are contractive. A simple model-theoretic argument also shows that analogues of Shoesmith

Proposition 2.10.8.

$$(\text{RC}) \quad \frac{\langle\Gamma_R \cup \{\phi\}, \Gamma_P\rangle \vdash_{RT} \Delta \qquad \langle\Gamma_R, \Gamma_P\rangle \vdash_{RT} \{\phi\} \cup \Delta}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{RT} \Delta}$$

$$(\text{PC}) \quad \frac{\langle\Gamma_R, \Gamma_P \cup \{\phi\}\rangle \vdash_{PT} \Delta \qquad \langle\Gamma_R, \Gamma_P\rangle \vdash_{PT} \{\phi\} \cup \Delta}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{PT} \Delta}$$

$$(\text{RRW}) \quad \frac{\langle\Gamma_R, \Gamma_P\rangle \vdash_{RT} \Delta}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{RT} \Theta \cup \Delta} \qquad\qquad (\text{PRW}) \quad \frac{\langle\Gamma_R, \Gamma_P\rangle \vdash_{PT} \Delta}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{PT} \Theta \cup \Delta}$$

$$(\text{RLW}) \quad \frac{\langle\Gamma_R, \Gamma_P\rangle \vdash_{RT} \Delta}{\langle\Gamma_R \cup \Theta, \Gamma_P\rangle \vdash_{RT} \Delta} \qquad\qquad (\text{PLW}) \quad \frac{\langle\Gamma_R, \Gamma_P\rangle \vdash_{PT} \Delta}{\langle\Gamma_R, \Gamma_P \cup \Theta\rangle \vdash_{PT} \Theta \cup \Delta}$$

$$(\text{RI}) \quad \frac{}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{RT} \Gamma_R} \qquad\qquad (\text{PI}) \quad \frac{}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{PT} \Gamma_P}$$

Proof. (RC). Suppose $\langle\Gamma_R \cup \{\phi\}, \Gamma_P\rangle \vdash_{RT} \Delta$ and $\langle\Gamma_R, \Gamma_P\rangle \vdash_{PT} \{\phi\} \cup \Delta$. Then $\Gamma_R \cup \{\phi\} \vdash_{K3T}$ $\neg\Gamma_P \cup \Delta$ and $\Gamma_R \vdash_{K3T} \neg\Gamma_P \cup \{\phi\} \cup \Delta$, by Theorem 2.10.5.(a). Since we have Cut for $K3T$, we get $\Gamma_R \vdash_{K3T} \neg\Gamma_P \cup \Delta$. Then:

$$\begin{array}{l} (\text{A1}_T) \\ (\text{R1}_T) \end{array} \frac{\langle\Gamma_R, \Gamma_P\rangle \vdash_{RT} \neg\Gamma_P \cup \Delta}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{RT} \Delta}$$

(PC). Suppose $\langle\Gamma_R, \Gamma_P \cup \{\phi\}\rangle \vdash_{PT} \Delta$ and $\langle\Gamma_R, \Gamma_P\rangle \vdash_{PT} \{\phi\} \cup \Delta$. Then $\Gamma_P \cup \{\phi\} \vdash_{LPT}$ $\neg\Gamma_R \cup \Delta$ and $\Gamma_P \vdash_{LPT} \neg\Gamma_R \cup \{\phi\} \cup \Delta$, by Theorem 2.10.5.(b). Since we have Cut for LPT, we get $\Gamma_P \vdash_{LPT} \neg\Gamma_R \cup \Delta_2$. Then:

$$\begin{array}{l} (\text{A2}_T) \\ (\text{R2}_m) \end{array} \frac{\langle\Gamma_R, \Gamma_P\rangle \vdash_{PT} \neg\Gamma_R \cup \Delta}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{PT} \Delta}$$

(RRW). Suppose $\langle\Gamma_R, \Gamma_P\rangle \vdash_{RT} \Delta$. Then $\Gamma_R \vdash_{K3T} \neg\Gamma_P \cup \Delta$, by Theorem 2.10.5.(a). Since we have Right Weakening for $K3T$, we get $\Gamma_R \vdash_{K3T} \Theta \cup \neg\Gamma_P \cup \Delta$. Then:

---

and Smiley's (1978, 29) Cut for Sets are valid in RT and PT, from which it follows by Theorem 2.10.7 that we have the following:

$$(\text{RCS}) \quad \frac{\langle\Gamma_R \cup \Theta_1, \Gamma_P\rangle \vdash_{RT} \Theta_2 \cup \Delta, \text{ for each partition } \Theta_1, \Theta_2 \text{ of } \Theta}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{RT} \Delta}$$

$$(\text{PCS}) \quad \frac{\langle\Gamma_R, \Gamma_P \cup \Theta_1\rangle \vdash_{PT} \Theta_2 \cup \Delta, \text{ for each partition } \Theta_1, \Theta_2 \text{ of } \Theta}{\langle\Gamma_R, \Gamma_P\rangle \vdash_{PT} \Delta}$$

$$(\text{A1}_T) \; \frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{RT} \Theta \cup \neg \Gamma_P \cup \Delta}$$
$$(\text{R1}_T) \; \frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{RT} \Theta \cup \Delta}$$

(PRW). Suppose $\langle \Gamma_R, \Gamma_P \rangle \vdash_{PT} \Delta$. Then $\Gamma_P \vdash_{LPT} \neg \Gamma_R \cup \Delta$, by Theorem 2.10.5.(b). Since we have Right Weakening for LPT, we get $\Gamma_P \vdash_{LPT} \Theta \cup \neg \Gamma_R \cup \Delta$. Then:

$$(\text{A2}_T) \; \frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{PT} \Theta \cup \neg \Gamma_R \cup \Delta}$$
$$(\text{R2}_T) \; \frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{PT} \Theta \cup \Delta}$$

(RLW). Suppose $\langle \Gamma_R, \Gamma_P \rangle \vdash_{RT} \Delta$. Then $\Gamma_R \vdash_{K3T} \neg \Gamma_P \cup \Delta$, by Theorem 2.10.5.(a). Since we have Left Weakening for $K3T$, we get $\Gamma_R \cup \Theta \vdash_{K3T} \neg \Gamma_P \cup \Delta$.Then:

$$(\text{A1}_T) \; \frac{}{\langle \Gamma_R \cup \Theta, \Gamma_P \rangle \vdash_{RT} \neg \Gamma_P \cup \Delta}$$
$$(\text{R1}_T) \; \frac{}{\langle \Gamma_R \cup \Theta, \Gamma_P \rangle \vdash_{RT} \Delta}$$

(PLW). Suppose $\langle \Gamma_R, \Gamma_P \rangle \vdash_{PT} \Delta$. Then $\Gamma_P \vdash_{LPT} \neg \Gamma_R \cup \Delta$, by Theorem 2.10.5.(b). Since we have Left Weakening for LPT, we get $\Gamma_P \cup \Theta \vdash_{LPT} \neg \Gamma_R \cup \Delta$. Then:

$$(\text{A2}_T) \; \frac{}{\langle \Gamma_R, \Gamma_P \cup \Theta \rangle \vdash_{PT} \neg \Gamma_R \cup \Delta}$$
$$(\text{R2}_T) \; \frac{}{\langle \Gamma_R, \Gamma_P \cup \Theta \rangle \vdash_{PT} \Delta}$$

(RI). Follows immediately from (A1_T) and the fact that $\Gamma_R \vdash_{K3T} \Gamma_R$.

(PI). Follows immediately from (A2_T) and the fact that $\Gamma_P \vdash_{LPT} \Gamma_P$. $\boxtimes$

2.10.4 Comparisons

What distinguishes glap theory from all other non-classical or substructural approaches to transparent truth is that its logic assumes that there are two premise sets instead of one and that there are two consequence relations and two derivability relations. As already discussed at length, this additional structure gives us a logic that's stronger than $K3$ and LP combined, but it retains the latter's relative simplicity vis-à-vis the logics of Field (2008) and Beall (2009).

We also just saw in subsection 2.10.3 that the logic of glap theory is fully structural in that it includes the rule of Cut for both $\vdash_{RT}$ and $\vdash_{PT}$. This contrasts with the Cut-free logic $STTT$ of Cobreros et al. (2013).[29] Since the absence of Cut means that $\vdash_{STTT}$ isn't transitive, this means that we can't string together proofs in $STTT$. There is thus a concrete sense in which Feferman's

---

[29]See also Ripley (2012, 2013a,b). As mentioned in footnotes 24 and 28, glap theory is also contractive, unlike the logic developed by Zardini (2011).

complaint applies here too: ordinary reasoning can't be *sustained* in $STTT$. Of course, in giving up Cut, $STTT$ gains the ability to have transparent truth while preserving all classical validities. Thus, while in glap theory we don't have restrictive identity ($\nvdash_{RT} \phi \to \phi$) and we only have one non-hybrid form of modus ponens ($\langle\{\phi \to \psi, \phi\}, \emptyset\rangle \vdash_{RT} \psi$ but $\langle\emptyset, \{\phi \to \psi, \phi\}\rangle \nvdash_{PT} \psi$), we have unqualified identity and modus ponens in $STTT$ ($\vdash_{STTT} \phi \to \phi$ and $\{\phi \to \psi, \phi\} \vdash_{STTT} \psi$).

While it's not immediately obvious that the advantages of glap theory outweigh its disadvantages vis-à-vis the approaches of Field (2008), Beall (2009), and Cobreros et al. (2013), glap theory is an unqualified improvement over an approach recently recommended by Beall (2017). As we saw at the end of section 2.3, gap theorists and glut theorists have expressive reasons to help themselves to extra speech acts, be it denial or a second kind of assertion. But once they enrich their expresisve resources, they become functionally indistinguishable. For, they now both have a permissive speech act governed by a paraconsistent logic and a restrictive speech act governed by a paracomplete logic, and they can express all the same things. This might be seen to show that the choice between gap and glut theory is arbitrary.[30] Beall concludes from this that we should opt for Anderson and Belnap's (1961) *four*-valued logic of First Degree Entailment (FDE), which is weaker than $K3$ and LP combined.[31] While FDE is exceedingly weak, Beall (2015a) would argue that we can always assert the law of identity for certain domains or deny that certain domains are inconsistent to recapture stronger modes of reasoning. But our discussion shows that this reaction to the fact that we lack a clear tie-breaker to choose between $K3$ and LP is unwarranted. Rather than admitting both gaps and gluts, the champion of either gaps or gluts who doesn't want to bear the cost of adding a new conditional or giving up structural rules should embrace glaps, as the logic of glap theory is *stronger* than both $K3$ and LP.

2.11 Subervaluationism

Because the logic of gaps and the logic of gluts are so weak, van Fraassen (1966) and Fine (1975b) opt for supervaluationism, and Varzi (1994, 1997) and Hyde (1997) opt for subvaluationism. Super-

---

[30]See also Parsons (1984).
[31]See Woodruff (1984) for the theory of Kripke fixed points in an FDE setting.

and subvaluationism have in common that they associate with each $K3/LP$-model a set of admissible classical models where the extensions and anti-extensions of vague predicates are precisified so as to yield extensions and anti-extensions that don't overlap and that exhaust the domain. Supervaluationism introduces the concept of *supertruth*, which universally quantifies over the admissible classical models, while subvaluationism introduces the concept of *subtruth*, which existentially quantifies over the admissible classical models. We can use these ideas to do the same work for glap theory, yielding su*b*ervaluationism.[32]

2.11.1 Supervaluationism and subvaluationism

But first, a quick review of supervaluationism and subvaluationism. We'll work with a standard quantified language without a truth predicate and the class $\mathfrak{M}$ of all classical models for the language. An *Sv-model* is a non-empty set $\mathcal{M} \subseteq \mathfrak{M}$.[33] We say that ϕ is *supertrue* in $\mathcal{M}$ iff it is true in every classical model $\mathscr{M} \in \mathcal{M}$ and ϕ is *subtrue* in $\mathcal{M}$ iff it is true in some classical model $\mathscr{M} \in \mathcal{M}$.

We can now define the two usual consequence relations $\vDash_{K3V}$ and $\vDash_{LP_V}$ of supervaluationism and subvaluationism, respectively, as well as the classical consequence relation $\vDash_{CL}$: $\Gamma \vDash_{K3V} \Delta$ iff, for every Sv-model $\mathcal{M}$, if every member of Γ is supertrue in $\mathcal{M}$, then some member of Δ is supertrue in $\mathcal{M}$. And $\Gamma \vDash_{LP_V} \Delta$ iff, for every Sv-model $\mathcal{M}$, if every member of Γ is subtrue in $\mathcal{M}$, then some member of Δ is subtrue in $\mathcal{M}$. And $\Gamma \vDash_{CL} \Delta$ iff, for every $\mathscr{M} \in \mathfrak{M}$, if every member of Γ is true in $\mathscr{M}$, then some member of Δ is true in $\mathscr{M}$.

$K3^V$ and LP_V come quite close to being fully classical, for Hyde (1997, 655) tells us that the following correspondences hold:

Fact 2.11.1. Let Γ and Δ be finite. Then:

(a) $\Gamma \vDash_{CL} \Delta$ iff $\Gamma \vDash_{K3V} \bigvee \Delta$.

(b) $\Gamma \vDash_{CL} \Delta$ iff $\bigwedge \Gamma \vDash_{LP_V} \Delta$.

Thus, for arguments with just one premise ϕ and just one conclusion ψ, we have that $\phi \vDash_{CL} \psi$ iff $\phi \vDash_{K3V} \psi$ iff $\phi \vDash_{LP_V} \psi$. We also have the following for any Γ and Δ, finite or infinite.

---

[32]*Not* named after Peter Suber, the editor of Bertlett and Suber (1987).

[33]I loosely follow the presentation in Cobreros et al. (2012a) here.

Proposition 2.11.1.   (a) $\Gamma \vDash_{K3V} \Delta$ iff $\Gamma \vDash_{CL} \delta$, for some $\delta \in \Delta$.

(b) $\Gamma \vDash_{LP_V} \Delta$ iff $\gamma \vDash_{CL} \Delta$, for some $\gamma \in \Gamma$.

Proof. (a). ($\Rightarrow$). Suppose $\Gamma \nvDash_{CL} \delta$, for every $\delta \in \Delta$. Then for every $\delta \in \Delta$, there is a classical model $\mathcal{M} \in \mathfrak{M}$ such that every $\gamma \in \Gamma$ is true in $\mathcal{M}$ and δ is false in $\mathcal{M}$. Let M be the set of all those classical models. M is an Sv-model. Every $\gamma \in \Gamma$ is supertrue in M, but no $\delta \in \Delta$ is supertrue in M, and so $\Gamma \nvDash_{K3V} \Delta$. ($\Leftarrow$). Suppose $\Gamma \vDash_{CL} \delta$, for some $\delta \in \Delta$. Then we immediately have that for any Sv-model M, if every $\gamma \in \Gamma$ is supertrue in M, then there is some $\delta \in \Delta$ that is supertrue in M, and so $\Gamma \vDash_{K3V} \Delta$.

(b). ($\Rightarrow$). Suppose $\gamma \nvDash_{CL} \Delta$, for every $\gamma \in \Gamma$. Then for every $\gamma \in \Gamma$, there is a classical model $\mathcal{M} \in \mathfrak{M}$ such that γ is true in $\mathcal{M}$ and no $\delta \in \Delta$ is true in $\mathcal{M}$. Let M be the set of all those classical models. M is an Sv-model. Every $\gamma \in \Gamma$ is subtrue in M, but no $\delta \in \Delta$ is subtrue in M, and so $\Gamma \nvDash_{LP_V} \Delta$. ($\Leftarrow$). Suppose $\gamma \vDash_{CL} \Delta$, for some $\gamma \in \Gamma$. Then we immediately have that for any Sv-model M, if some $\gamma \in \Gamma$ is subtrue in M, then there is some $\delta \in \Delta$ that is subtrue in M, and so $\Gamma \vDash_{LP_V} \Delta$.   ⊠

But we still have certain failures of classicality. While we have $\emptyset \vDash_{K3V} \phi \vee \neg\phi$, we also have $\emptyset \nvDash_{K3V} \{\phi, \neg\phi\}$. And while we have $\phi \wedge (\phi \rightarrow \psi) \vDash_{LP_V} \psi$ and $\phi \wedge \neg\phi \vDash_{LP_V} \emptyset$, we also have $\{\phi, \phi \rightarrow \psi\} \nvDash_{LP_V} \psi$ and $\{\phi, \neg\phi\} \nvDash_{LP_V} \emptyset$. More generally, we have $\phi \vee \psi \nvDash_{K3V} \{\phi, \psi\}$ and $\{\phi, \psi\} \nvDash_{LP_V} \phi \wedge \psi$.

2.11.2   Subervaluationist consequence and derivability

We can improve on this using ideas analogous to those of glap theory. We define the restrictive and permissive consequence relations $\vDash_{RV}$ and $\vDash_{PV}$ of subervaluationism: $\langle \Gamma_R, \Gamma_P \rangle \vDash_{RV} \Delta$ iff, for every Sv-model M, if every member of Γ_R is supertrue in M and every member of Γ_P is subtrue in M, then some member of Δ is supertrue in M. And $\langle \Gamma_R, \Gamma_P \rangle \vDash_{PV} \Delta$ iff, for every Sv-model M, if every member of Γ_R is supertrue in M and every member of Γ_P is subtrue in M, then some member of Δ is subtrue in M. We then get analogues of Theorem 2.10.1 and Theorem 2.10.2.

Proposition 2.11.2. $\langle \Gamma_R, \Gamma_P \rangle \vDash_{RV} \Delta$ iff $\Gamma_R \vDash_{K3V} \neg\Gamma_P \cup \Delta$.

Proof. The following statements are equivalent:

- $\langle \Gamma_R, \Gamma_P \rangle \vDash_{RV} \Delta$.

- For every Sv-model $\mathcal{M}$, if every member of Γ_R is supertrue in $\mathcal{M}$ and every member of Γ_P is subtrue in $\mathcal{M}$, then some member of Δ is supertrue in $\mathcal{M}$.

- For every Sv-model $\mathcal{M}$, if every member of Γ_R is supertrue in $\mathcal{M}$, then some member of Γ_P isn't subtrue in $\mathcal{M}$ or some member of Δ is supertrue in $\mathcal{M}$.

- For every Sv-model $\mathcal{M}$, if every member of Γ_R is supertrue in $\mathcal{M}$, then some member of $\neg\Gamma_P$ is supertrue in $\mathcal{M}$ or some member of Δ is supertrue in $\mathcal{M}$.

- $\Gamma_R \vDash_{K3^V} \neg\Gamma_P \cup \Delta$.

$$\boxtimes$$

Proposition 2.11.3. $\langle \Gamma_R, \Gamma_P \rangle \vDash_{PV} \Delta$ iff $\Gamma_P \vDash_{LP_V} \neg\Gamma_R \cup \Delta$.

Proof. The following statements are equivalent:

- $\langle \Gamma_R, \Gamma_P \rangle \vDash_{PV} \Delta$.

- For every Sv-model $\mathcal{M}$, if every member of Γ_R is supertrue in $\mathcal{M}$ and every member of Γ_P is subtrue in $\mathcal{M}$, then some member of Δ is subtrue in $\mathcal{M}$.

- For every Sv-model $\mathcal{M}$, if every member of Γ_P is subtrue in $\mathcal{M}$, then some member of Γ_R isn't supertrue in $\mathcal{M}$ or some member of Δ is subtrue in $\mathcal{M}$.

- For every Sv-model $\mathcal{M}$, if every member of Γ_P is subtrue in $\mathcal{M}$, then some member of $\neg\Gamma_R$ is subtrue in $\mathcal{M}$ or some member of Δ is subtrue in $\mathcal{M}$.

- $\Gamma_P \vDash_{LP_V} \neg\Gamma_R \cup \Delta$.

$$\boxtimes$$

Next, we define derivability for subvaluationism. $\vdash_{RV}$ and $\vdash_{PV}$ are the smallest relations that are closed under the following axioms and rules (where $\vdash_{CL}$ is classical derivability):

$$(A1_V) \frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{RV} \Delta} \text{ whenever } \Gamma_R \vdash_{CL} \delta, \text{ for some } \delta \in \Delta$$

$$(A2_V) \frac{}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{PV} \Delta} \text{ whenever } \gamma \vdash_{CL} \Delta, \text{ for some } \gamma \in \Gamma_P$$

$$(\text{R1}_V) \; \frac{\langle \Gamma_R, \Gamma_P \rangle \vdash_{RV} \neg \Gamma_P \cup \Delta}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{RV} \Delta} \qquad\qquad (\text{R2}_V) \; \frac{\langle \Gamma_R, \Gamma_P \rangle \vdash_{PV} \neg \Gamma_R \cup \Delta}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{PV} \Delta}$$

Proposition 2.11.4. $\langle \Gamma_R, \Gamma_P \rangle \vDash_{RV} \Delta$ iff $\langle \Gamma_R, \Gamma_P \rangle \vdash_{RV} \Delta$.

Proof. ($\Rightarrow$). Suppose $\langle \Gamma_R, \Gamma_P \rangle \vDash_{RV} \Delta$. Then $\Gamma_R \vDash_{K3V} \neg \Gamma_P \cup \Delta$, by Theorem 2.11.2, and so $\Gamma_R \vdash_{K3V} \neg \Gamma_P \cup \Delta$. Then:

$$\begin{array}{l} (\text{A1}_V) \; \dfrac{}{} \\ (\text{R1}_V) \; \dfrac{\langle \Gamma_R, \Gamma_P \rangle \vdash_{RV} \neg \Gamma_P \cup \Delta}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{RV} \Delta} \end{array}$$

($\Leftarrow$). (A1$_V$) is immediate, given Theorem 2.11.1.(a) and the fact that $\Gamma \vDash_{CL} \Delta$ iff $\Gamma \vdash_{CL} \Delta$. For (R1$_V$), suppose $\langle \Gamma_R, \Gamma_P \rangle \vDash_R \neg \Gamma_P \cup \Delta$ and suppose that some Sv-model $\mathcal{M}$ is such that every member of Γ_R is supertrue in $\mathcal{M}$ and every member of Γ_P is subtrue in $\mathcal{M}$. Then some member of $\neg \Gamma_P \cup \Delta$ is supertrue in $\mathcal{M}$. But since every member of Γ_P is subtrue in $\mathcal{M}$, no member of $\neg \Gamma_P$ is supertrue in $\mathcal{M}$, and so some member of Δ is supertrue in $\mathcal{M}$, as desired. $\boxtimes$

Proposition 2.11.5. $\langle \Gamma_R, \Gamma_P \rangle \vDash_{PV} \Delta$ iff $\langle \Gamma_R, \Gamma_P \rangle \vdash_{PV} \Delta$.

Proof. ($\Rightarrow$). Suppose $\langle \Gamma_R, \Gamma_P \rangle \vDash_{PV} \Delta$. Then $\Gamma_P \vDash_{LP_V} \neg \Gamma_R \cup \Delta$, by Theorem 2.11.3, and so $\Gamma_P \vdash_{LP_V} \neg \Gamma_R \cup \Delta$. Then:

$$\begin{array}{l} (\text{A2}_V) \; \dfrac{}{} \\ (\text{R2}_V) \; \dfrac{\langle \Gamma_R, \Gamma_P \rangle \vdash_{PV} \neg \Gamma_R \cup \Delta}{\langle \Gamma_R, \Gamma_P \rangle \vdash_{PV} \Delta} \end{array}$$

($\Leftarrow$). (A2$_V$) is immediate, given Theorem 2.11.1.(b) and the fact that $\Gamma \vDash_{CL} \Delta$ iff $\Gamma \vdash_{CL} \Delta$. For (R2$_V$), suppose $\langle \Gamma_R, \Gamma_P \rangle \vDash_P \neg \Gamma_P \cup \Delta$ and suppose that some Sv-model $\mathcal{M}$ is such that every member of Γ_R is supertrue in $\mathcal{M}$ and every member of Γ_P is subtrue in $\mathcal{M}$. Then some member of $\neg \Gamma_R \cup \Delta$ is subtrue in $\mathcal{M}$. But since every member of Γ_R is supertrue in $\mathcal{M}$, no member of $\neg \Gamma_R$ is subtrue in $\mathcal{M}$, and so some member of Δ is subtrue in $\mathcal{M}$, as desired. $\boxtimes$

Corollary 2.11.6.   (a) $\langle \Gamma_R, \Gamma_P \rangle \vdash_{RV} \Delta$ iff $\Gamma_R \vdash_{K3V} \neg \Gamma_P \cup \Delta$.

  (b) $\langle \Gamma_R, \Gamma_P \rangle \vdash_{PV} \Delta$ iff $\Gamma_P \vdash_{LP_V} \neg \Gamma_R \cup \Delta$.

Proof. (a) follows from Theorem 2.11.2 and Theorem 2.11.4, (b) follows from Theorem 2.11.3 and Theorem 2.11.5. $\boxtimes$

2.11.3 Discussion

Although just like in the case of $K3^V$ and LP_V we have $\langle \emptyset, \emptyset \rangle \nvDash_{RV} \{\phi, \neg\phi\}$, $\langle \emptyset, \{\phi, \phi \to \psi\} \rangle \nvDash_{PV}$ ψ, $\langle \emptyset, \{\phi, \neg\phi\} \rangle \nvDash_{PV} \emptyset$, $\langle \phi \vee \psi, \emptyset \rangle \nvDash_{RV} \{\phi, \psi\}$, and $\langle \emptyset, \{\phi, \psi\} \rangle \nvDash_{PV} \phi \wedge \psi$, we do have the following variants:

Proposition 2.11.7.   (a) $\langle \Gamma_R, \Gamma_P \rangle \vdash_{PV} \{\phi, \neg\phi\}$

  (b) $\langle \phi, \phi \to \psi \rangle \vdash_{PV} \psi$

  (c) $\langle \{\phi, \neg\phi\}, \Gamma_P \rangle \vdash_{RV} \emptyset$

  (d) $\langle \phi \vee \psi, \Gamma_P \rangle \vdash_{PV} \{\phi, \psi\}$

  (e) $\langle \phi, \psi \rangle \vdash_{PV} \phi \wedge \psi$

Proof. It's straightforward to check that the analogues hold for $\vDash_{RV}$ and $\vDash_{RV}$, and so the result follows from Theorem 2.11.4 and Theorem 2.11.5.   ⊠

We can identify precisely how classical logic relates to RV and PV:

Proposition 2.11.8.   (a) $\langle \Gamma_R, \Gamma_P \rangle \vdash_{RV} \phi$ iff $\Gamma_R \vdash_{CL} \phi$ or $\Gamma_R \vdash_{CL} \neg\gamma_P$, for some $\gamma_P \in \Gamma_P$.

  (b) $\langle \Gamma_R, \Gamma_P \rangle \vdash_{PV} \phi$ iff $\Gamma_R \vdash_{CL} \gamma_P \to \phi$, for some $\gamma_P \in \Gamma_P$.

  (c) $\langle \Gamma_R, \Gamma_P \rangle \vdash_{RV} \Delta$ iff $\Gamma_R \vdash_{CL} \delta$, for some $\delta \in \Delta$, or $\Gamma_R \vdash_{CL} \neg\gamma_P$, for some $\gamma_P \in \Gamma_P$.

  (d) $\langle \Gamma_R, \Gamma_P \rangle \vdash_{PV} \Delta$ iff $\Gamma_R \vdash_{CL} \gamma_P \to \delta$, for some $\gamma_P \in \Gamma_P$ and some $\delta \in \Delta$.

Proof. (a). ($\Rightarrow$). Suppose $\Gamma_R \nvdash_{CL} \phi$ and $\Gamma_R \nvdash_{CL} \neg\gamma_P$, for every $\gamma_P \in \Gamma_P$. Then $\Gamma_R \nvDash_{CL} \phi$ and $\Gamma_R \nvDash_{CL} \neg\gamma_P$, for every $\gamma_P \in \Gamma_P$. The latter tells us that for every $\gamma_P \in \Gamma_P$, there is a classical model $\mathscr{M} \in \mathfrak{M}$ such that every $\gamma_R \in \Gamma_R$ is true in $\mathscr{M}$ and γ_P is true in $\mathscr{M}$. Let $\mathcal{M}$ be the set of all those classical models. $\Gamma_R \nvDash_{CL} \phi$ tells us that there is a classical model $\mathscr{M}' \in \mathfrak{M}$ such that every $\gamma_R \in \Gamma_R$ is true in $\mathscr{M}'$ and ϕ is false in $\mathscr{M}'$. $\mathcal{M} \cup \{\mathscr{M}'\}$ is an Sv-model. Every $\gamma_R \in \Gamma_R$ is supertrue in $\mathcal{M} \cup \{\mathscr{M}'\}$, every $\gamma_P \in \Gamma_P$ is subtrue in $\mathcal{M} \cup \{\mathscr{M}'\}$, and ϕ is not supertrue in $\mathcal{M} \cup \{\mathscr{M}'\}$, and so $\langle \Gamma_R, \Gamma_P \rangle \nvDash_{RV} \phi$. Then $\langle \Gamma_R, \Gamma_P \rangle \nvdash_{RV} \phi$, by Theorem 2.11.4. ($\Leftarrow$). Suppose

first $\Gamma_R \vdash_{CL} \phi$. Then $\langle \Gamma_R, \Gamma_P \rangle \vdash_{RV} \phi$, by (A1$_V$). Suppose next $\Gamma_R \vdash_{CL} \neg\gamma_P$, for some $\gamma_P \in \Gamma_P$. Then $\Gamma_R \vDash_{CL} \neg\gamma_P$. Then there is no classical model $\mathcal{M} \in \mathfrak{M}$ such that every $\gamma_R \in \Gamma_R$ is true in $\mathcal{M}$ and γ_P is true in $\mathcal{M}$. Then there is no Sv-model $\mathcal{M}$ such that every $\gamma_R \in \Gamma_R$ is supertrue in $\mathcal{M}$ and every $\gamma_P \in \Gamma_P$ is subtrue in $\mathcal{M}$, and so $\langle \Gamma_R, \Gamma_P \rangle \nvDash_{RV} \phi$. Then $\langle \Gamma_R, \Gamma_P \rangle \nvdash_{RV} \phi$, by Theorem 2.11.4.

(b). ($\Rightarrow$). Suppose $\Gamma_R \nvdash_{CL} \gamma_P \to \phi$, for every $\gamma_P \in \Gamma_P$. Then $\Gamma_R \nvDash_{CL} \gamma_P \to \phi$, for every $\gamma_P \in \Gamma_P$. Then for every $\gamma_P \in \Gamma_P$, there is a classical model $\mathcal{M} \in \mathfrak{M}$ such that every $\gamma_R \in \Gamma_R$ is true in $\mathcal{M}$, γ_P is true in $\mathcal{M}$, and ϕ is false in $\mathcal{M}$. Let $\mathcal{M}$ be the set of all those classical models. $\mathcal{M}$ is an Sv-model. Every $\gamma_R \in \Gamma_R$ is supertrue in $\mathcal{M}$, every $\gamma_P \in \Gamma_P$ is subtrue in $\mathcal{M}$, and ϕ is not subtrue in $\mathcal{M}$, and so $\langle \Gamma_R, \Gamma_P \rangle \nvDash_{PV} \phi$. Then $\langle \Gamma_R, \Gamma_P \rangle \nvdash_{PV} \phi$, by Theorem 2.11.5. ($\Leftarrow$). Suppose $\Gamma_R \vdash_{CL} \gamma_P \to \phi$, for some $\gamma_P \in \Gamma_P$. Then $\Gamma_R \vDash_{CL} \gamma_P \to \phi$. Then every classical model $\mathcal{M} \in \mathfrak{M}$ is such that if every $\gamma_R \in \Gamma_R$ is true in $\mathcal{M}$ and γ_P is true in $\mathcal{M}$, then ϕ is true in $\mathcal{M}$. Let $\mathcal{M}$ be an arbitrary Sv-model such that every $\gamma_R \in \Gamma_R$ is supertrue in $\mathcal{M}$ and every $\gamma'_P \in \Gamma_P$ is subtrue in $\mathcal{M}$. Then there is a classical model $\mathcal{M}' \in \mathcal{M}$ such that γ_P is true in $\mathcal{M}'$. Then, given the above, ϕ is true in $\mathcal{M}'$. So ϕ is subtrue in $\mathcal{M}$. Since $\mathcal{M}$ was arbitrary, $\langle \Gamma_R, \Gamma_P \rangle \vDash_{PV} \phi$, and so $\langle \Gamma_R, \Gamma_P \rangle \vdash_{PV} \phi$, by Theorem 2.11.5.

(c). Analogous to the proof of (a).

(d). Analogous to the proof of (b). $\boxtimes$

2.11.4   Truth

We can add a truth predicate to subervaluationism in much the same way that Kripke (1975) adds a truth predicate to supervaluationism to obtain $\vdash_{RVT}$ and $\vdash_{PVT}$.

2.12   Comparing glap theory and subervaluationism

To wrap up, we compare glap theory and subervaluationism with regards to transparency, tolerance, and classicality.

Transparency. Where Γ'_R, Γ'_P, and Δ' are just like Γ_R, Γ_P, and Δ, respectively, except that some

subsentences ϕ of some members of Γ_R, Γ_P, and Δ are replaced with $Tr(\ulcorner\phi\urcorner)$:

$$\frac{\langle\Gamma_R,\Gamma_P\rangle \vdash_X \Delta}{\langle\Gamma'_R,\Gamma'_P\rangle \vdash_X \Delta'}$$

T-biconditionals. $\langle\emptyset,\emptyset\rangle \vdash_X \phi \leftrightarrow Tr(\ulcorner\phi\urcorner)$.

Left tolerance instances. $\langle\{$'Isaiah Thomas is short,' 'If Isaiah Thomas is short, then Michael Adams is short,' 'If Michael Adams is short, then Terrell Brandon is short,' ..., 'If Ralph Sampson is short, then Yao Ming is short'$\},\emptyset\rangle \nvdash_X \{$'Yao Ming is short'$\}$.

Right tolerance instances. $\langle\emptyset,\{$'Isaiah Thomas is short,' 'If Isaiah Thomas is short, then Michael Adams is short,' 'If Michael Adams is short, then Terrell Brandon is short,' ..., 'If Ralph Sampson is short, then Yao Ming is short'$\}\rangle \nvdash_X \{$'Yao Ming is short'$\}$.

Left tolerance conjunction. $\langle\{$'Isaiah Thomas is short,' 'If Isaiah Thomas is short, then Michael Adams is short, and if Michael Adams is short, then Terrell Brandon is short, and ...and if Ralph Sampson is short, then Yao Ming is short'$\},\emptyset\rangle \nvdash_X \{$'Yao Ming is short'$\}$.

Right tolerance instances. $\langle\emptyset,\{$'Isaiah Thomas is short,' 'If Isaiah Thomas is short, then Michael Adams is short, and if Michael Adams is short, then Terrell Brandon is short, and ...and if Ralph Sampson is short, then Yao Ming is short'$\}\rangle \nvdash_X \{$'Yao Ming is short'$\}$.

Left single-conclusion classicality. $\langle\Gamma,\emptyset\rangle \vdash_X \phi$ iff $\Gamma \vdash_{CL} \phi$.

Right single-conclusion classicality. $\langle\emptyset,\Gamma\rangle \vdash_X \phi$ iff $\Gamma \vdash_{CL} \phi$.

Left single-premise classicality. $\langle\phi,\emptyset\rangle \vdash_X \Delta$ iff $\phi \vdash_{CL} \Delta$.

Right single-premise classicality. $\langle\emptyset,\phi\rangle \vdash_X \Delta$ iff $\phi \vdash_{CL} \Delta$.

Left full classicality. $\langle\Gamma,\emptyset\rangle \vdash_X \Delta$ iff $\Gamma \vdash_{CL} \Delta$.

Right full classicality. $\langle\emptyset,\Gamma\rangle \vdash_X \Delta$ iff $\Gamma \vdash_{CL} \Delta$.

	Glap theory		Subervaluationism	
	$X = RT$	$X = PT$	$X = RVT$	$X = PVT$
Transparency	✓	✓	✗	✗
T-biconditionals	✗	✓	✓	✓
Left tolerance instances	✗	✗	✗	✗
Right tolerance instances	✗	✓	✗	✓
Left tolerance conjunction	✗	✗	✗	✗
Right tolerance conjunction	✗	✓	✗	✗
Left single-conclusion classicality	✗	✓	✓	✗
Right single-conclusion classicality	✗	✗	✗	✗
Left single-premise classicality	✗	✓	✗	✓
Right single-premise classicality	✗	✗	✗	✓
Left full classicality	✗	✓	✗	✓
Right full classicality	✗	✗	✗	✗

It is straightforward to check most of these with model-theoretic reasoning. Transparency / $X = RVT$ and Transparency $X = PVT$ fail to hold because in subervaluationism, the logical connectives aren't supertruth-/subtruth functional. Left full classicality / $X = PT$ and Left full classicality / $X = PVT$ are proved in Cobreros et al. (2012b) and Cobreros et al. (2012a), respectively. It is because these two hold that Left tolerance instances / $X = PT$, Left tolerance conjunction / $X = PT$, and Left tolerance instances / $X = PVT$ don't hold.

Finally, let's compare all possible hybrid forms of modus ponens that can be formulated for glap theory and subervaluationism.

$$(\text{MP1}) \quad \frac{\langle \Gamma_R, \Gamma_P \rangle \vdash_X \phi \to \psi \qquad \langle \Gamma_R, \Gamma_P \rangle \vdash_Y \phi}{\langle \Gamma_R, \Gamma_P \rangle \vdash_X \psi}$$

$$(\text{MP2}) \quad \frac{\langle \Gamma_R, \Gamma_P \rangle \vdash_X \phi \to \psi \qquad \langle \Gamma_R, \Gamma_P \rangle \vdash_Y \phi}{\langle \Gamma_R, \Gamma_P \rangle \vdash_Y \psi}$$

	Glap theory		Subervaluationism	
	$X = RT$ $Y = PT$	$X = PT$ $Y = RT$	$X = RVT$ $Y = PVT$	$X = PVT$ $Y = RVT$
(MP1)	✓	✓	✗	✓
(MP2)	✓	✗	✓	✗

(MP1) / $X = RT, Y = PT$ and (MP1) / $X = PT, Y = RT$ are valid since they are just the two rules (R1) and (R2) for $\vdash_R$ and $\vdash_P$. And (MP2) / $X = RT, Y = PT$ follows from (MP1) / $X = RT, Y = PT$ by Theorem 2.8.3. To see that (MP2) / $X = PT, Y = RT$ is invalid, note that if ϕ has value 1 and ψ value $\frac{1}{2}$, then $\phi \rightarrow \psi$ has value $\frac{1}{2}$, and so $\phi \rightarrow \psi$ is permissively assertable and ϕ is restrictively assertable but ψ isn't restrictively assertable. To see that (MP1) / $X = PVT, Y = RVT$ is valid, take an arbitrary Sv-model $\mathcal{M}$ such that $\phi \rightarrow \psi$ is subtrue in $\mathcal{M}$ and ϕ is supertrue in $\mathcal{M}$. Then ψ is true in the classical model $\mathscr{M} \in \mathcal{M}$ in which $\phi \rightarrow \psi$ is true, and so ψ is subtrue in $\mathcal{M}$. To see that (MP2) / $X = RVT, Y = PVT$ is valid, take an arbitrary Sv-model $\mathcal{M}$ such that $\phi \rightarrow \psi$ is supertrue in $\mathcal{M}$ and ϕ is subtrue in $\mathcal{M}$. Then ψ is true in the classical model $\mathscr{M} \in \mathcal{M}$ in which ϕ is true, and so ψ is subtrue in $\mathcal{M}$. To see that (MP1) / $X = RVT, Y = PVT$ is invalid, consider the classical models $\mathscr{M}_1$ and $\mathscr{M}_2$ such that ϕ and ψ are true in $\mathscr{M}_1$ but false in $\mathscr{M}_2$. Then $\phi \rightarrow \psi$ is true in both $\mathscr{M}_1$ and $\mathscr{M}_2$, and so $\phi \rightarrow \psi$ is supertrue in the Sv-model $\{\mathscr{M}_1, \mathscr{M}_2\}$ and ϕ is subtrue in $\{\mathscr{M}_1, \mathscr{M}_2\}$, but ψ isn't supertrue $\{\mathscr{M}_1, \mathscr{M}_2\}$. To see that (MP2) / $X = PVT, Y = RVT$ is invalid, consider the classical models $\mathscr{M}_1$ and $\mathscr{M}_2$ such that ϕ is true in both $\mathscr{M}_1$ and $\mathscr{M}_2$ and ψ is true in $\mathscr{M}_1$ but false in $\mathscr{M}_2$. Then $\phi \rightarrow \psi$ is true in $\mathscr{M}_1$ but false in $\mathscr{M}_2$, and so $\phi \rightarrow \psi$ is subtrue in the Sv-model $\{\mathscr{M}_1, \mathscr{M}_2\}$ and ϕ is supertrue in $\{\mathscr{M}_1, \mathscr{M}_2\}$, but ψ isn't supertrue $\{\mathscr{M}_1, \mathscr{M}_2\}$.

Part II

Taming the Undecidable

Chapter 3

Counterpossibles in Science: The Case of Relative Computability

3.1 Introduction

It is a well known feature of the orthodox possible-worlds approach to counterfactual conditionals due to Robert Stalnaker (1968) and David Lewis (1973) that it makes all counterfactuals with metaphysically impossible antecedents come out vacuously true. Many have pointed out that this so-called *vacuity thesis* runs counter to our initial judgments about the truth-values of many such counterpossibles. Some of the proposed counterexamples to the vacuity thesis concern philosophical questions such as what would be the case if the laws of metaphysics had failed or if certain moral principles had been different, while others are about more ordinary topics such as whether anyone would've cared if Hobbes had squared the circle or what I would do if I were you.[1] Against such proposed counterexamples, Timothy Williamson (2007; 2010; 2015) has recently mounted a fresh defense of the vacuity thesis by making a strong case for its many theoretical virtues.[2]

In this chapter, I discuss a new source of trouble for the orthodoxy: relative computability theory. Textbook writers often introduce relative computability with the help of counterfactual conditionals. For example, Martin Davis writes that relative computability theory is concerned with the following:

---

[1]See Nolan (1997) and Brogaard and Salerno (2013) for influential papers and Berto (2013, §5.1) for more references.
[2]See Berto et al. (2017) for a detailed discussion of Williamson's arguments.

We may ask, of a given problem P,

If we could solve P, what else could we solve?

And, we may ask,

The solutions to which problems would also furnish solutions to P?

(Davis, 1958, 179, emphasis in the original)

After providing some background on relative computability theory, I will argue that, just like other mathematical facts, the facts uncovered by relative computability theorists are metaphysically necessary. So, on the assumption that P is not in fact solvable, the vacuity thesis would have it that any answer to Martin's first question is true.

But the vacuity thesis doesn't just find counterexamples in the way relative computability theorists talk about their discipline in ordinary language. I will argue that non-vacuous counterpossibles play a central role in how relative computability theory is developed in canonical textbooks on the subject. The vacuity thesis thus threatens to undermine how practitioners of an established science think about their discipline.

Instead of abandoning relative computability theory in light of this, I will instead draw from its resources to patch up the orthodoxy about counterfactuals. Like previous attempts to revise the theory of counterfactuals, I will present a model theory that makes use of so called "impossible worlds," world-like entities where metaphysical impossibilities can hold.[3] However, unlike earlier attempts, which have run into trouble when it comes to expanding Lewis' comparative similarity relation to these new entities,[4] I show that with the right choice of the set of "worlds," a comparative similarity relation immediately falls out of the mathematical theory of relative computability that gives the right results for counterfactuals about this theory. Questions remain about how to interpret my proposed model theory, especially the "worlds" involved. But given the continuity with the comparative similarity models for ordinary counterfactuals, these questions become tractable.

---

[3]See again Berto (2013) for an overview of previous proposals.
[4]See Baras (MS) for a discussion of Brogaard and Salerno's (2013) proposal.

3.2 Background

Computability theory studies what sets of natural numbers are algorithmically decidable (or "solvable," as in the above quote by Davis). By algorithmic decidability we mean that a computing agent could, in principle, decide for any natural number whether it is a member of the set by mechanically following a completely explicit algorithm that terminates in the right answer in finite time and after finitely many steps. An example of an algorithm is the truth table method, which allows us to decide for any sentence of the propositional calculus whether it is a tautology. The sets of natural numbers whose algorithmic decidability or lack thereof is of particular interest are those that represent certain well-formed problems. *The validity problem* (sometimes simply called *the decision problem*) is the set that encodes the sentences of the predicate calculus that are logically valid.[5] To say that the validity problem is algorithmically decidable would be to say that there is an algorithm that would allow us to decide for any number representing a sentence of the language of the predicate calculus whether it is a member of the set of the validity problem and so whether it is logically valid. It was a significant discovery by Alonzo Church (1936a; 1936b) and Alan Turing (1936) that the validity problem is not algorithmically decidable. Other sets that aren't algorithmically decidable are the *the halting problem*, which encodes the problem of deciding whether a computer will eventually halt when it's given a certain input, and *arithmetical truth*, which encodes the true sentences of the language of arithmetic.

As we already saw, *relative* computability theory is introduced by Martin Davis using counterfactuals. Similarly, Hartley Rogers says (where to *calculate* the characteristic function of a set amounts to algorithmically deciding the set):

> Intuitively, A is *reducible* to B if, given any method for calculating [the characteristic function of B], we could then obtain a method for calculating [the characteristic function of A.] (Rogers, Jr., 1967, 127, emphasis in the original)

And, for a more recent example, Herbert Enderton writes:

---

[5]For the sake of simplicity, I am straining traditional usage a bit here. Traditionally, the decision problem was so-called because it called for an algorithm for deciding membership in the set containing the logical validities; it wasn't the set itself that was called 'the decision problem.' See Mancosu and Zach (2015).

On the one hand, we might be able to show that if, hypothetically speaking, we could somehow decide membership in B, then we could decide membership in A. This would lead us to the opinion that A is no more undecidable than B is. (Enderton, 2011, 121)

The study of relative computability was spearheaded by Turing (1939) and Emil Post (1944) and later developed into a mature mathematical discipline using the usual extensional tools of set theory and first-order logic by the likes of Richard Friedberg, Stephen Kleene, Albert Muchnik, Rózsa Péter, and Post.[6] In formal regimentations of mathematics, the only conditional available is of course the material conditional. We know now that counterfactual conditionals behave very differently from material conditionals, but it wasn't until a few years after Turing's and Post's early papers appeared that counterfactual conditionals were identified as interesting objects of study.[7] And it took another twenty years after that until the now standard possible worlds model theory for counterfactuals was worked out by Robert Stalnaker and David Lewis.[8] But with hindsight, we can ask the kind of questions that we will be presently concerned with.

A basic result of relative computability theory is that the halting problem is *reducible* to the validity problem.[9] This fact can be expressed as follows:

(valid > halt) If the validity problem were algorithmically decidable, then the halting problem would also be algorithmically decidable.

By contrast, arithmetical truth is *not* reducible to the validity problem. This means that the following is false:

(valid > arith) If the validity problem were algorithmically decidable, then arithmetical truth would also be algorithmically decidable.

How do we know this? And how, for that matter, do we know that the validity problem, the halting problem, and arithmetical truth aren't algorithmically decidable? We know all of this due to

---

[6]See Soare (1996, 2009) for historical overviews of and Soare (2016) for an up-to-date introduction to the theory of relative computability. I will appeal to facts proven in Soare (2016) throughout here. Note that Soare (1996) initiated the change in usage from 'recusion theory' to '(relative) computability theory.'

[7]See Chisholm (1946) and Goodman (1947).

[8]See Stalnaker (1968), Stalnaker and Thomason (1970), and Lewis (1971, 1973). See also Todd (1964) and Sprigge (1970) for early statements of ideas similar to Stalnaker's and Lewis'.

[9]As it happens, the validity problem is also reducible to the halting problem; but the reducibility relation isn't in general symmetrical.

a combination of mathematical theorems and two principles connecting the mathematical apparatus with the notions of algorithmic decidability and reducibility. Take first the fact the validity problem, the halting problem, and arithmetical truth aren't arithmetically decidable. Church and Turing established certain mathematical theorems that get us halfway towards establishing this fact. It will be most illuminating to follow Turing's presentation of the result. Turing introduced a class of abstract machines that are now called *Turing machines*. He then showed that the assumption that, say, the halting problem is decidable by a Turing machine leads to a contradiction, akin to the contradiction Cantor derived from the assumption that the cardinality of the natural numbers is equal to the cardinality of the real numbers. Therefore, there isn't a Turing machine that decides the halting problem, or the validity problem or arithmetical truth for that matter. This is the mathematical part. The other part involves what is nowadays called *the Church-Turing thesis*. This thesis says that the sets that are algorithmically decidable in the informal sense, i.e. the sets that are decidable by any algorithmic means, are just the sets that are decidable by a Turing machine. Interpreted most conservatively, this thesis claims that Turing machines are an adequate model of algorithmic decidability.[10] Putting the Church-Turing thesis together with the fact that there's no Turing machine that decides the halting problem, the validity problem, or arithmetical truth gives us that these sets are algorithmically undecidable.

We know that the halting problem is reducible to the validity problem but arithmetical truth isn't for similar reasons, but with a twist. To establish these results, we need *oracle Turing machines*. An oracle Turing machine is just like a Turing machine, except that it has access to an "oracle." Oracles can be thought of as external storage devices that contain the correct answer to any "yes" or "no" question about a particular decision problem we may ask them. For example, an oracle for the validity problem contains, for arbitrary sentences of the predicate calculus, the answer to the question whether they are logically valid or not. Think of an oracle Turing machine as just like an ordinary Turing machine, except that it has an extra port where we can plug in an oracular storage device.[11] We can now show that an oracle Turing machine with an oracle for the validity problem can algo-

---

[10]See Shapiro (1981) and Rescorla (2007) for discussions of different interpretations of the thesis.

[11]Of course, talk of such storage devices is purely metaphorical; recall that Turing machines are abstract objects instead of concrete computing devices. So strictly speaking, an oracle is the abstract analogue of a concrete storage device.

rithmically transform the answers it gets about the validity problem into answers about the halting problem. That's how the halting problem is *Turing reducible* to the validity problem. However, even if the oracle Turing machine can ask the oracle questions about the validity problem, it won't be able to transform these answers into answers about arithmetical truth. That's how arithmetical truth isn't Turing reducible to the validity problem. To get from these results, which can be stated and proved purely mathematically, to the results that the halting problem is reducible simpliciter to the validity problem but that arithmetical truth isn't, we need an analogue of the Church-Turing thesis. This thesis, which is variously called *the Post-Turing thesis* or *the relativized Church-Turing thesis*, says that a set B is reducible simpliciter to a set A iff B is Turing reducible to A.[12]

But what is this relation of reducibility simpliciter? We may understand the claim that B is reducible to A as saying that if A were algorithmically decidable, then B would be algorithmically decidable—hence the counterfactual locutions in the above quotes from Davis, Rogers, and Enderton. In fact, I will argue that this is *the* way of understanding the claim. This understanding runs into philosophical trouble, however. For it is plausible that facts about what is and isn't algorithmically decidable are metaphysically necessary. The mathematical theorems involved in showing that none of our three sets can be decided by a Turing machine hold of course as a matter of metaphysical necessity. That it's metaphysically necessary that none of the sets are algorithmically decidable then follows by the fact that the Church-Turing thesis is metaphysically necessary.

What reasons do we have for thinking that the Church-Turing thesis is metaphysically necessary? Note that the limits of computation that Church and Turing discovered aren't merely technological. Church and Turing didn't merely show that we haven't built the right kind of computer or discovered the right kind of algorithm to decide the validity problem. In fact, Church and Turing's result *predates* the modern computer. Before anyone had built anything resembling a modern computer, Church and Turing had already identified computational problems that no computer could ever decide. And since the invention of the first computer, all technological innovations in computing, including innovations involving quantum computers that are yet to be realized,[13] have merely lead to an increase in computing speed and efficiency; they never have and never will lead to an

---

[12] See Soare (2009, 382) and Cooper (2004, 142).
[13] See Piccinini (2015, §4.3).

improvement in what can be algorithmically decided. Furthermore, the limits of computation that Church and Turing discovered also aren't merely limits imposed by the actual laws of nature. Church and Turing don't argue for their conclusion that the validity problem isn't algorithmically decidable by showing that the laws of nature rule out a computer that decides the validity problem.[14] This suggests that the degree to which it's impossible to algorithmically decide the validity problem is stronger than both technological or nomic impossibility. This suggests, but doesn't yet prove, that the Church-Turing thesis is indeed metaphysically necessary.[15]

There is also a direct argument for the metaphysical impossibility of the claim that, say, the validity problem is algorithmically decidable. To say that the validity problem isn't algorithmically decidable is just to say that there isn't an algorithm to decide the validity problem. But algorithms are abstract objects.[16] As such, they are the kinds of thing that either exist of metaphysical necessity or else don't exist at all; and if they don't exist, then it's metaphysically impossible that they exist. So if there isn't an algorithm to decide the validity problem, then it's metaphysically impossible that there exists such an algorithm, and so it's metaphysically impossible that the validity problem is algorithmically decidable. Thus, (valid > halt) and (valid > arith) are indeed counterpossibles.

The argument just presented relies on certain assumptions about metaphysical possibility and the modal metaphysics of abstracta, assumptions that may be doubted. Nevertheless, the assumptions are perfectly in line with orthodox thinking about these issues. So it follows from orthodox thinking about metaphysical possibility and the modal metaphysics of abstracta that it's metaphysically impossible that the validity problem is algorithmically decidable.[17]

It is important to be clear on what I am and am not claiming. I'm not claiming that it's metaphysically impossible to determine the members of the set corresponding to the validity problem. It is entirely compatible with everything I've said that some deity would be able to tell us for any natural number whether it is a member of that set. But if what I've argued for is right, then even

---

[14]I skip over some complications here; see Piccinini (2015, §4) for a more detailed discussion. In particular, I interpret the Church-Turing thesis as what Piccinini calls the *mathematical* Church-Turing thesis; I take this to be historically accurate. Note also that I discuss the issue of hypercomputation in section 3.4.

[15]McGee (2006, 111), one of the very few discussions of the modal status of algorithmic decidability, concurs.

[16]See Knuth (1966).

[17]These considerations also suggest that Cleland's (1993) thinking about the Church-Turing thesis is even more revisionary than Cleland herself suggests.

such a deity wouldn't be able to *algorithmically decide* the validity problem, because there is no algorithm that the deity could rely on. But since the antecedent of (valid > halt) and (valid > arith) claims that the validity problem is algorithmically decidable, the metaphysically possible existence of such a deity wouldn't pose a threat to my claim that these counterfactuals are indeed counterpossibles. Of course, given my concession that such a deity may be metaphysically possible, it may be worried immediately that the status of these counterfactuals as counterpossibles aren't significant, for perhaps we can reinterpret counterfactuals about relative computability as about such deities. However, things aren't that simple, as the extended argument in section 3.4 will show. I'll argue there that such a reinterpretation and many more like it would amount to a revision of what relative computability theorists take themselves to be doing.

But before we move on, let's state precisely what the present challenge to the Stalnaker-Lewis approach to counterfactuals is: we have counterfactuals about relative computability, such as (valid > halt) and (valid > arith) above, some of which appear to be true and some of which appear to be false, but we also have that these counterfactuals have metaphysically impossible antecedents. Now, the usual way of understanding the Stalnaker-Lewis approach to counterfactuals is as follows: a counterfactual ⌜If ϕ had been the case, then ψ would've been the case⌝ is true at a metaphysically possible world w iff all metaphysically possible worlds sufficiently similar to w where ϕ is true are such that ψ is true in them as well. Since there are no metaphysically possible worlds where the validity problem is algorithmically decidable, any counterfactual that starts with 'If the validity problem were algorithmically decidable...' is vacuously true. Given that with (valid > arith) we have such a counterfactual that appears to be false, we seem to have a counterexample to the semantics just sketched. And not just that: given that (valid > halt) appears to be true, we also immediately see that we can't just change the orthodoxy so that counterpossibles are all false.[18] And given that the halting problem and arithmetical truth are algorithmically decidable at all the same metaphysically possible worlds—namely none—, we also have a counterexample to the part of orthodoxy that says that taking a counterfactual sentence and replacing any of its subsentences with a sentence that's true at all the same metaphysically possible worlds yields a necessarily equivalent

---

[18]See Kment (2014, 25, 220) for a theory that treats all counterfactuals with *logically* impossible antecedents as vacuously false.

sentence.

I said that (valid > arith) *appears* to be false and that (valid > halt) *appears* to be true. In the next two sections, I argue that these appearances aren't deceiving: we ought to understand counterfactuals about relative computability literally; in fact, they play a central role in the definition of the reducibility relation.

3.3 Philosophical humility

Researchers in relative computability theory are authorities on the reducibility relation. However, they are generally not experts on the semantics of counterfactuals.[19] So the mere fact that they are disposed to assert some counterfactuals about relative computability and deny others doesn't indefeasibly undermine the orthodoxy about counterfactuals. On the face of it, this fact is simply another piece of evidence that needs to be weighed against the considerations that speak in favor of the orthodoxy, to be filed away with the well-known fact that ordinary speakers are disposed to assert some ordinary counterpossibles and deny others. Perhaps we can hold on to the orthodoxy and excuse relative computability theorists' dispositions by appealing to similar considerations with which we may excuse the dispositions of ordinary speakers. Timothy Williamson (2015, §4), for example, develops an error theory about the dispositions of ordinary speakers. So perhaps we can simply co-opt Williamson's error theory and conclude that counterfactuals about relative computability pose no threat to the orthodoxy, especially in light of the considerable theoretical pressures to hold on to the orthodoxy, also discussed by Williamson (2015, §2).

However, I am going to argue now and in the next section that this response on behalf of the orthodoxy runs counter to a certain kind of philosophical humility. This philosophical humility says that whenever an established mathematical or scientific discipline purports to study a certain phenomenon, we shouldn't give in to philosophical considerations that suggest that there is no such phenomenon to be studied.[20] Relative computability theory, which is certainly an established mathematical discipline, purports to study the reducibility relation. In the previous section, I mentioned

---

[19]Thanks to Alex Byrne for helpful discussion here.

[20]This attitude is related to Lewis' (1991, §2.8) Credo about set theory and Shapiro's (1997, ch. 1) "philosophy-last" approach to philosophy of mathematics.

that *a* way of understanding the claim that A is reducible to B is as saying that A would be algorithmically decidable if B were algorithmically decidable. I now want to argue that this is *the* way of understanding the reducibility relation. If that's right, and if the orthodoxy about counterfactuals is correct, then the reducibility relation holds between any two algorithmically undecidable sets. It would also mean that A isn't reducible to B iff B is algorithmically decidable and A isn't. But then the reducibility relation would carve out the same distinction among sets of natural numbers that the property of algorithmic decidability does. The study of the reducibility relation would thus become nothing other than the study of algorithmic decidability, and so relative computability theory is robbed of its own subject matter. In light of this fact, philosophical humility recommends that we reject the vacuity thesis.

Some might argue that philosophical humility should be understood slightly differently. The philosophy of mathematics that emerges from Stephen Yablo's *Aboutness* (2014, esp. §5.3) is a case in point. Astronomers study, among other things, the number of planets. However, nominalists think that numbers don't exist. So nominalism threatens to rob astronomy of one of its subjects. Yablo, who is a nominalist, agrees that astronomers speak falsely when they say that the number of planets in our solar system is eight. However, Yablo thinks that these astronomers nonetheless speak *correctly*, because what they say is partially, and non-vacuously, true—it has a true part, the part that is about the concrete world. Thus, with Yablo's theory of partial truth, we can hold on to a kind of philosophical humility, diminished though it may be, in allowing that that there is a phenomenon that astronomers study: how things stand concretely with the planets. Likewise, perhaps we can extract some non-vacuous core from the claim that A would be algorithmically decidable if B were algorithmically decidable, even when A and B are both algorithmically undecidable. This core would then be the proper phenomenon that relative computability theorists study.

Unfortunately though, Yablo's theory doesn't help in rescuing the orthodoxy. In order for this theory to yield the result that it's partially true that the number of planets is eight, Yablo needs there to be a possible world where the astronomers' statement is fully true, i.e. a world where numbers exist. Now, think about how we would develop a Yablovian theory of counterfactuals about relative computability. We would say that the statement (valid > halt) is vacuously true, but its assertion

86

is correct, perhaps because it has a non-vacuously true part that talks about certain structural relationships between the validity problem and the halting problem. But now if we wanted to follow Yablo's theory of partial truth, we would need there to be a possible world where the statement is fully true, and non-vacuously so. In such a world, we would need there to be a possible world where the validity problem is algorithmically decidable. So it looks like our Yablovian theory would require the claim that the validity problem is algorithmically decidable to be possibly possibly true. Now, it's true that unless we help ourselves to the characteristic axiom of the modal logic $S4$, 'possibly, possibly, the validity problem is algorithmically decidable' isn't quite the same as 'possibly, the validity problem is algorithmically decidable.' But it also isn't so far removed from it that we can be said to have made genuine progress on behalf of the orthodox approach to counterfactuals. What's more, both Stalnaker and Lewis as well as the the model theory I will present later validate the $S4$ axiom.

In sum, philosophical humility does indeed recommend that we reject the orthodoxy about counterfactuals, on the assumption that the counterfactual way of understanding the reducibility relation is indeed *the* way of understanding it. I now turn to a defense of this latter claim.

3.4   Understanding and misunderstanding reducibility

An immediate reason for thinking that the counterfactual way of understanding the reducibility relation is indeed *the* way of understanding it is that that's exactly how Davis, Rogers, and Enderton characterize the relation in our quotes above (see pages 78–79). But of course, this needn't be decisive. Perhaps we want to say that when relative computability theorists assert (valid > halt), what they're *really* saying is _____. Let's look at seven proposals of how to fill in this blank. The first five are instances of quite general proposals of how to respond to purported counterexamples to the vacuity thesis whereas the final two are specific proposals about our counterfactuals about relative computability. I will argue that none of these proposals work. This suggests that when relative computability theorists assert (valid > halt), they really mean it, which in turn suggests that the reducibility relation is indeed to be understood in terms of counterfactuals.

Idioms. Here's a proposal for filling in the blank above: when relative computability theorists assert (valid $>$ halt), what they're *really* saying is that the halting problem is reducible to the validity problem; the counterfactual locution (valid $>$ halt) and its variants in the quotes from Davis, Rogers, and Enderton are merely idiomatic ways of gesturing towards the notion of reducibility. Perhaps the counterfactual locution is particularly evocative of some of the ideas behind the notion of reducibility, but sentences such as (valid $>$ arith) aren't to be taken literally.

However, counterfactuals about relative computability don't behave linguistically the way idioms do. In general, idioms, though syntactically complex, are not semantically complex. Take the idiom 'to keep an eye out for.' While the sentence 'I'm keeping an eye out for you' is perfectly linguistically appropriate, its cleft analogue 'It's an eye that I'm keeping out for you' strikes us as odd. This despite the fact that with non-idiomatic expressions, a cleft sentence is very close in meaning to its non-cleft variant; viz. 'I gave her an umbrella' and 'It's an umbrella that I gave her.' The reason for this is that the meaning of 'to keep an eye out for,' unlike the meaning of 'to give an umbrella to,' is not derived compositionally from the meanings of its parts. Rather, its meaning is directly lexically encoded by the whole expression. This means that, on the level of semantics, 'to keep an eye out for' is a single unit that can't be broken up by, say, cleft constructions.[21] In contrast, counterfactuals about relative computability interact with other sentence constructions just like ordinary counterfactuals do. For example, not only is (valid $>$ arith) false, but the following where we add a negation is true:

(valid $> \overline{\text{arith}}$) (Even) if the validity problem were algorithmically decidable, arithmetical truth would (still) not be algorithmically decidable.[22]

We will see more examples of how these counterfactuals interact with quantifiers and conjunction shortly. From this, it emerges that the compositional behavior of counterfactuals about relative com-

---

[21]Although this is the received view of the semantics of idioms, it has been challenged: see Nunberg et al. (1994) and Egan (2008). If the received view is in fact false, then the present propoosal collapses into the one involving the idea of glosses discussed next.

[22]Note that (valid $> \overline{\text{arith}}$) is only *equivalent* to the negation of (valid $>$ arith) if we assume conditional excluded middle, which, as we'll see later, is not in general valid. Note also that I assume that 'even' and 'still' don't make any truth-conditional contributions to the counterfactuals in which they occur, which is why I have put them in parentheses; see Bennett (2003, §§102–7) for a defense of this assumption.

putability is just like that of ordinary counterfactuals, so that they can't be merely idiomatic ways of speaking.

Glosses. A related option would be to treat the counterfactuals used by relative computability theorists as imperfect glosses or paraphrases of claims about reducibility. It may be thought that the case is analogous to the case of causation.[23] When asked to explain what we mean by 'causation,' we make free use of counterfactual locutions. But, so the proposal goes, the failure of the program of analyzing causation in terms of counterfactuals should teach us that we shouldn't take counterfactual locutions as they appear in writings on relative computability too seriously. So, on this proposal, counterfactuals don't characterize or define the reducibility relation, they merely illuminate it.

There are two problems with this analogy with causation. First, the problem with counterfactual analyses of causation is that they notoriously either over- or undergenerate cases of genuine causation. Things are different in the case of relative computability. If we bracket the violations of the vacuity thesis—which it is fair to bracket, since the status of the vacuity thesis is the very thing that's at issue—, counterfactual glosses on the notion of reducibility seem to get things *exactly* right.

Secondly, we seem to have an understanding of causation that's independent of our understanding of counterfactuals. In fact, several authors have recently argued that we should give a semantics for counterfactuals in terms of causal models, the latter of which treat causation as a primitive notion.[24] In contrast, it's implausible that the notion of reducibility that's at the core of relative computability theory is primitive. We simply don't have a pretheoretical notion of reducibility that's not understood by way of some auxiliary notions. My present claim is that reducibility is understood in terms of counterfactuals, and that some of these counterfactuals are counterpossibles. It would be entirely mysterious how such an understanding could be achieved if the vacuity thesis were correct. Of course, whether my claim about how we understand reducibility is true will in part depend on whether there are ways of understanding the reducibility relation that don't involve counterpossibles. I will discuss some potential definitions presently.

---

[23]Thanks to Bradford Skow for suggesting this analogy and to Justin Khoo and an anonymous referee for urging me to give this proposal more serious consideration.
[24]See Briggs (2012) and the references therein.

Conceptual possibility. One tempting response to counterpossibles that appear to be non-trivial is to interpret them as talking about what's *conceptually* possible. The notion of conceptual possibility is a famously fraught one, since it is tied to the notions of apriority and analyticity. ϕ is sometimes said to be epistemically possible iff $\ulcorner\neg\phi\urcorner$ isn't knowable a priori, and sometimes it's said that ϕ is epistemically possible iff $\ulcorner\neg\phi\urcorner$ isn't true in virtue of meaning. We needn't be concerned with the details here. Let's just grant that there is a notion of conceptual possibility according to which it's conceptually possible that water is an element and that Ms. Marvel, the heroine of the eponymous comic book series, isn't Kamala Khan. A conceptually possible world can then be defined as a maximal consistent set of sentences that includes all a priori knowable or analytic truths.[25] Conceptually possible worlds may be used to give a model theory for counterfactuals such as the following:

(water) If water had been an element, then water splitting would've been impossible.

(marvel) If Ms. Marvel hadn't been Kamala Khan, we would've seen them together at some point or another.

Since the sentences 'Water isn't an element' and 'Ms. Marvel is Kamala Khan,' though true, are neither a priori knowable nor analytic, there will be conceptually possible worlds where the antecedents of (water) and (marvel) are true. Note that it's crucial for this general strategy to be promising that the building blocks out of which we construct the worlds are sentences, or perhaps Fregean senses, and not something more worldly such as Russellian propositions. The Russellian proposition corresponding to 'Ms. Marvel is Kamala Khan' is the same as the Russellian proposition corresponding to the logical truth 'Ms. Marvel is Ms. Marvel,' and so there isn't any consistent set of Russellian propositions that contains the Russellian proposition corresponding to 'Ms. Marvel is Kamala Khan.'[26]

Given the promise of conceptually possible worlds constructed out of sentences in giving a model theory for (water) and (marvel), it's tempting to also use them to give a model theory for counterfactuals about relative computability. After all, it's plausible that it's conceptually possible

---

[25]I assume here that failures of the laws of logic aren't conceptually possible. This is a harmless assumption in the present context since we're not concerned with counterfactuals with explicit violations of the laws of logic in the antecedent. See Brogaard and Salerno (2013) for an account along the lines I'm imagining here that dispenses with this assumption.

[26]I am grateful to an anonymous referee for pressing me to be clearer on this.

that the validity problem is algorithmically decidable.[27] However, conceptual possibility notoriously run into difficulties when it comes to quantifying-in.[28] Indeed, it is commonly assumed that it is illegitimate to quantify into sentential contexts that involve conceptual possibility. But now note that there are certain results about relative computability that require quantifying-in when we express them using counterfactuals. Take Gerald Sacks' (1964) Density Theorem. It states that the Turing reducibility relation is dense.[29] Where $A \leq_T B$ says that A is Turing reducible to B and $A <_T B$ says that $A \leq_T B$ and $B \nleq_T A$, this theorem can be expressed as follows: for any two sets A, B, if $A <_T B$, there is a set C such that $A <_T C <_T B$. Using the Post-Turing thesis, we can express this theorem as follows:

(sacks) For any A, B, if it's the case that A would be computable if B were computable but not vice versa, then there's some C such that: A would be computable if C were computable but not vice versa and C would be computable if B were computable but not vice versa.

So we see that quantifying-in is very natural for counterfactuals about relative computability. This sets these counterfactuals apart from the kinds of examples commonly discussed in the literature on counterpossibles.[30]

Note that the present claim isn't that the fact that sentences such as (sacks) involve quantification into counterfactuals prohibits the use of *any* world-like entities in their analysis. In fact, the model theory I will present later also involves world-like entities. The present claim is just that the presence, and indeed indispensability, of quantifying-in in some counterfactuals about relative computability calls for more elaborate resources than just conceptually possible worlds qua maximal consistent sets of sentences.

---

[27]This is assuming, perhaps contrary to Smith (2007, §35), Sieg (2008), and Kripke (2012), that the Church-Turing thesis isn't a conceptual truth. If you disagree, then so much the worse for the present proposal on behalf of the orthodoxy.

[28]These difficulties are most famously noted by Quine (1953). Yalcin (2015) develops a Fregean compositional semantics for quantifying into belief contexts. It's not obvious that his ideas can be adapted to the present case.

[29]Strictly speaking, Sacks shows that the relation on the Turing *degrees* is dense. That the Turing reducibility relation is dense is an immediate corollary. For ease of exposition, I'll put off discussion of Turing degrees until the next section.

[30]For example, Brogaard and Salerno (2013) don't even tell us how to extend their model theory to a language with quantifiers.

Semantic ascent.   Perhaps counterfactuals about relative computability are best understood as making meta-linguistic remarks about the predicate 'algorithmically decidable': (valid > halt) says that if the extension of 'algorithmically decidable' had included the validity problem, then it would also have included the halting problem. In discussing a proposal like this, Berit Brogaard and Joe Salerno (2013) assert that they "highly doubt that there is an elegant and convincing pragmatic story to be told" about why we would ascend semantically in such a way (p. 645). Contrary to this, I submit that we can tell at least a partial story using Stalnaker's (1978) apparatus of diagonalization. Without going into too many details, this apparatus could be extended quite straightforwardly to predict that counterfactuals with impossible antecedents receive a non-standard reading on which they make meta-linguistic remarks such as the above.

Nevertheless, this story would remain incomplete. Suppose for simplicity that we give a simple Stalnakerian semantics for the reinterpreted counterfactual: the closest world where the extension of 'algorithmically decidable' includes the validity problem is such that at that world, the extension also includes the halting problem. We may ask why this would be so. Surely, the extension of 'algorithmically decidable' could have differed in all sorts of ways. For example, the minimal way of changing the extension so as to include the validity problem would be to *just* add the validity problem and nothing else. Surely, it isn't a brute fact about the predicate 'algorithmically decidable' that this minimal change isn't what happens at the closest world. The reason as to why this minimal change is ruled out must lie in the fact that the halting problem is reducible to the validity problem. But now we're taking the notion of reducibility as more basic than the counterfactuals in terms of which we had originally defined that notion. So now it looks like the best we can do to explain why the closest world where the extension of 'algorithmically decidable' includes the validity problem is such that the extension also includes the halting problem is by appealing to the truth of (valid > halt). This suggests that we have a better grip on the literal interpretation of (valid > halt) than on its meta-linguistic reinterpretation.

The *reductio* analogy.   Maybe we can understand counterfactuals about relative computability along the lines of counterfactuals found in informal *reductio* proofs.[31] Consider Euclid's proof that

---

[31]Thanks to Stephen Yablo for pushing me to think harder about this strategy.

there are infinitely many primes. We start by supposing that there are exactly n many primes. Let $p_1, \ldots, p_n$ be them. It follows that there will be a prime p that divides $p_1 \times \cdots \times p_n + 1$. The crucial next step in the proof can then be put in counterfactual terms:

(euclid) If p were one of $p_1, \ldots, p_n$, then p would divide $(p_1 \times \cdots \times p_n + 1) - p_1 \times \cdots \times p_n$.

Since nothing divides $(p_1 \times \cdots \times p_n + 1) - p_1 \times \cdots \times p_n = 1$, we conclude by modus tollens that p isn't one of $p_1, \ldots, p_n$, and so that $p_1, \ldots, p_n$ aren't all of the primes after all. Now, there is some debate over whether counterfactuals such as (euclid) pose a serious challenge to the standard approach to counterfactuals.[32] Suppose they don't. And suppose that counterfactuals such as (euclid) are best understood either as material conditionals or as strict conditionals. This may be particularly plausible in cases where the material conditional is a logical truth, for in that case a normal modal logic proves the corresponding strict conditional, and both Stalnaker's and Lewis' counterfactual logics then prove the corresponding counterfactual. In any case, whatever the details of the story may be that we tell about (euclid), the present proposal on behalf of the orthodoxy suggests that we treat counterfactuals about relative computability along the same lines. (valid $>$ halt) and (valid $>$ arith), the proposal goes, are merely disguised material or strict conditionals.

The problem with this proposal is that counterfactuals about relative computability don't behave like material or strict conditionals. The reason why Stalnaker and Lewis developed their model theory for counterfactuals is that natural language counterfactuals fail to conform to antecedent strengthening, which is valid for material and strict conditionals. Focusing on the case of strict conditional, this principle reads:

$$\frac{\Box(\phi \to \psi)}{\Box((\phi \land \chi) \to \psi)}$$

This rule seems adequate for (euclid). No matter what else we put in its antecedent to strengthen it, the resulting sentence still seems true, though perhaps misleading.[33] However, there are counterexamples to antecedent strengthening in the case of counterfactuals about relative computability. Consider:

---

[32]Nolan (1997, 537–8) doesn't think so whereas Dutilh Novaes (2016) does. See also Williamson (2015, §3) for discussion.

[33]If you disagree, then so much the worse for the present proposal on behalf of the orthodoxy.

(valid > $\overline{\text{arith}}$) (Even) if the validity problem were algorithmically decidable, arithmetical truth would (still) not be algorithmically decidable.

(valid&arith > $\overline{\text{arith}}$) (Even) if the validity problem and arithmetical truth were algorithmically decidable, arithmetical truth would (still) not be algorithmically decidable.

On the strict conditional interpretation, the inference from (valid > $\overline{\text{arith}}$) to (valid&arith > $\overline{\text{arith}}$) is an instance of antecedent strengthening. But (valid > $\overline{\text{arith}}$) is true and (valid&arith > $\overline{\text{arith}}$) is false.

In response, it may be suggested that the negation in (valid > $\overline{\text{arith}}$) and (valid&arith > $\overline{\text{arith}}$) is a wide-scope negation so that (valid > $\overline{\text{arith}}$) and (valid&arith > $\overline{\text{arith}}$) become '$\neg\Box(V \to A)$' and '$\neg\Box\big((V \land A) \to A\big)$' respectively. Perhaps some story can be told according to which the added 'even' and 'still,' which make (valid > $\overline{\text{arith}}$) and (valid&arith > $\overline{\text{arith}}$) sound more natural, force such a wide-scope interpretation.[34] On this regimentation, the inference from (valid > $\overline{\text{arith}}$) to (valid&arith > $\overline{\text{arith}}$) isn't an instance of antecedent strengthening anymore.

However, this response won't work in full generality. For consider:

(valid > halt&$\overline{\text{arith}}$) If the validity problem were algorithmically decidable, then the halting problem would be algorithmically decidable but arithmetical truth would (still) not be algorithmically decidable.

(valid&arith > halt&$\overline{\text{arith}}$) If the validity problem and arithmetical truth were algorithmically decidable, then the halting problem would be algorithmically decidable but arithmetical truth would (still) not be algorithmically decidable.

As before, (valid > halt&$\overline{\text{arith}}$) is true but (valid&arith > halt&$\overline{\text{arith}}$) is false. But here, there is no temptation whatsoever to treat the negation embedded within the consequent as taking wide scope over the whole counterfactual.

What's more, there are even counterexamples to the claim that a negation that appears unembedded in the consequent of a counterfactual should always be read as taking wide scope. It follows

---

[34]Though see footnote 22.

from Corollary 1 in §2.2 of Kleene and Post (1954) that there are sets of natural numbers A and B neither of which is reducible simpliciter to the other. This means that we should be inclined to reject the following:

(A ∨ B > $\overline{\text{B}}$) If A or B were algorithmically decidable, then B wouldn't be algorithmically decidable.

But now if it were mandatory to read the negation in (A ∨ B > $\overline{\text{B}}$) as taking wide scope, we should expect to accept the following:

(A ∨ B > B) If A or B were algorithmically decidable, then B would be algorithmically decidable.

In fact, however, we should reject (A ∨ B > B) for the same reason that leads us to reject (A ∨ B > $\overline{\text{B}}$).[35]

   In short, the claim that counterfactuals about relative computability are material or strict conditionals is untenable.[36]

The primacy of oracles.   Here's a proposal on behalf of the orthodoxy that exploits the particulars of what these counterfactuals are about. The proposal is that, for example, (valid > halt) is

---

[35] It might be worried that we are only inclined to reject (A ∨ B > B) because simplification of disjunctive antecedents is a valid rule of inference for counterfactuals. This rule reads:

$$\frac{(\phi \lor \psi) \mathbin{\square\!\!\rightarrow} \chi}{(\phi \mathbin{\square\!\!\rightarrow} \chi) \land (\psi \mathbin{\square\!\!\rightarrow} \chi)}$$

Simplification isn't valid in Stalnaker's and Lewis' logics of counterfactuals, but Fine (1975a, 2012), Ellis et al. (1977), and Santorio (2017) have argued that that's a defect of these logics. But even accepting simplification doesn't help in the current situation. For, Kleene and Post's result also leads us to reject the following.

(A > B) If A were algorithmically decidable, then B would be algorithmically decidable.

But then by simplififaction, we should also reject (AB > B). So we should accept the negation of (A ∨ B > B). But then on the assumption that a negation in the consequent of a counterfactual takes wide scope, we should accept (A ∨ B > $\overline{\text{B}}$) as well, contrary to what we just observed.

[36] von Fintel (2001) and Gillies (2007) have recently argued that natural language counterfactuals only *dynamically* fail to validate antecedent strengthening. Whether they are right is subject to ongoing debate; see Moss (2012) and Lewis (2017) for criticism. But even if von Fintel and Gillies turn out to be right, their dynamic semantics is still very different from the static strict conditional treatment that the current proposal argues is adequate for counterfactuals in *reductio* proofs. For example, von Fintel and Gillies need something like a comparative similarity relation to model the evolution of the context, whereas a static strict conditional treatment only needs an accessibility relation for the modal operators. This also means, in turn, that if we wish to model counterfactuals about relative computability in a dynamic framework, we could just borrow the comparative similarity relation of the model theory that I describe in the appendix.

merely shorthand for saying that if we had an oracle for the validity problem, then we could figure out the right answer to any question we may ask about the halting problem.[37] This proposal is inspired by the way we study relative computability, namely by way of oracle Turing machines. What may further motivate this proposal is the thought that there isn't a clear phenomenon, relative computability, that we have a grasp of independently of studying it with oracle Turing machines. Perhaps all we have in relative computability theory is a mathematically rich and thus mathematically interesting structure that doesn't correspond to anything non-mathematical. Don't we all know that mathematicians can become interested in just about any arcane phenomenon as long as it gives rise to a mathematically interesting structure? What's more, understanding counterfactuals such as the above as merely shorthand for saying that if we had an oracle for the validity problem would make its antecedent metaphysically possible. For certainly, the proposal continues, though perhaps nomically impossible, oracles by themselves surely aren't metaphysically impossible. Perhaps there could have popped up out of nowhere an oracle that intuits facts about the validity problem. In fact, look back at the quotes from Davis and Enderton (see pages 78 and 79). Davis' counterfactual begins with 'If we could solve $P\dots$' and Enderton's begins with 'If, hypothetically speaking, we could somehow decide membership in $B\dots$' Regarding the quote from Davis, I said that to solve a problem just is to algorithmically decide it. Perhaps I was too quick here. Perhaps Davis has in mind a more general notion of solving, and Enderton has in mind a more general notion of deciding, one that allows reference to metaphysically possible oracles that pop up out of nowhere. A more sober rendition of the present proposal is the following:

(info) When relative computability theorists assert (valid > halt), what they're *really* saying is that there is an algorithm that would allow us to decide which natural numbers are members of the halting problem if we were given complete information about which natural numbers are members of the validity problem.[38]

Just like oracles are metaphysically possible, it's metaphysically possible to be given complete information about which natural numbers are members of the validity problem.

---

[37]Agustín Rayo suggested this to me in personal communication.

[38]Thanks to an anonymous referee for suggesting this formulation.

The claim that there isn't any phenomenon to be studied that we understand independently of the notion of an oracle Turing machine runs directly counter to how Rogers develops the subject in his book. In chapter 8, Rogers describes a relation of "reducibility" (the scare quotes are Rogers') among sets that is similar to Turing reducibility, called truth-table reducibility, but which is not defined in terms of oracle Turing machines. After describing truth-table reducibility, Rogers argues for the need for the stronger relation of Turing reducibility in chapter 9, which of course is defined in terms of oracle Turing machines. His argument goes as follows. He produces two sets, the first of which he argues is reducible to the second. He then shows that the first set isn't *truth-table* reducible to the second, but that it is *Turing* reducible to it. Rogers concludes that using truth-table reducibility to analyze what he explicitly calls the *intuitive* notion of reducibility would be inadequate, for this would leave out certain sets, and that an analysis in terms of Turing reducibility fares better. To arrive at this verdict, Rogers clearly assumes that he and his readers have an understanding of the notion of reducibility that's independent of talk about oracle Turing machines. And the understanding of reducibility that Rogers provides is in terms of counterfactuals. In fact, looking back at his quote reveals that it's more difficult to read Rogers in such a way that he's talking about something metaphysically possible. For Rogers' (syntactically non-standard) counterfactual begins with 'given any method for calculating [the characteristic function of B]....' And simply being given information doesn't involve any calculating; after all, calculating the validity problem is metaphysically impossible.

Note that the present claim isn't that Rogers assumes that his use of counterpossibles allows himself and his readers to gain an explicit knowledge of the full extension of the relation of reducibility and that he then holds up this extension against the extension of Turing reducibility. Rather, the claim is that Rogers assumes that his use of counterfactuals allows himself and his readers to have an implicit grasp of the notion of reducibility. It may well be, and in fact it is quite plausible, that to pin down the exact boundaries of the extension of the relation of reducibility, the notion of Turing reducibility, which allows for a precise mathematical analysis, is indispensable. But to admit this is consistent with claiming that counterpossibles are essential in pinning down the subject matter of

relative computability theory.[39]

Regarding the analysis of reducibility in terms of (info), I don't deny that this analysis succeeds, just like I don't deny that the analysis of reducibility simpliciter in terms of Turing reducibility succeeds. However, (info) crucially appeals to the notion of a *relative* algorithm, i.e. an algorithm that is given complete information about a certain set of natural numbers. While working relative computability theorists of course have a grasp of this notion, the fact that Rogers sees the need to introduce the notion of reducibility in terms of counterpossibles that don't appeal to the notion of a relative algorithm suggests that the conceptual building blocks that are required for an understanding of relative computability theory are the notion of a non-relative algorithm on the one hand and counterfactuals on the other. But these building blocks only succeed in facilitating an understanding of relative computability theory if the vacuity thesis is false.

That Rogers assumes that he and his readers come to have an understanding of reducibility by way of his use of counterpossibles may be dismissed if Rogers were a minor figure in relative computability theory and if his readers were few. However, from its initial release in 1967 until at least the release of Robert Soare's (1986) textbook, Rogers' book was *the* main textbook with the help of which a whole generation of mathematicians was raised.

The primacy of hypercomputers.   Another topic specific proposal suggests that the study of relative computability is the study of metaphysically possible *hypercomputers*. Hypercomputers are hypothesized machines that overcome the finiteness of actual computers in one way or another. Many such machines have been described in the literature.[40] One is a so called *accelerating Turing machine*, also called *Zeus machine*.[41] This is a machine that completes an infinite number of computational steps in a finite amount of time. One way it could do this is by completing a supertask, e.g. by completing the first computational step in one minute, the second step in half a minute, the third step in fifteen seconds, and so on. In other words, the machine completes each computational step after the first one in half the time it took to complete the previous one. After two minutes

---

[39]Thanks to an anonymous referee for urging me to be clearer on this.
[40]See Davis (2004) and Piccinini (2015, §4.3) for critical discussions and references.
[41]See Boolos et al. (2007, 19).

have passed, the machine will have completed an infinite number of steps. There's some debate about whether accelerating Turing machines and the supertasks that they require are physically possible.[42] But they surely seem to be *metaphysically* possible.[43] Now, accelerating Turing machines could "decide" the validity problem. That's because that set, though algorithmically undecidable, is *computably enumerable*. This means that the set of predicate logic validities is such that a Turing machine, given an infinite amount of time, could list all of its members. Consequently, an accelerating Turing machine of the sort described above could list all and only the members of that set in two minutes. In order to decide in a finite amount of time whether a sentence of the predicate calculus is logically valid, this machine would then just have generate the list and determine whether the sentence appears on it or not.

So, perhaps talk about relative computability could be cashed out in terms of talk about hyper-computers: a set A is reducible to a set B iff if the laws of nature allowed for a hypercomputer that could decide membership in B, then the laws would also allow for a hypercomputer that could decide membership in A. The only modalities involved here are metaphysical.

However, this proposal makes false predictions. It predicts that there will be true counterfactuals of the form,

($A \leq_{hyper} B$) If the laws of nature would allow for a hypercomputer that could decide membership in B, then the laws would also allow for a hypercomputer that could decide membership in A,

even though the corresponding claim about Turing reducibility,

($A \leq_T B$) A is Turing reducible to B,

is false. To see this, note that there are some algorithmically undecidable but computably enumerable sets A and B that are such that it's neither the case that A is reducible to B nor vice versa.[44] Since A is computably enumerable, a Zeus machine could "decide" A. But presumably if the laws

[42]See Earman (1995, ch. 4), Davis (2004, 197), and Romero (2014) for discussion.
[43]See Shagrir (2004) for an argument that accelerating Turing machines don't fall prey to Thomson's (1954) paradox.
[44]We know that such sets exists due to Friedberg (1957) and Muchnik's (1956) solution to Post's (1944) Problem.

of nature allowed for there to be a Zeus machine that could "decide" A, then they would also allow for there to be a Zeus machine that could "decide" B, since B is computably enumerable as well. In other words, if we had a hypercomputer to decide membership in A, then we could also have a hypercomputer to decide membership in B. Consequently, ($A \leq_{hyper} B$) is true, even though ($A \leq_T B$) is false. So the explanation of claims involving the Turing reducibility relation, and in turn of relative computability, in terms of what a Zeus machine could do yields the wrong results. And in fact, according to the theory of supertask computation as developed by Joel David Hamkins (2004) and Philip Welch (2004), Zeus machines are vastly more powerful than many oracle machines.[45]

The failure of these seven proposals suggests that the reducibility relation is indeed to be understood in terms of counterfactuals. This means that the orthodoxy about counterfactuals does indeed rob relative computability theory of its subject. Philosophical humility thus recommends that we reject the orthodoxy. But perhaps we think that philosophical humility has its limits. Perhaps we want to dig in our heels and insist that counterfactuals such as (valid > halt) and (valid > arith) are indeed both true. This attitude owes us a story as to why these counterfactuals strike us as *prima facie* non-vacuous. We can take a cue from Williamson's (2015) discussion here.[46]

The following is a version of an argument of Williamson's that purports to put pressure on our inclination to treat (valid > arith) as false using general principles of the logic of counterfactuals. The argument, which is adapted for our purposes, starts by claiming that (valid > arith) is equivalent to (valid&$\overline{\text{valid}}$ > arith):

(valid&$\overline{\text{valid}}$ > arith) If the validity problem were and weren't algorithmically decidable, then arithmetical truth would be algorithmically decidable.

Why should this equivalence hold? It's metaphysically necessary that the validity problem isn't algo-

---

[45]It may be argued, perhaps with Shapiro (2006), that the informal notion of decidability, and in turn the informal notion of relative computability, can be precisified in a number of ways, one of which coincides with the notion of hypercomputers. But this would still leave us with the result that there's *a* notion of relative computability on which counterfactuals about relative computability are non-vacuous yet have metaphysically impossible antecedents. Thanks to Kieran Setiya for discussion here.

[46]See also Williamson (2010, 95–6) for an earlier discussion.

rithmically decidable. And since ϕ is metaphysically equivalent to $\ulcorner \phi \wedge \psi \urcorner$ whenever ψ is metaphysically necessary, 'the validity problem is algorithmically decidable' is metaphysically equivalent to 'the validity problem is and isn't algorithmically decidable.' In worlds talk, 'the validity problem is algorithmically decidable' is true at all the same metaphysically possible worlds as 'the validity problem is and isn't algorithmically decidable.' Next, it is claimed that counterfactuals allow for substitution of necessary equivalents; i.e. if ϕ and ψ are true at all the same metaphysically possible worlds, then $\ulcorner \phi \mathbin{\square\!\!\rightarrow} \chi \urcorner$ and $\ulcorner \psi \mathbin{\square\!\!\rightarrow} \chi \urcorner$ are equivalent. This gives us the desired equivalence between (valid $>$ arith) and (valid&$\overline{\text{valid}}$ $>$ arith). Now, surely (valid&$\overline{\text{valid}}$ $>$ arith), with its logically impossible antecedent, is much less obviously false than (valid $>$ arith). So perhaps we are merely tricked into thinking that (valid $>$ arith) is false because we don't realize that it's equivalent to (valid&$\overline{\text{valid}}$ $>$ arith).

However, a closer look at this argument reveals that it rests on an assumption that we ought to reject for the same reason that we ought to accept counterfactuals about relative computability as non-vacuous. Let's look at how we would derive the supposed equivalence between (valid $>$ arith) and (valid&$\overline{\text{valid}}$ $>$ arith). Stalnaker's (1968, 106) counterfactual logic $C2$ contains the following axioms:

(a3) $\square(\phi \rightarrow \psi) \rightarrow (\phi \mathbin{\square\!\!\rightarrow} \psi)$

(a7) $\big((\phi \mathbin{\square\!\!\rightarrow} \psi) \wedge (\psi \mathbin{\square\!\!\rightarrow} \phi)\big) \rightarrow \big((\phi \mathbin{\square\!\!\rightarrow} \chi) \leftrightarrow (\psi \mathbin{\square\!\!\rightarrow} \chi)\big)$

Now, since it's metaphysically necessary that the validity problem isn't algorithmically decidable, we have:

$$\square\big(V \leftrightarrow (V \wedge \neg V)\big)$$

Using (a3), this gives us:

$$\big(V \mathbin{\square\!\!\rightarrow} (V \wedge \neg V)\big)$$

and

$$\big((V \wedge \neg V) \mathbin{\square\!\!\rightarrow} V\big)$$

And so (a7) gives us:

$$\big(V \mathbin{\square\!\!\rightarrow} A\big) \leftrightarrow \big((V \wedge \neg V) \mathbin{\square\!\!\rightarrow} A\big)$$

That (a3) gives us '$V \boxright (V \wedge \neg V)$' is suspicious. For it follows from this that any counterfactual that assumes in its antecedent that the validity problem is algorithmically decidable is vacuous. And whether that's the case is exactly what's at issue. So if '$\Box$' is interpreted as metaphysical necessity, then we ought to reject (a3). In counterfactual logic, '$\Box$' is usually defined such that $\ulcorner \Box \phi \urcorner$ abbreviates $\ulcorner \neg \phi \boxright \phi \urcorner$. That's how Lewis (1973, §1.5) defines it; he calls '$\Box$' *outer necessity*. (a3) is valid in the model theory I present in the appendix if that's how we understand '$\Box$,' since outer necessity is now broader than metaphysical necessity. But if that's how we understand '$\Box$,' then we can't accept '$\Box(V \leftrightarrow (V \wedge \neg V))$.' The latter is true only where '$\Box$' is understood as metaphysical necessity. So the logic of counterfactuals doesn't force upon us the equivalence of (valid > arith) and (valid&$\overline{\text{valid}}$ > arith). And without this equivalence, it becomes less plausible that we are tricked into thinking that (valid > arith) is false. Note that given this notion of outer necessity, the debate over the vacuity thesis can be rephrased as follows: is outer necessity the same as metaphysical necessity? Stalnaker, Lewis, and Williamson think that it is, whereas I argue that outer necessity is stronger than metaphysical necessity.

A final way of holding on to the orthodoxy is to argue that despite its shortcomings, it's the only game in town, since all alternative approaches such as for example that of Brogaard and Salerno (2013) run into serious trouble. And indeed, perhaps there's a way of amending the orthodoxy by providing an error theory about our judgments about counterpossibles. Williamson (2015, §4), for example, proposes that we use certain heuristics when evaluating counterfactuals that lead us astray in cases of counterpossibles. However, in the next section, I describe a model theory for counterfactuals about relative computability, which I describe in more detail in the appendix, which I hope demonstrates that the orthodoxy isn't the only game in town.

3.5 Patching up the orthodoxy

Williamson likens the supposed folly of rejecting the vacuity thesis to the Aristotelian rejection of vacuously true universal generalizations:

> The logic of quantifiers was confused and retarded for centuries by unwillingness to
> recognize vacuously true universal generalizations; we should not allow the logic of

counterfactuals to be similarly confused by unwillingness to recognize vacuously true

counterpossibles. (Williamson, 2007, 175)

Given the fact that the standard model theory of counterfactuals treats counterfactuals as universal quantifiers over worlds, Williamson's analogy is of course particularly apt. Do we risk entering a kind of logical Dark Age if we accept that counterfactuals such as (valid > halt) and (valid > arith) are non-vacuous? Fortunately, there is no such risk. On the model theory for counterfactuals about relative computability presented in the appendix, these counterfactuals are still universal quantifiers over indices and they still admit of vacuously true instances. In fact, the model theory is of a piece with Lewis' similarity models; it incorporates a version of the vacuity thesis insofar as it treats counterfactuals with outright logical falsehoods in the antecedents as vacuously true.

The basic idea of the model theory is simple. Relative computability theory provides us with an abstract structure called the *Turing degrees*. Informally, we can say that this structure classifies sets of natural numbers into complexity classes. The halting problem and the validity problem belong to the same complexity class, which is why (valid > halt) and its converse are true, but arithmetical truth belongs to class of problems of much higher complexity, which is why (valid > arith) is false. The Turing degrees form a hierarchy that has the form of an infinite tree originating from a single point.[47] This point of origin is the class of least complex sets, i.e. the sets that are in fact computable. For example, the set ω of all natural numbers belongs to this class, since we can easily come up with an algorithm for deciding it: for any number n, to decide whether n is in ω, compute nothing and output 'yes.' Another way of thinking of this least class is that it represents something like the actual world: everything that's actually algorithmically decidable is represented by this class as algorithmically decidable. This is the class where the Church-Turing holds and so where the laws of computation are as they actually are. So it's tempting to just have the Turing degrees play the role of worlds, where all of the Turing degrees except for the one that stands for the actual one are thought of as non-actual worlds where the laws of computation are different. The further you move up the tree, the more violations of the Church-Turing thesis you get, since more and more sets that aren't actually algorithmically decidable become represented as algorithmically decidable.

---

[47]I use 'tree' in an informal sense here. In its technical sense, trees are well-founded, which the Turing degrees aren't, due to the Sacks Density Theorem.

This tree-like structure gives us everything we need for Lewis' comparative similarity semantics for the counterfactual connective. Unfortunately, this isn't quite right, for reasons explained in the appendix. What we need for our worlds are rather *ideals* on Turing degrees. The ideals still form a tree-like structure on which we can build Lewis' comparative similarity semantics. A simple counterfactual ⌜If B were algorithmically decidable, then A would be algorithmically decidable⌝ is true at a world w (i.e. an ideal on Turing degrees) iff all worlds closest to w that represents B as algorithmically decidable also represent A as algorithmically decidable.[48] We can turn this into a fully general semantics for the counterfactual connective by incorporating the standard semantic clauses for the Boolean connectives and the quantifiers. As long as the semantic clauses for the connectives are classical, ⌜$(\phi \wedge \neg\phi) \:\Box\!\!\rightarrow\: \psi$⌝ comes out vacuously true, for any ψ, since there's no world where ⌜$\phi \wedge \neg\phi$⌝ is true. Again, for more details, see the appendix, and for a complete axiomatization of a propositional fragment of what I call *a conditional logic of Turing reducibility*, see chapter 4.

Let's take stock. Not only do we have positive reasons for interpreting counterfactuals about relative computability literally, as seen in the previous section, but we can also see now that nothing stands in the way of extending Lewis' similarity models to give a model theory for these counterfactuals. The resulting theory doesn't have us falling back into a logical Dark Age that Williamson has warned us of. Our job isn't done, however. One big remaining question is how to interpret our model theory. Even though the ideals on Turing degrees in the model theory just sketched act like worlds as far as the model theory is concerned, they are of course a very different kind of object than what we usually think of when we think of worlds, possible or impossible. I take up this issue in the next section.

3.6 Interpreting the indices

The reason why the ideals on Turing degrees, which are just sets of sets of sets of natural numbers, act like worlds as far as the above model theory is concerned says more about the model theory than

---

[48] Since for any set, there's a closest world where that set is algorithmically decidable, this way of glossing the semantic clause is apt. But since the structure of the Turing degrees is dense, a fully accurate statement, such as the one given in the appendix, will have to be slightly more complicated in that we can't rely on the limit assumption.

about the ideals. As is well known, so-called "possible worlds" model theory doesn't presuppose any kind of realism about possible worlds. As a piece of mathematics, the model theory doesn't care what the "worlds" are that we use. These worlds are just indices at which we evaluate sentences. So there's nothing mysterious about the fact that ideals on Turing degrees can act as indices.

However, we may still ask what possible worlds model theory is *for*, and depending on what we think it's for, we may want to ask some more probing questions about how to interpret the role of the ideals on Turing degrees in the above model theory. Of course, it is beyond the scope of this chapter to develop a theory of model theory. But I want to make a few remarks about how my proposed model theory fits into two alternative pictures of the role of model theory.

On an *instrumentalist* understanding of possible worlds model theory, possible worlds models are merely a useful tool to study the logic of the object languages in question. There's no doubt that possible worlds model theory has greatly advanced our understanding of modal and counterfactual logic. But an appreciation of the usefulness of model theory is consistent with the rejection of any sort of realism about possible worlds. One form of such instrumentalism is *modalism*.[49] Modalism claims that the modal operators and counterfactual connectives are in some sense more basic than the possible worlds used in their model theory. Kit Fine (1977) explicitly speaks of the *construction* of possible worlds. So the rough idea is that we "construct" possible worlds using our modal and counterfactual language and then use them to obtain a more precise understanding of that language. This take on possible worlds model theory fits particularly well with the way we use the ideals on Turing degrees in the above model theory. For after all, as described in the appendix, the ideals on Turing degrees are selected from among the mathematical universe to play the role of worlds with the help of the Turing reducibility relation. The Turing reducibility relation in turn corresponds to the relation of reducibility simpliciter, via the Post-Turing thesis. And as we've seen, reducibility simpliciter is best cashed out in counterfactual talk. So on a modalist-instrumentalist understanding of possible worlds model theory, there is no puzzle about the role of the ideals on Turing degrees in our model theory.

---

[49]See Fine (1977), Forbes (1989, 1992), and Williamson (2013, §8.4) for modalism about metaphysical possibility. See also Williamson's (2009, 9) related remarks about the role of Lewis' comparative similarity relation in the analysis of counterfactuals, as well as Stalnaker's (1984, ch. 8) related remarks about the role of his selection functions.

There is also a more inflationary understanding of possible worlds model theory, the *representa-tional* understanding. The idea here is that there is a privileged possible worlds model, the one that corresponds to the semantics of our language, and that model represents the truth conditions of our sentences.[50] Such a representational understanding of course presupposes a kind of realism about possible worlds. But that realism needn't be as strong as Lewis' (1986); a weaker realism, such as perhaps Stalnaker's (2003; 2012), suffices.[51] Given such realism, the question how the ideals on Turing degrees *qua* indices relate to possible worlds becomes pressing. Whatever possible worlds are, they surely aren't sets of sets of sets of natural numbers. So if we want to give genuine truth conditions for counterfactuals about relative computability, an appeal to ideals on Turing degrees is bound to be unilluminating. However, a representational understanding of our model theory may be available. Suppose there are worlds, possible or otherwise, where the laws of computability are different from what they actually are. And suppose that for any set that appears somewhere in the structure of the ideals on Turing degrees, there's such a world where that set is algorithmically de-cidable. Then we can define a partition on the set of all of these worlds such that two worlds are in the same cell iff they agree on the laws of computability. We will then be able to define a model that's isomorphic to the model I present in the appendix where the indices are the cells of the par-tition. What's more, the Post-Turing will guarantee that the resulting truth conditions for sentences such as (valid > halt) and (valid > arith) will be adequate. And if we want to provide an intended model for a language in which we can talk about more than just algorithmic decidability, we can take this new model and extend the comparative similarity relation to the members of the cells of the partition. This will allow us to assign truth-conditions to counterfactuals whose component sen-tences talk both about algorithmic decidability as well as about all things other than algorithmic decidability. Of course, this may lead us to assign truth conditions to counterfactuals that involve odd, gerrymandered pairings of sentences about algorithmic decidability and sentences having noth-ing whatsoever to do with algorithmic decidability. But we can of course have counterfactuals with similarly odd pairings even in the absence of an ability to talk about algorithmic decidability. Such

---

[50]I borrow the expression 'representational' from Etchemendy's (1990, ch. 1) closely related notion of a representational semantics. Note that on a supervaluational treatment of vagueness, we would have a *class* of privileged models, not a single one.

[51]See Berto (2013, §3) for an overview of various theories of possible and impossible worlds.

is the nature of compositionality. Perhaps some such pairings will lead us to adopt, say, a model theory that allows for truth-value gaps so that we aren't required to count every counterpossible as either true or false. But there's no reason for thinking that the introduction of an ability to talk about algorithmic decidability will put any pressure on us to go in for such maneuvers that wasn't already there before.

Of course, some will doubt the intelligibility of metaphysically impossible worlds where the laws of computability are different, given the metaphysical necessity of the Church-Turing thesis. Echoing Bertrand Russell's (1905) and W. V. Quine's (1948) criticisms of Meinongian ontology, Lewis (1986, 7 n. 3) and Stalnaker (1996) are suspicious of *logically* impossible worlds where contradictions hold. They argue as follows: suppose that there's an impossible world w at which p and '$\neg p$' are true. Then given that '$\neg p$' is true at w, it's not the case that p is true at w. So it both is and isn't the case that p is true at w. Contradiction. So w can't exist. Whatever the force of this objection may be, it clearly doesn't apply to the present use of metaphysically impossible worlds. For none of the worlds required by our model theory are logically impossible.[52] And clearly, a version of the Stalnaker-Lewis argument against our impossible worlds won't go through. Essentially, we are saying that there are worlds where the Church-Turing thesis fails. To get a contradiction from this, we would need the assumption that the Church-Turing thesis holds in every world. But all I've argued is that the Church-Turing thesis holds in every *metaphysically possible* world. More generally, if we're representationalists about worlds model theory, then our metaphysically impossible worlds earn their keep for much the same reason that metaphysically possible worlds earned their keep: as we saw, they allow us to develop truth conditions for a certain class of counterfactuals.

We thus see that no matter whether we're instrumentalists or representationalists about our model theory, there's no serious worry about its use of indices that represent the laws of computation as different from what they actually are.

---

[52]Of course, if we want to allow for non-vacuous counterfactuals with logically inconsistent antecedents, we will have to face this objection head on. See Berto (2013, §6) for an overview of responses to this objection. In any case, accepting non-vacuous counterfactuals with merely metaphysically impossible antecedents doesn't immediately commit us to such stronger failures of the vacuity thesis.

3.7 Conclusion

The case for the vacuity of counterfactuals about relative computability looks feeble. We've seen that the reducibility relation, which is the subject of study of relative computability theory, is to be understood in terms of counterfactuals. These counterfactuals have metaphysically impossible antecedents, and so the vacuity thesis threatens to undermine a whole mathematical discipline. Philosophical humility recommends that we revise our theory of counterfactuals before we propose to put our colleagues in mathematics out of a job.

Some questions still remain, however. First, the representational understanding of worlds model theory gives rise to general questions about the metaphysics of worlds, and about whether metaphysically possible worlds are the same kind of thing as metaphysically impossible worlds. These questions are beyond the scope of the present chapter.

Another question concerns the status of the outer necessity operator that I briefly discussed at the end of section 3.4. Is there a theoretically important modality corresponding to this operator that's of the same kind as metaphysical necessity, though more strict? Accepting an ideology of outer necessity would arguably be the most conservative way of amending the orthodoxy, since it would allow us to hold on to a version of the vacuity thesis. In fact, if we accept this ideology, then counterfactuals about relative computability turn out not to be counterpossibles at all, at least not as far as outer possibility is concerned. Whether the ideology of outer possibility is worth accepting for this and other reasons will have to be judged against the same kind of criteria that are used to answer questions about ideological commitment in general.

Finally, one may wonder how the theory I've developed extends to counterpossibles that aren't about relative computability, such as perhaps (water) and (marvel) mentioned on page 90. I submit that my discussion gives us reason to take seriously the suggestion, made of course by many in the literature, that there are other non-vacuous counterpossibles. But I also hope that my discussion has shown that careful investigation is required to establish that a given counterpossible is indeed non-vacuous. In particular, since many of the purported counterexamples to the vacuity thesis mentioned in the literature such as those involving claims about what would have happened

if the laws of metaphysics had failed are about philosophical topics, an appeal to philosophical humility such as the one I invoke above may not always be available. What we have here is a classic case where one philosophical domain, i.e. metaphysics, is in tension with another, i.e. philosophical semantics. To move beyond the gridlock in the debate over counterpossibles, we need to look for uses of counterpossibles outside of philosophy. I therefore suggest that we seek to find established scientific disciplines other than relative computability theory where counterpossibles play a central role. It is my hope that the present study has taken a first step towards such a case-by-case study of counterpossibles. Once we have a clearer picture of the areas where non-vacuous counterpossibles are indispensable and once we have model theories for these various classes of counterpossibles, we may then investigate to what extent we can integrate these model theories to come up with a unified and fully general theory of non-vacuous counterpossibles.

3.8   Appendix: Model theory

In this appendix, I describe a model for a quantified language of relative computability with a designated predicate 'D' for algorithmic decidability.[53]

The Turing degree of some set A is $\deg(A) = \{B : A \leq_T B \text{ and } B \leq_T A\}$. We can define an ordering $\leq$ on the Turing degrees $\mathbf{D}$ so that for $\mathbf{a}, \mathbf{b}$ Turing degrees, $\mathbf{a} \leq \mathbf{b}$ iff there's some $A \in \mathbf{a}$ and some $B \in \mathbf{b}$ such that $A \leq_T B$. Informally, the Turing degree of A is its complexity class. I mentioned that it's tempting to think of Turing degrees as worlds, where a degree-world would represent a set as decidable iff it contains that set. However, this would mean, for example, that there is no world where the decidable sets are all *and only* the arithmetically definable sets. This follows from Corollary 1 of §4.4 in Kleene and Post (1954) that there's no degree that contains all *and only* the arithmetical sets, since $\mathbf{0}^{(\omega)}$ isn't a minimal upper bound to the arithmetical degrees $\mathbf{0}, \mathbf{0}', \mathbf{0}'', \ldots$.[54] We can avoid this undesirable result if we use *ideals* on Turing degrees instead. For any $\mathbf{a}, \mathbf{b} \in \mathbf{D}$ and for $\mathbf{0}$ the degree of the algorithmically decidable sets, an ideal $\mathbf{i}$ on the Turing

---

[53]A construction similar to the present one that's based on the enumeration degrees would yield a model for a language with a designated predicate for computable enumerability. See Odifreddi (1992, ch. XIV) for an introduction to enumeration degrees.

[54]This is how Rogers, Jr. (1967, 276) puts the result in Corollary XVI of §13.4.

degrees is a non-empty set of Turing degrees such that if $\mathbf{a}, \mathbf{b} \in \mathbf{i}$, then their join $\mathbf{a} \oplus \mathbf{b}$ is in $\mathbf{i}$ as well; and if $\mathbf{a} \in \mathbf{i}$ and $\mathbf{b} \leq \mathbf{a}$, then $\mathbf{b} \in \mathbf{i}$. Since the join of two arithmetically definable Turing degrees is arithmetically definable and since anything reducible to an arithmetical set is arithmetical, the arithmetical sets form an ideal.

The starting point for our model theory are the frames for Lewis' (1971; 1973) comparative similarity models, which consist of a set of indices (worlds) $\mathfrak{W}$ and a ternary relation $\mathfrak{R}$ on $\mathfrak{W}$ such that for each $w \in \mathfrak{W}$, $\mathfrak{R}_w$ is a total binary preordering on $\mathfrak{W}$. $v\mathfrak{R}_w u$ is informally understood as saying that world v is at least as similar to w as u is to w.[55]

The structure of the Turing degrees $\langle \mathbf{D}, \leq \rangle$ is very similar to such frames. It is easily seen that $\leq$ partially orders $\mathbf{D}$. What is more difficult to see is that $\leq$ isn't total; there are $\mathbf{a}$ and $\mathbf{b}$ in $\mathbf{D}$ such that $\mathbf{a} \nleq \mathbf{b}$ and $\mathbf{b} \nleq \mathbf{a}$. This is Corollary 1 in §2.2 of Kleene and Post (1954), which we've already encountered. But we already saw that we can't use $\mathbf{D}$ to serve as the set of worlds. Rather, we need to use the set $\mathbf{I}$ of ideals on Turing degrees. This set already comes partially ordered by the subset relation. But still, a difference between $\langle \mathbf{I}, \subseteq \rangle$ (besides the fact that $\subseteq$ isn't total, due to the non-totality of $\leq$) is that $\subseteq$ is a binary relation whereas Lewis' $\mathfrak{R}$ is ternary. This turns out not to be a problem, however.

For $\langle \mathbf{I}, \subseteq \rangle$ a frame, our model is the tuple $\mathfrak{M} = \langle \wp(\omega), \mathbf{I}, \subseteq, \mathfrak{I} \rangle$, where $\wp(\omega)$ is the power set of the set of natural numbers and $\mathfrak{I}$ takes 'D' to functions from members of $\mathbf{I}$ to subsets of $\wp(\omega)$ such that for $w \in \mathbf{I}$ and $x \in \wp(\omega)$, $x \in \mathbf{I}('D')(w)$ iff for some $y \in \bigcup w$, $x \leq_T y$. For g a function that assigns members of $\wp(\omega)$ to the variables of the language, we then have that 'Dx' is satisfied at a world w iff $g(x) \in \mathfrak{I}('D')(w)$.[56] The counterfactual connective '$\square\rightarrow$' is defined as follows (where $\mathfrak{W}_w = \{v \in \mathfrak{W} : w \subseteq v\}$ and $\llbracket \phi \rrbracket_{\mathfrak{M}}^g$ is shorthand for $\{w \in \mathfrak{W} : \llbracket \phi \rrbracket_{\mathfrak{M},w}^g = 1\}$): $\llbracket \ulcorner \phi \square\rightarrow \psi \urcorner \rrbracket_{\mathfrak{M},w}^g = 1$ iff for all $v \in \mathfrak{W}_w \cap \llbracket \phi \rrbracket_{\mathfrak{M}}^g$, there is some $u \in \mathfrak{W}_w \cap \llbracket \phi \rrbracket_{\mathfrak{M}}^g$ such that $u \subseteq v$ and such that for any $t \in \mathfrak{W}_w$ such that $t \subseteq u$, $\llbracket \ulcorner \phi \rightarrow \psi \urcorner \rrbracket_{\mathfrak{M},t}^g = 1$. Note that this clause for '$\square\rightarrow$,' which is adapted from Burgess' (1981), differs from Lewis' clause in that it contains an additional

---

[55]Of course, by building on Lewis' model theory, we also inherit some of the potential problems of the Stalnaker-Lewis approach to counterfactuals. For example, it doesn't validate simplification of disjunctive antecedents (see footnote 35). If simplification is indeed desirable, the present model theory can be adapted along the lines developed by Fine (2012) or Santorio (2017) to accommodate it.

[56]Note that for reasons of perspicuity, I use the usual letters 'w,' 'v,' 'u,' and 't' to denote "world" variables here, even though I previously used boldface letters as variables for the members of $\mathbf{I}$.

initial universal quantifier. This is required because our partial order isn't total, whereas Lewis' comparative similarity relations are. Note also that our binary partial order can be turned into a ternary comparative similarity relation in a canonical way: we define the ternary relation $\subseteq^*$ such that $\mathbf{j} \subseteq^*_i \mathbf{k}$ iff $\mathbf{i} \subseteq \mathbf{j}$ and $\mathbf{j} \subseteq \mathbf{k}$. This gives us the frame $\langle \mathbf{I}, \subseteq^* \rangle$, on which we can build models each of which belongs to (the quantified version of) John Burgess' (1981) model class $\mathcal{M}_{0,1}$. If we then redefine $\mathfrak{W}_w$ as $\{v \in \mathfrak{W} : v \subseteq_w v\}$ and take over the above clause for '$\square\!\rightarrow$,' we immediately get that the ternary version of our model on $\langle \mathbf{I}, \subseteq^* \rangle$ validates all axioms and rules of Burgess' (1981) logic $S_{0,1}$. $S_{0,1}$ is strictly weaker than Lewis' (1971) favored counterfactual logic $C1$, which we obtain from $S_{0,1}$ by adding:[57]

D′. $\left((\phi \vee \psi) \square\!\rightarrow \neg\phi \right) \rightarrow \left(\left((\phi \vee \chi) \square\!\rightarrow \neg\phi \right) \vee \left((\psi \vee \chi) \square\!\rightarrow \neg\chi \right) \right)$

And of course from $C1$ we can get Stalnaker's (1968) logic $C2$ by adding conditional excluded middle:

CEM. $(\phi \square\!\rightarrow \psi) \vee (\phi \square\!\rightarrow \neg\psi)$

Neither D′ nor CEM are valid in $\mathfrak{M}$, due to the fact that Corollary 1 in §2.2 of Kleene and Post (1954) makes $\subseteq$ non-total. Regarding the quantifiers, since these models have a fixed domain, the Barcan (1946) formula and its converse come out valid.

Of course, the frame $\langle \mathbf{I}, \subseteq \rangle$ has certain features that not all frames have on which the models in Burgess' $\mathcal{M}_{0,1}$ are built. In fact, the structure of the ideals on Turing degrees is an upper semi-lattice with a zero-element, and it has many other features that we may wish to capture axiomatically. I provide a complete axiomatization of a propositional fragment of the conditional logic of Turing reducibility as well as a decision procedure in chapter 4. For the quantificational case, we may want to enrich our language with a predicate for computable enumerability and with function signs for the complementation, jump, and join operations on sets of natural numbers. Whether the structure of the ideals on the Turing degrees can be completely axiomatized is unknown. What's important

---

[57]See also Pollock (1976, 43) for a related logic SS, which can be turned into $C1$ by adding:

$\left((\phi \square\!\rightarrow \psi) \wedge \neg(\phi \square\!\rightarrow \neg\chi) \right) \rightarrow \left((\phi \wedge \chi) \square\!\rightarrow \psi \right)$

for present purposes is that with our model $\mathfrak{M}$ we have what we need to correctly interpret (regimentations of) our counterfactuals (valid > halt) and (valid > arith) (see page 80), quantified counterfactuals such as (sacks) (page 91), as well as many more.

Before we can regiment (valid > halt) and (valid > arith), we should expand our language to include individual constants 'v,' 'h,' and 'a' that $\mathfrak{I}$ assigns to the validity problem, the halting problem, and arithmetical truth respectively. Then (valid > halt) and (valid > arith) become '$Dv \boxbox\to Dh$' and '$Dv \boxbox\to Da$' respectively. Given that the degree of both the validity and the halting problem is the degree $\mathbf{0}'$ and the degree of arithmetical truth is $\mathbf{0}^{(\omega)}$ and given that $\mathbf{0}' \leq \mathbf{0}^{(\omega)}$, '$Dv \boxbox\to Dh$' comes out true at the zero-element "world" in $\mathfrak{M}$ and '$Dv \boxbox\to Da$' comes out false, as desired. In fact, as long as we have a model on the frame $\langle \mathbf{I}, \subseteq^* \rangle$ that assigns to the atomic sentences of our language the intended set ideals on of Turing degrees, our model theory gives exactly the results we want. For example, it is routine to verify that the relevant regimentations of (valid > $\overline{\text{arith}}$), (valid&arith > $\overline{\text{arith}}$), (valid > halt&$\overline{\text{arith}}$), and (valid&arith > halt&$\overline{\text{arith}}$) (see page 94) have all the desired properties in our model. And since the structure of the Turing degrees is dense, the relevant regimentation of (sacks) is true at any world in our model as well:

$$\forall x \forall y \big((Dy \boxbox\to Dx) \wedge \neg(Dx \boxbox\to Dy) \big) \to$$

$$\exists z \big((Dz \boxbox\to Dx) \wedge \neg(Dx \boxbox\to Dz) \wedge (Dy \boxbox\to Dz) \wedge \neg(Dz \boxbox\to Dy) \big)$$

Chapter 4

A Conditional Logic of Turing Reducibility

4.1 Introduction

Informal expositions of the theory of Turing reducibility often make use of counterfactual conditionals such as the following (where Val, K_0, and K_ω are the sets of natural numbers that code the validity problem in the predicate calculus, the halting problem for Turing machines, and arithmetical truth, respectively):

> If K_0 were algorithmically decidable, then Val would be algorithmically decidable as well.

> Even if K_0 were algorithmically decidable, K_ω would still not be algorithmically decidable.[1]

Chapter 3 argues that such conditionals play an ineliminable role in the development of the theory of Turing reducibility. Its appendix also presents models for a quantified counterfactual language that can express claims such as the above using the counterfactual connective $\Box\!\!\rightarrow$. The theory of Turing reducibility supplies us with the structure of the Turing degrees $\langle \mathbf{D}, \leq \rangle$ that we can use to build models that are close relatives of the the comparative similarity models studied in Lewis (1971),

---

[1] See for example Davis (1958, p. 179), Rogers, Jr. (1967, p. 127), and Enderton (2011, p. 121). When talking about the theory of Turing reducibility here, we adopt the notation of Soare (2016). We also appeal to many facts proven there.

Burgess (1981), and Veltman (1985). First, we define the set $\mathbf{I}$ of ideals on $\mathbf{D}$, where an ideal $\mathbf{i}$ is a non-empty set of degrees that is closed under reducibility $\leq$ and the join operation $\oplus$. That is, for $\mathbf{i} \in \mathbf{I}$ and for $\mathbf{a}, \mathbf{b} \in \mathbf{D}, \mathbf{i} \neq \emptyset$, if $\mathbf{a} \in \mathbf{i}$ and $\mathbf{b} \leq \mathbf{a}$ then $\mathbf{b} \in \mathbf{i}$, and if $\mathbf{a}, \mathbf{b} \in \mathbf{i}$ then $\mathbf{a} \oplus \mathbf{b} \in \mathbf{i}$. We can treat $\mathbf{I}$ as the set of indices, or "worlds."

Intuitively, a world in our models can be understood as representing what sets of natural numbers are algorithmically decidable at that world, namely all of those that are a member of one of its members. For example, the world $\{\mathbf{0}\}$ represents the decidability facts as they actually are since its only member contains exactly those sets that are actually algorithmically decidable. In other words, $\{\mathbf{0}\}$ represents the decidability facts in accordance with the Church-Turing thesis, and none of the other worlds do. Furthermore, because $\langle \mathbf{D}, \leq \rangle$ is an upper semilattice with a minimal element, $\langle \mathbf{I}, \subseteq \rangle$ is also an upper semilattice with a minimal element. The set inclusion relation can be understood as representing the comparative similarity among worlds: if $\mathbf{a} \subseteq \mathbf{b}$, then either $\mathbf{a} = \mathbf{b}$ or else $\mathbf{a}$ is more similar to the bottom world $\{\mathbf{0}\}$ than $\mathbf{b}$ because $\mathbf{a}$ represents fewer sets as algorithimcally decidable that are not actually algorithimcally decidable than $\mathbf{b}$ and so it violates the Church-Turing thesis less dramatically.

A word on the choice of ideals on Turing degrees as our "worlds": it is tempting to think of the Turing degrees themselves as "worlds," where such a world would represent a set as decidable iff it contains that set. However, this would mean, for example, that there is no world where the decidable sets are all *and only* the arithmetically definable sets. This follows from Corollary 1 of §4.4 in Kleene and Post Kleene and Post (1954) that there is no degree that contains all *and only* the arithmetical sets, since $\mathbf{0}^{(\omega)}$ is not a minimal upper bound to the arithmetical degrees $\mathbf{0}, \mathbf{0}', \mathbf{0}'', \ldots$ (this is how the theorem is put in Corollary XVI of §13.4 in Rogers, Jr. (1967)). We can avoid this undesirable result if we use ideals on Turing degrees instead. However, it is easily verified that using Turing degrees instead of their ideals as worlds would not change the present logic because the simple language we are working with cannot distinguish between the two upper semilattices.

Given the frame $\langle \mathbf{I}, \subseteq \rangle$ and given a function f that maps the atomic sentences of our language to sets of natural numbers, we can define a model $\langle \mathbf{I}, \{\mathbf{0}\}, \subseteq, \mathfrak{I} \rangle$, where $\{\mathbf{0}\}$ is thought of as the designated "actual world" and where $\mathfrak{I}$ is defined so that for α atomic, $\mathfrak{I}(\alpha) = \{\mathbf{i} \in \mathbf{I} : \text{for some}$

$\mathbf{d} \in \mathbf{i}, f(\alpha) \in \mathbf{d}\}$. We then have that α is true at an ideal/world w iff $f(\alpha)$ is a member of one of the members of w. Relative to such a model, α can then be read as saying that the set of natural numbers $f(\alpha)$ is algorithmically decidable.

We immediately have that if an atomic sentence is true at a world, it remains true as we move up the partial order induced by $\subseteq$. Note also that since every $N \in \omega$ is in some Turing degree, namely the degree $\{M \in \omega : N \leq_T M$ and $M \leq_T N\}$, there is some world for every atomic α where α is true. The conditional connective $\square\rightarrow$ is defined in the same way in which Burgess (1981), Lewis (1981), and Veltman (1985) define it given a ternary comparative similarity relation: where ϕ is said to be a *possibility* at a world $\mathbf{a}$ iff there is some world $\mathbf{b}$ such that $\mathbf{a} \subseteq \mathbf{b}$ where ϕ is true, we say that $\phi \square\rightarrow \psi$ is valid in a model $\langle \mathbf{I}, \{\mathbf{0}\}, \subseteq, \mathfrak{I} \rangle$ iff either $\phi \wedge \psi$ is a less remote possibility than $\phi \wedge \neg\psi$ at the designated, minimal world $\{\mathbf{0}\}$ or ϕ is impossible at $\{\mathbf{0}\}$.

Based on the above understanding of $\square\rightarrow$, we present a complete axiomatization of the propositional fragment of a conditional logic of Turing reducibility for a language where the counterfactual conditionals are restricted so that they are not allowed to contain any conditional antecedents (but they are allowed in the consequents of conditionals). Our logic is called PT, in honor of Post (1944) and Turing (1939). In addition to being a complete axiomatization of the sentences of the language that are valid in all intended models, PT is decidable.

We do not mean to present PT as *the* conditional logic of Turing reducibility. In fact, we have to make certain non-trivial choices when devising our model theory. For example, we assume that algorithmic decidability is closed under joins not just actually, but also counterfactually. We also assume that algorithmic decidability is a property of sets of natural numbers, not of "problems" in a more informal sense. Let S be the set of natural numbers n such that the power of the continuum is greater than $\aleph^n$. It is algorithmically decidable whether $1 \in S$. However, it is famously impossible to decide whether the power of the continuum is greater than $\aleph^1$ using currently accepted mathematical means. Now, if we had treated algorithmic decidability as a property of problems instead of as a property of sets of natural numbers, then we would have the following argument, a version of which was suggested to us by an anonymous reviewer, for the claim that if the halting problem were algorithmically decidable, then arithmetical truth would be algorithmically decidable as well:

117

Suppose that the halting problem is algorithmically decidable. Now consider the following algorithm A: for an algorithm B, A tests whether B halts on input 0. If it does not, then A halts. If it does, then A tests whether B halts on input 1. If it does not, then A halts. And so on. Thus, A halts iff the function computed by B is not total. But this means that to test whether the function computed by some algorithm B is total, we only need to determine whether A halts when given (the code of) B as input. And since we assume that the halting problem is algorithmically decidable, this means that we also have that the problem of determining whether a computable function is total is algorithmically decidable. Now, the halting problem is Σ_1^0-complete, while the problem of determining whether a computable function is total is Π_2^0-complete. So, the foregoing shows that if the halting problem were algorithmically decidable, then Π_2^0-complete problems would algorithmically decidable as well. And an analogous argument would show that if the halting problem were algorithmically decidable, then any problem in the arithmetical hierarchy would be algorithmically decidable, in which case arithmetical truth would be decidable.

This is a clever argument, but it does not go through on the assumption that algorithmic decidability is a property of sets of natural numbers. 'K_0' is a *rigid designator* in the sense of Kripke (1980). In other words, the referent of the term does not change as we change the facts about algorithmic decidability: even if K_0 were algorithmically decidable, the set K_0 would still have the same members that it actually does. In contrast, the above argument rests on the assumption that if the halting problem were algorithmically decidable, then the problem of determining whether a machine halts on a given input would encompass not only the algorithms that compute functions that are actually computable but also algorithms that exist only counterfactually and that do not compute functions that are actually computable.

That devising our logic involved such non-trivial choices should not come as too much of a surprise. For, it is well known from the following pair of conditionals due to Quine (1960, 222), that judgments about conditionals are heavily influenced by contextual factors:

If Caesar had been in command in Korea, then he would have used the atom bomb.

If Caesar had been in command in Korea, then he would have used catapults.

We can easily imagine contexts in which the first one is true but not the second, but also ones where the second one is true but not the first. Given this malleability of conditionals, we cannot expect the conditional logic of Turing reducibility to be as straightforward as, for example, the logic of provability Boolos (1993) becomes once we settle on *Bew* as representing provability.

Recall that the present syntax does not allow for nestings of conditionals inside the antecedents of conditionals. Here we find a precedent in Briggs' logic of counterfactuals Briggs (2012). What is more, it is not obvious that there is an urgent need for such an axiomatization, for it is not easy to get a good handle on versions of the above conditionals about algorithmic decidability that have conditional antecedents. In any case, it is not known whether the present model theory, applied to the full propositional language without syntactic restrictions, can be completely axiomatized. Furthermore, the appendix to chapter 3, also discusses a quantified language with a designated predicate for algorithmic decidability; it is another open problem whether the quantified conditional logic of Turing reducibility can be completely axiomatized, that is whether it is computably enumerable.

While the present study is concerned with conditionals about algorithmic decidability, the formalism developed here also lends itself to a study of conditionals about polynomial reducibility or enumeration reducibility:[2]

> If this set is decidable in polynomial time, then that set is decidable in polynomial time as well.

> If this set were computably enumerable, then that set would be computably enumerable as well.

4.2 Syntax

We first define the language $\mathscr{L}$ of PT.

Definition 4.2.1.   (i) The *atomic sentences* α of $\mathscr{L}$ are the members of $\{p_i : i \in \omega\}$.

---

[2]See Papadimitriou (1994) and Rogers, Jr. (1967, §9.7), respectively.

(ii) The *conditional-free sentences* β of $\mathscr{L}$ have the following BNF definition:

(4.1)
$$\beta ::= \alpha \mid \bot \mid \neg\beta \mid (\beta \wedge \beta)$$

(iii) The *sentences* ϕ of $\mathscr{L}$ have the following BNF definition:

$$\phi ::= \beta \mid \neg\phi \mid (\phi \wedge \phi) \mid (\beta \mathbin{\Box\!\!\rightarrow} \phi)$$

Remark 4.2.1. $(\phi \vee \psi)$, $(\phi \rightarrow \psi)$, $(\phi \leftrightarrow \psi)$, $\top$, $\Box\phi$, and $\Diamond\phi$ are the usual metalinguistic abbreviations for $\neg(\neg\phi \wedge \neg\psi)$, $(\neg\phi \vee \psi)$, $((\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi))$, $\neg\bot$, $(\neg\phi \mathbin{\Box\!\!\rightarrow} \bot)$, and $\neg\Box\neg\phi$. In the metalanguage, we also often omit parentheses where there is no threat of ambiguity. For finite $\Gamma \subseteq \mathscr{L}$, $\bigwedge\Gamma$ and $\bigvee\Gamma$ denote a suitably grouped conjunction and a suitably grouped disjunction of the elements of Γ respectively, where $\bigwedge\emptyset = \top$ and $\bigvee\emptyset = \bot$.

Remark 4.2.2. Since $\mathscr{L}$ does not contain the material conditional connective $\rightarrow$ as a primitive connective, in what follows, we will invariably use 'conditional sentence' or just 'conditional' to refer to sentences of the form $\phi \mathbin{\Box\!\!\rightarrow} \psi$.

Definition 4.2.2. For $\Gamma \subseteq \mathscr{L}$, $\mathscr{L}_\Gamma \subseteq \mathscr{L}$ is the set of sentences that contains only atomic sentences that are also contained in some member of Γ.

4.3 Models

Definition 4.3.1. (i) A PT-*frame* $\mathfrak{F}$ is a pair $\langle \mathfrak{W}, \leq \rangle$ that is an upper semilattice with a minimal element.

(ii) For $\mathfrak{F} = \langle \mathfrak{W}, \leq \rangle$ a PT-frame, a PT-*model* $\mathfrak{M}$ is a tuple $\langle \mathfrak{W}, w_@, \leq, \mathfrak{I} \rangle$ such that $w_@$ is the minimal element of $\mathfrak{W}$ and such that $\mathfrak{I}$ is a function from atomic sentences $\alpha \in \mathscr{L}$ to subsets of $\mathfrak{W}$ such that:

   (a) for all atomic sentences $\alpha \in \mathscr{L}$ and for all $w \in \mathfrak{W}$, there is exactly one $\leq$-least $v \in \mathfrak{W}$ such that $w \leq v$ and $v \in \mathfrak{I}(\alpha)$,

   (b) for all $w, v \in \mathfrak{W}$, for $\alpha \in \mathscr{L}$ atomic, if $w \in \mathfrak{I}(\alpha)$ and $w \leq v$, then $v \in \mathfrak{I}(\alpha)$.

We record, without proof, two immediate consequences of this definition:

Lemma 4.3.1. Let $\mathfrak{M}$ be a PT-model and let $\mathfrak{W}_w = \{v : w \leq v\}$. Then the following hold:

(i) For all $w \in \mathfrak{W}$, $\mathfrak{W}_w \neq \emptyset$,

(ii) for $\alpha_1, \ldots, \alpha_n \in \mathscr{L}$ atomic sentences and for all $w \in \mathfrak{W}$, there is exactly one $\leq$-least $v \in \mathfrak{W}$ such that $w \leq v$ and $v \in \mathfrak{I}(\alpha_1) \cap \cdots \cap \mathfrak{I}(\alpha_n)$.

Definition 4.3.2. $[\![\cdot]\!]_{\mathfrak{M}}$ is a function from $\mathscr{L} \times \mathfrak{W}$ to $\{0, 1\}$. Where $\mathfrak{W}_w = \{v : w \leq v\}$, $[\![\phi]\!]_{\mathfrak{M},w}$ is shorthand for $[\![\cdot]\!]_{\mathfrak{M}}(\phi, w)$, and $[\![\phi]\!]_{\mathfrak{M}}$ is shorthand for $\{w \in \mathfrak{W} : [\![\phi]\!]_{\mathfrak{M},w} = 1\}$, $[\![\cdot]\!]_{\mathfrak{M}}$ is defined such that for each $w \in \mathfrak{W}$:

(a) $[\![\bot]\!]_{\mathfrak{M},w} = 0$.

(b) For $\alpha \in \mathscr{L}$ atomic, $[\![\alpha]\!]_{\mathfrak{M},w} = 1$ iff $w \in \mathfrak{I}(\alpha)$.

(c) $[\![\neg\phi]\!]_{\mathfrak{M},w} = 1$ iff $[\![\phi]\!]_{\mathfrak{M},w} = 0$.

(d) $[\![\phi \wedge \psi]\!]_{\mathfrak{M},w} = 1$ iff $[\![\phi]\!]_{\mathfrak{M},w} = 1$ and $[\![\psi]\!]_{\mathfrak{M},w} = 1$.

(e) $[\![\phi \boxbox\to \psi]\!]_{\mathfrak{M},w} = 1$ iff for all $v \in \mathfrak{W}_w \cap [\![\phi]\!]_{\mathfrak{M}}$, there is some $u \in \mathfrak{W}_w \cap [\![\phi]\!]_{\mathfrak{M}}$, $u \leq v$, such that for any $t \in \mathfrak{W}_w$ such that $t \leq u$, $[\![\phi \to \psi]\!]_{\mathfrak{M},t} = 1$.

Lemma 4.3.2. (i) $[\![\Box\phi]\!]_{\mathfrak{M},w} = 1$ iff $\mathfrak{W}_w \subseteq [\![\phi]\!]_{\mathfrak{M}}$.

(ii) $[\![\Diamond\phi]\!]_{\mathfrak{M},w} = 1$ iff $\mathfrak{W}_w \cap [\![\phi]\!]_{\mathfrak{M}} \neq \emptyset$.

Proof. (i). ($\Rightarrow$). Suppose $[\![\Box\phi]\!]_{\mathfrak{M},w} = 1$. Then $[\![\neg\phi \boxbox\to \bot]\!]_{\mathfrak{M},w} = 1$. Then for all $v \in \mathfrak{W}_w \cap [\![\neg\phi]\!]_{\mathfrak{M}}$, there is some $u \in \mathfrak{W}_w \cap [\![\neg\phi]\!]_{\mathfrak{M}}$ such that $w \leq u$ and $u \leq v$ and such that for any $t \in \mathfrak{W}_w$ such that $w \leq t$ and $t \leq u$, $[\![\neg\phi \to \bot]\!]_{\mathfrak{M},t} = 1$. Then, setting $t = u$, which we know we can do due to the reflexivity of $\leq$, $[\![\phi]\!]_{\mathfrak{M},u} = 1$. But that cannot be, because $u \in [\![\neg\phi]\!]_{\mathfrak{M}}$. So there cannot be any $v \in \mathfrak{W}_w \cap [\![\neg\phi]\!]_{\mathfrak{M}}$, and so $\mathfrak{W}_w \subseteq [\![\phi]\!]_{\mathfrak{M}}$.

($\Leftarrow$). Suppose $\mathfrak{W}_w \subseteq [\![\phi]\!]_{\mathfrak{M}}$. Then it vacuously holds that for all $v \in \mathfrak{W}_w \cap [\![\neg\phi]\!]_{\mathfrak{M}}$, there is some $u \in \mathfrak{W}_w \cap [\![\neg\phi]\!]_{\mathfrak{M}}$, $u \leq v$, such that for any $t \in \mathfrak{W}_w$ such that $t \leq u$, $[\![\neg\phi \to \bot]\!]_{\mathfrak{M},t} = 1$. Thus, $[\![\neg\phi \boxbox\to \bot]\!]_{\mathfrak{M},w} = 1$, and so $[\![\Box\phi]\!]_{\mathfrak{M},w} = 1$.

(ii). $[\![\Diamond\phi]\!]_{\mathfrak{M},w} = 1$ iff $[\![\neg\Box\neg\phi]\!]_{\mathfrak{M},w} = 1$ iff $\mathfrak{W}_w \not\subseteq [\![\neg\phi]\!]_{\mathfrak{M}}$ iff $\mathfrak{W}_w \cap [\![\phi]\!]_{\mathfrak{M}} \neq \emptyset$. $\boxtimes$

Definition 4.3.3. For $w \in \mathfrak{W}$ and condition C, $v \in \mathfrak{W}_w$ is a *minimal world relative to w meeting C* iff v meets C and there is no $u \in \mathfrak{W}_w$, $u \neq v$, that meets C. (Note that a world can be a minimal world relative to w meeting C even if it is not a unique such world.)

Lemma 4.3.3. (i) Let ϕ be a positive Boolean combination of atomic sentences. Then if $[\![\phi]\!]_{\mathfrak{M},w} = 1$, then for all $v \in \mathfrak{W}_w$, $[\![\phi]\!]_{\mathfrak{M},v} = 1$.

  (ii) Let ϕ again be a positive Boolean combination of atomic sentences. Then for any $w \in \mathfrak{W}$, there is a non-empty finite set $\Gamma \subseteq \mathfrak{W}_w$ of minimal worlds such that for all $v \in \Gamma$, $[\![\phi]\!]_{\mathfrak{M},v} = 1$, and for all $u \in \mathfrak{W}_w$, $[\![\phi]\!]_{\mathfrak{M},u} = 1$ iff for some $v \in \Gamma$, $v \leq u$.

  (iii) Let ϕ be a Boolean combination of atomic sentences, positive or not. Then, for any $w \in \mathfrak{W}$, there is a finite set $\Gamma \subseteq \mathfrak{W}_w$ of minimal worlds such that for all $v \in \Gamma$, $[\![\phi]\!]_{\mathfrak{M},v} = 1$, and such that for every world $u \in \mathfrak{W}_w$, if $[\![\phi]\!]_{\mathfrak{M},u} = 1$, then there is some $v \in \Gamma$ such that $v \leq u$.

  (iv) Let ϕ be a conjunction of literals (i.e. an atomic sentences or negated atomic sentences). Then, for all $w \in \mathfrak{W}$, if there is a world $v \in \mathfrak{W}_w$ such that $[\![\phi]\!]_{\mathfrak{M},v} = 1$, then there is a unique minimal such world.

Proof. Follows immediately from 4.3.1. $\boxtimes$

Definition 4.3.4. (i) $\phi \in \mathscr{L}$ is *valid in a* PT-*model* $\mathfrak{M}$ iff $[\![\phi]\!]_{\mathfrak{M},w_@} = 1$. (We also say that $\mathfrak{M}$ *validates* ϕ.)

  (ii) $\Gamma \subseteq \mathscr{L}$ is *valid in a* PT-*model* $\mathfrak{M}$ iff for all $\gamma \in \Gamma$, γ is valid in $\mathfrak{M}$. (We also say that $\mathfrak{M}$ *validates* Γ.)

  (iii) $\phi \in \mathscr{L}$ is PT-*valid* ($\vDash_{\mathsf{PT}} \phi$) iff it is valid in every PT-model.

  (iv) $\Gamma \vDash_{\mathsf{PT}} \phi$ iff for any PT-model $\mathfrak{M}$, if Γ is valid in $\mathfrak{M}$, then ϕ is valid in $\mathfrak{M}$.

  (v) $\phi \in \mathscr{L}$ is PT-*satisfiable* iff $\neg\phi$ is not PT-valid.

  (vi) $\Gamma \subseteq \mathscr{L}$ is PT-*satisfiable* iff there is a PT-model $\mathfrak{M}$ such that Γ is valid in $\mathfrak{M}$.

(vii) For $\Gamma \subseteq \mathscr{L}$ and $\phi \in \mathscr{L}$, a PT-model $\mathfrak{M}$ *validates* $\langle \Gamma, \phi \rangle$ iff either Γ is not valid in $\mathfrak{M}$ or ϕ is valid in $\mathfrak{M}$.

4.4 Axioms and rules

The axioms of PT are all truth-functional tautologies plus all instances of the following:

A0. $\phi \mathbin{\square\!\!\rightarrow} \phi$,

A1. $\big((\phi \mathbin{\square\!\!\rightarrow} \psi) \wedge (\phi \mathbin{\square\!\!\rightarrow} \chi)\big) \rightarrow \big(\phi \mathbin{\square\!\!\rightarrow} (\psi \wedge \chi)\big)$,

A2. $\big(\phi \mathbin{\square\!\!\rightarrow} (\psi \wedge \chi)\big) \rightarrow \big(\phi \mathbin{\square\!\!\rightarrow} \psi\big)$,

A3. $\big((\phi \mathbin{\square\!\!\rightarrow} \psi) \wedge (\phi \mathbin{\square\!\!\rightarrow} \chi)\big) \rightarrow \big((\phi \wedge \psi) \mathbin{\square\!\!\rightarrow} \chi\big)$,

A4. $\big((\phi \mathbin{\square\!\!\rightarrow} \chi) \wedge (\psi \mathbin{\square\!\!\rightarrow} \chi)\big) \rightarrow \big((\phi \vee \psi) \mathbin{\square\!\!\rightarrow} \chi\big)$,

A5. $(\phi \wedge \psi) \rightarrow (\phi \mathbin{\square\!\!\rightarrow} \psi)$,

A6. $(\phi \mathbin{\square\!\!\rightarrow} \psi) \rightarrow (\phi \rightarrow \psi)$,

A7. $\square(\chi \leftrightarrow \theta) \rightarrow (\phi \leftrightarrow \psi)$, where ψ differs from ϕ only by replacing some subsentences of ϕ of the form χ by θ,

A8. $\lozenge\phi$, for ϕ a conjunction of atomic sentences,

A9. $\big(\phi \mathbin{\square\!\!\rightarrow} (\psi \mathbin{\square\!\!\rightarrow} \chi)\big) \leftrightarrow \big((\phi \wedge \psi) \mathbin{\square\!\!\rightarrow} \chi\big)$, for ϕ a conjunction of atomic sentences,

A10. $(\phi \mathbin{\square\!\!\rightarrow} \psi) \rightarrow \square(\phi \rightarrow \psi)$, for ψ a positive Boolean combination of atomic sentences,

A11. $\big(\phi \mathbin{\square\!\!\rightarrow} (\psi \vee \chi)\big) \leftrightarrow \big((\phi \mathbin{\square\!\!\rightarrow} \psi) \vee (\phi \mathbin{\square\!\!\rightarrow} \chi)\big)$, for ϕ a conjunction of literals,

A12.
$$\left(\neg\big((\phi \vee \psi) \mathbin{\square\!\!\rightarrow} \phi\big) \wedge \neg\big((\phi \vee \psi) \mathbin{\square\!\!\rightarrow} \psi\big)\right) \rightarrow$$
$$\left(\big((\phi \vee \psi) \mathbin{\square\!\!\rightarrow} \chi\big) \rightarrow \big((\phi \mathbin{\square\!\!\rightarrow} \chi) \wedge (\psi \mathbin{\square\!\!\rightarrow} \chi)\big)\right),$$

for ϕ, ψ conjunctions of literals.

Definition 4.4.1. A PT-*proof* is a finite list of sentences of $\mathscr{L}$ such that for every member χ of the list:

   (i) either χ is a truth-functional tautology,

   (ii) or χ is an instance of one of A0–A12,

   (iii) or χ has the form ψ where ϕ and $\psi \rightarrow \psi$ appear earlier in the list,

   (iv) or χ has the form $\Box\phi$ and ϕ appears earlier on the list.

If ϕ is the last member of a PT-proof, then we say that ϕ is a *theorem of* PT and we write $\vdash_{\mathsf{PT}} \phi$.

   Thus, in addition to the above axioms, the rules of PT are:

Modus Ponens (MP). If $\vdash_{\mathsf{PT}} \phi$ and $\vdash_{\mathsf{PT}} \phi \rightarrow \psi$, then $\vdash_{\mathsf{PT}} \psi$.

Necessitation (NEC). If $\vdash_{\mathsf{PT}} \phi$, then $\vdash_{\mathsf{PT}} \Box\phi$.

Definition 4.4.2.   (i) For $\Gamma \subseteq \mathscr{L}$, $\Gamma \vdash_{\mathsf{PT}} \phi$ iff for some finite $\Gamma' \subseteq \Gamma$, $\vdash_{\mathsf{PT}} \bigwedge \Gamma' \rightarrow \phi$.

   (ii) Γ is PT-*consistent* iff for some $\phi \in \mathscr{L}$, $\Gamma \nvdash_{\mathsf{PT}} \phi$.

   Next, we record two central facts about $\vdash_{\mathsf{PT}}$.

Lemma 4.4.1.   (i) If $\vdash_{\mathsf{PT}} \chi \leftrightarrow \theta$, $\vdash_{\mathsf{PT}} \phi$, and ψ differs from ϕ only by replacing some subsentences of ϕ of the form χ by θ, then $\vdash_{\mathsf{PT}} \psi$.

   (ii) $\vdash_{\mathsf{PT}} \Box\phi \rightarrow \Box\Box\phi$. (This is the characteristic axiom of the modal logic S4.)

Proof. (i): Suppose that $\vdash_{\mathsf{PT}} \chi \leftrightarrow \theta$. Then $\vdash_{\mathsf{PT}} \Box(\chi \leftrightarrow \theta)$, by NEC. Now suppose that $\vdash_{\mathsf{PT}} \phi$ and that ψ differs from ϕ only by replacing some subsentences of ϕ of the form χ by θ. Then $\vdash_{\mathsf{PT}} \psi$ immediately follows by A7 and MP.

   (ii):

     1. $\Box(\phi \leftrightarrow \top) \rightarrow (\Box\Box\phi \leftrightarrow \Box\Box\top)$ A7

     2. $\Box\phi \rightarrow \Box\phi$ taut.

     3. $\Box\phi \rightarrow \Box(\phi \leftrightarrow \top)$ 2, Theorem 4.4.1.(i).

     4. $\Box\Box\top$ taut., NEC

     5. $\Box\phi \rightarrow \Box\Box\phi$ 1, 3, 4, taut., MP

4.5 Discussion

The following series of remarks discuss certain features of PT and compare it to a number of well known conditional and modal logics.

Remark 4.5.1. Note that in 4.3.4.(i), we defined validity in a model as truth at the minimal element of that model. This captures the idea PT is a logic whose theorems are, relative to an interpretation, true at the ideal $\{\mathbf{0}\}$, i.e. the ideal whose sole member is the computable degree.

Remark 4.5.2. Any PT-frame $\langle \mathfrak{W}, \leq \rangle$ can be transformed into a frame with a ternary comparative similarity relation $\leq^*$ where for $w, v, u \in \mathfrak{W}$, $v \leq^*_w u$ iff $w \leq v$ and $v \leq u$. Conversely, call a ternary relation $\mathfrak{R}$ *base-invariant* iff it does not vary as we vary first argument except for adjustments of the domain. Then given a base-invariant ternary frame $\langle \mathfrak{W}, \leq^* \rangle$ such that $\langle \mathfrak{W}, \leq^*_{w_@} \rangle$ is an upper semilattice with a minimal element $w_@$ where for any $w \in \mathfrak{W}$, $\{\langle v, u \rangle : v \leq_{w_@} u\}$ includes $\{\langle v, u \rangle : v \leq_w u\}$, any model on $\langle \mathfrak{W}, \leq^* \rangle$ is in Burgess' model class $\mathcal{M}_1$. The models in $\mathcal{M}_1$ are $\mathcal{M}_1$ are tuples $\langle \mathfrak{W}, \mathfrak{R} \rangle$, subject to the following constraints. Let $\mathfrak{W}^b_w = \{v \in \mathfrak{W} : \mathfrak{R}wvv\}$. Then we require that $\mathfrak{R}$ is a ternary relation on $\mathfrak{W} \neq \emptyset$ such that for all $w \in \mathfrak{W}$, $\{\langle v, u \rangle : \mathfrak{R}wvu\}$ is a partial order and such that for all $w \in \mathfrak{W}$, $w \in \mathfrak{W}^b_w$ and for all $v \in \mathfrak{W}^b_w$, $\mathfrak{R}wwv$. $\square\rightarrow$ is defined such that $[\![\phi \mathbin{\square\rightarrow} \psi]\!]_{\mathfrak{M},w} = 1$ iff for all $v \in \mathfrak{W}^b_w \cap [\![\phi]\!]_{\mathfrak{M}}$, there is some $u \in \mathfrak{W}^b_w \cap [\![\phi]\!]_{\mathfrak{M}}$ such that $\mathfrak{R}wuv$ and such that for any $t \in \mathfrak{W}^b_w$ such that $\mathfrak{R}wtu$, $[\![\phi \rightarrow \psi]\!]_{\mathfrak{M},t} = 1$.

Remark 4.5.3. 4.3.2.(e), the semantic clause for $\square\rightarrow$, is closely related to the clause used in Burgess (1981), Lewis (1981), and Veltman (1985). This clause is a generalization of Lewis' clause in his γ-models Lewis (1971). Lewis' clause is the following:

(e′) $[\![\phi \mathbin{\square\rightarrow} \psi]\!]_{\mathfrak{M},w} = 1$ iff there is some $u \in \mathfrak{W}_w \cap [\![\phi]\!]_{\mathfrak{M}}$ such that for any $t \in \mathfrak{W}_w$ such that $\mathfrak{R}wtu$, $[\![\phi \rightarrow \psi]\!]_{\mathfrak{M},t} = 1$.

We need the initial universal quantifier because the worlds in our frames are not required to be comparable, i.e. we do not require that for all $w, v \in \mathfrak{W}$, $w \leq v$ or $v \leq w$.

Remark 4.5.4. Given Theorem 4.3.3.(iii), we can say that $\phi \mathbin{\square\!\!\rightarrow} \psi$ is true at w is to say that all minimal ϕ-worlds relative to w are ψ-worlds, where a world v is a ϕ-*world* iff $[\![\phi]\!]_{\mathfrak{M},v} = 1$. So even though 4.3.1 does not require every non-empty subset of $\mathfrak{W}$ to have a set of minimal elements and so not all PT-models satisfy what is called the *limit assumption* in Lewis (1973), we can still help ourselves to a quite simple understanding of the semantic clause for the conditional connective that dispenses with the three quantifiers of the official wording of the clause. (In other words, PT-models satisfy the Smoothness Condition of Kraus et al. (1990, p. 182).) Finally, note that if ϕ is a conjunction of atomic sentences, Theorem 4.3.1.(ii) guarantees that there is exactly one minimal ϕ-world relative to any $w \in \mathfrak{W}$.

Remark 4.5.5. Given Theorem 4.5.4, it might be wondered why we did not simply require that PT-frames satisfy the limit assumption. While Theorem 4.5.4 shows that ϕ is valid in every PT-model iff ϕ is valid in every PT-model that satisfies the limit assumption, requiring that all PT-frames satisfy the limit assumption would have the undesirable consequence that the intended frame $\langle \mathbf{I}, \subseteq \rangle$ is not a PT-frame. For as shown in Sacks (1964), the computably enumerable degrees are dense. Nevertheless, Theorem 4.5.4 shows that our language cannot express this fact.

Remark 4.5.6. 4.3.1.(ii).(b) requires that if an atomic sentence is true at a world, it remains true as we move up the partial order induced by $\leq$, much as in Kripke models for intuitionistic logic Kripke (1965). However, since our connective $\neg$ is classical, we do not have the full hereditary condition, which would require for any ϕ that if ϕ is true at a world, it remains true as we move up the partial order.

Remark 4.5.7. Also unlike in Kripke models, we have that each atomic sentence α is true at some world (4.3.1.(ii).(a)). This together with 4.3.1.(ii).(b) gives us $\vDash_{\mathsf{PT}}$ A8, where A8 is the axiom that says that every atomic conjunction of atomic sentences is possibly true. This corresponds to the idea that we can feed any set $S \subseteq \omega$ into the oracle tape of an oracle Turing machine, which would allow that machine to decide S.

Remark 4.5.8. Burgess' system S_1 replaces A0–A6 with seven axioms that one gets by replacing $\phi, \psi,$ and χ in A0–A6 with $p, q,$ and r. Burgess (1981). The rules of S_1 are MP and Theo-

rem 4.4.1.(i), as well as Uniform Substitution (Burgess calls it just 'Substitution'). Burgess needs Uniform Substitution because his axioms are not schemas. We need to formulate our axioms as schemas because we would not be able to formulate A8–A12 otherwise. Thus unlike S_1, PT is not closed under Uniform Substitution. However, since we do not place any restrictions on A0–A6, we still have that if ϕ is a theorem of S_1, then $\vdash_{\mathsf{PT}} \phi$.

Remark 4.5.9. Burgess shows that the theorems of S_1 are all and only the sentences valid in every model in his model class $\mathcal{M}_1$. Since we saw in Remark 5 that every PT-model can be transformed into a model in $\mathcal{M}_1$ and A0–A6 are valid in every model in $\mathcal{M}_1$, we immediately have that $\vDash_{\mathsf{PT}}$ A0 $\wedge$ A1 $\wedge$ A3 $\wedge$ A4 $\wedge$ A5 $\wedge$ A6.

Remark 4.5.10. A7 is well known from modal logic, but it does not generally hold in conditional logics whose models contain a ternary comparative similarity relation such as those studied in Burgess (1981), Lewis (1981), and Veltman (1985). To see that, recall that in Theorem 4.4.1.(ii) we showed how to derive $\Box\phi \rightarrow \Box\Box\phi$ from A7, where other than A7 we only used axioms and rules found in Burgess' S_1. But $\Box\phi \rightarrow \Box\Box\phi$ will not hold in general if $\mathfrak{R}$ is a ternary comparative similarity relation that is not base-invariant. For suppose that $\Box\phi$ is true at w. Then ϕ holds at every $v \in \mathfrak{W}_w^b$ (see Theorem 4.5.2 for the definition of $\mathfrak{W}_w^b$). But that is compatible with there being some $v \in \mathfrak{W}_w^b$ such that there is some $u \in \mathfrak{W}_v^b$ where ϕ is false, in which case $\Box\phi$ is false at v, and so $\Box\Box\phi$ is false at w.

Remark 4.5.11. A9 is a restricted version of the *import-export principle*. In its unrestricted form, it makes a conditional collapse into the material conditional in the presence of MP; see Gibbard (1980), McGee (1985), Kratzer (1986), Fitelson (2013), and Khoo (2013) for discussion. Note furthermore that A10 entails a restricted version of *antecedent strengthening*: for ϕ, ψ, χ positive Boolean combinations of atomic sentences, $\vdash_{\mathsf{PT}} (\phi \boxbox\rightarrow \psi) \rightarrow ((\phi \wedge \chi) \boxbox\rightarrow \psi)$. In full generality, antecedent strengthening is invalid for the counterfactual conditional, as discussed in p. 10]lewis1973counterfactuals. A11 is a weak version of axiom (a5) of Stalnaker's logic Stalnaker (1968). Against the background of Stalnaker's other axioms (all of which theorems of PT), (a5) is equivalent to $(\phi \boxbox\rightarrow \psi) \vee (\phi \boxbox\rightarrow \neg\psi)$, which is the principle of conditional excluded middle

discussed in Stalnaker (1980). Its full version is not PT-valid. Finally, A12 is a restricted version of Lewis' p. 80]lewis1971completeness Axiom C. Its unrestricted version is invalid in PT as well as in Burgess' models because they allow for incomparable worlds.

Remark 4.5.12. We observe the following facts about $\models_{\mathsf{PT}}$:

VACUITY. $\models_{\mathsf{PT}} (\neg\phi \;\square\!\!\rightarrow\; \phi) \rightarrow (\psi \;\square\!\!\rightarrow\; \phi)$

CLOSURE. If $\models_{\mathsf{PT}} (\phi_1 \wedge \cdots \wedge \phi_n) \rightarrow \psi$, then $\models_{\mathsf{PT}} ((\chi \;\square\!\!\rightarrow\; \phi_1) \wedge \cdots \wedge (\chi \;\square\!\!\rightarrow\; \phi_n)) \rightarrow (\chi \;\square\!\!\rightarrow\; \psi)$.

EQUIVALENCE. If $\models_{\mathsf{PT}} \phi \leftrightarrow \psi$, then $\models_{\mathsf{PT}} (\phi \;\square\!\!\rightarrow\; \chi) \leftrightarrow (\psi \;\square\!\!\rightarrow\; \chi)$.

Together with A0 and MP, VACUITY, CLOSURE, and EQUIVALENCE constitute the minimal conditional logic discussed in Williamson (2010). Note also if we adopted CLOSURE and EQUIVALENCE as rules, we could dispense with A2, A7, and NEC.

Remark 4.5.13. The normal modal logic axiom K: $\square(\phi \rightarrow \psi) \rightarrow (\square\phi \rightarrow \square\psi)$ and the reflexivity axiom T: $\square\phi \rightarrow \phi$ are theorems of S_1 and thus we have $\vdash_{\mathsf{PT}}$ K $\wedge$ T. It thus follows from Theorem 4.4.1.(ii) that every theorem of the normal modal logic $\mathsf{S4}$ is a theorem of PT. (The models of $\mathsf{S4}$ have a reflexive and transitive accessibility relation. See Cresswell and Hughes (1996) for $\mathsf{S4}$ and the modal logics mentioned below.) As such, PT also contains the conditional logic CT4 studied in Boutilier (1990), which is equivalent to $\mathsf{S4}$. This contrasts with Burgess' system, some of whose models do not validate the characteristic axiom of $\mathsf{S4}$.

Remark 4.5.14. what is more, because PT-frames are upper semilattices, they are contained in the class of frames of the modal logic $\mathsf{S4.2}$, whose accessibility relations $\mathfrak{R}$ are not only reflexive and transitive, but also "convergent," which means that for any three worlds w, v, u, such that $\mathfrak{R}wv$ and $\mathfrak{R}wu$, there is some world t such that $\mathfrak{R}vt$ and $\mathfrak{R}vt$. The characteristic axiom of $\mathsf{S4.2}$ is M: $\Diamond\square\phi \rightarrow \square\Diamond\phi$, and so we have $\models_{\mathsf{PT}}$ M.

Remark 4.5.15. Because not all of our worlds are comparable, PT does *not* contain the modal logic $\mathsf{S4.3}$, whose accessibility relations $\mathfrak{R}$ are not only reflexive and transitive, but also "connected," which means that for any three worlds w, v, u, such that $\mathfrak{R}wv$ and $\mathfrak{R}wu$, either $\mathfrak{R}vu$ or $\mathfrak{R}uv$. The

characteristic axiom of S4.3 is D1: $\Box(\Box\phi \to \psi) \vee \Box(\Box\psi \to \phi)$. We thus have $\nvDash_{\mathsf{PT}}$ D1. As such, PT also does not contain the conditional logic CT4D studied in Boutilier (1990), which is equivalent to S4.3.

Remark 4.5.16. For ϕ a positive Boolean combination of atomic sentences, we have $\vDash_{\mathsf{PT}} (\phi \mathbin{\Box\!\!\!\to} \psi) \to \neg(\phi \mathbin{\Box\!\!\!\to} \neg\psi)$. In its unrestricted form, this is known as *Boethius' thesis* in the literature on connexive logics; see McCall (1966). In this unrestricted form, we only have $\vDash_{\mathsf{PT}} \Diamond\phi \to \big((\phi \mathbin{\Box\!\!\!\to} \psi) \to \neg(\phi \mathbin{\Box\!\!\!\to} \neg\psi)\big)$, which is axiom (a4) of Stalnaker's logic Stalnaker (1968).

4.6 Soundness

Theorem 4.6.1. If $\vdash_{\mathsf{PT}} \phi$, then $\vDash_{\mathsf{PT}} \phi$.

Proof. Recall Theorem 4.5.4 throughout this proof.

That A0–A6 are all PT-valid was already observed in Theorem 4.5.9.

The fact that $\leq$ is transitive and base-invariant (or rather that the ternary relation $\leq^*$ is base-invariant—see Theorem 4.5.2) guarantees that A7 is PT-valid. (See Theorem 4.5.10 for further discussion).

4.3.1.(ii).(a) guarantees that A8 is PT-valid.

A9 is PT-valid. ($\Rightarrow$). Suppose there is some conjunction of atomic sentences ϕ and some PT-model $\mathfrak{M}$ such that $[\![\phi \mathbin{\Box\!\!\!\to} (\psi \mathbin{\Box\!\!\!\to} \chi)]\!]_{\mathfrak{M}, w_@} = 1$ but $[\![(\phi \wedge \psi) \mathbin{\Box\!\!\!\to} \chi]\!]_{\mathfrak{M}, w_@} = 0$. The latter tells us that the set S_1 of $\phi \wedge \psi$-worlds that are minimal relative to $w_@$ contains a $\neg\chi$-world. The former tells us that the ϕ-world w that is minimal relative to $w_@$ is a $\psi \mathbin{\Box\!\!\!\to} \chi$-world. So the set S_2 of ψ-worlds that are minimal relative to w contains only χ-worlds. By 4.3.1.(ii).(b), S_2 contains only ϕ-worlds. Also, 4.3.1.(i) guarantees that $S_1 = S_2$. Contradiction. ($\Leftarrow$). Suppose there is some conjunction of atomic sentences ϕ and some PT-model $\mathfrak{M}$ such that $[\![(\phi \wedge \psi) \mathbin{\Box\!\!\!\to} \chi]\!]_{\mathfrak{M}, w_@} = 1$ but $[\![\phi \mathbin{\Box\!\!\!\to} (\psi \mathbin{\Box\!\!\!\to} \chi)]\!]_{\mathfrak{M}, w_@} = 0$. The latter tells us that the ϕ-world w that is minimal relative to $w_@$ is a $\neg(\psi \mathbin{\Box\!\!\!\to} \chi)$-world. So the set S_1 of ψ-worlds that are minimal relative to w contains a $\neg\chi$-world. By 4.3.1.(ii).(b), S_1 contains only ϕ-worlds. $[\![(\phi \wedge \psi) \mathbin{\Box\!\!\!\to} \chi]\!]_{\mathfrak{M}, w_@} = 1$ tells us that the set S_2 of $\phi \wedge \psi$-worlds that are minimal relative to $w_@$ contains only χ-worlds. But 4.3.1.(i) guarantees that

$S_1 = S_2$. Contradiction.

4.3.1.(ii).(b) guarantees that A10 is PT-valid.

As discussed in Theorem 4.5.11, A11 is a weak version of axiom (a5) of Stalnaker's logic Stalnaker (1968). 4.3.1.(i) and Theorem 4.3.3.(iv) guaranteee that it is PT-valid.

A12 is PT-valid. For suppose there are conjunctions of literals ϕ and ψ for which there is a PT-model $\mathfrak{M}$ such that $[\![\neg((\phi \lor \psi) \,\square\!\!\rightarrow \phi) \land \neg((\phi \lor \psi) \,\square\!\!\rightarrow \psi)]\!]_{\mathfrak{M},w_@} = 1$ but $[\![((\phi \lor \psi) \,\square\!\!\rightarrow \chi) \rightarrow ((\phi \,\square\!\!\rightarrow \chi) \land (\psi \,\square\!\!\rightarrow \chi))]\!]_{\mathfrak{M},w_@} = 0$. By Theorem 4.3.3.(iv), $[\![\neg((\phi \lor \psi) \,\square\!\!\rightarrow \phi) \land \neg((\phi \lor \psi) \,\square\!\!\rightarrow \psi)]\!]_{\mathfrak{M},w_@} = 1$ tells us that the minimal ϕ-world w relative to $w_@$ and the minimal ψ-world v relative to $w_@$ are such that $w \not\leq v$ and $v \not\leq w$. At the same time, $[\![((\phi \lor \psi) \,\square\!\!\rightarrow \chi) \rightarrow ((\phi \,\square\!\!\rightarrow \chi) \land (\psi \,\square\!\!\rightarrow \chi))]\!]_{\mathfrak{M},w_@} = 0$ tells us that $[\![(\phi \lor \psi) \,\square\!\!\rightarrow \chi]\!]_{\mathfrak{M},w_@} = 1$ and either $[\![\phi \,\square\!\!\rightarrow \chi]\!]_{\mathfrak{M},w_@} = 0$ or $[\![\psi \,\square\!\!\rightarrow \chi]\!]_{\mathfrak{M},w_@} = 0$. But that can only be if $w \leq v$ or $v \leq w$.

MP preserves PT-validity for the same reasons it preserves validity in all of Burgess' models. To see that NEC preserves PT-validity it is sufficient to note that the above proves regarding A0–A12 would go through unchanged for an arbitrary world instead of for $w_@$. $\boxtimes$

Corollary 4.6.2.   (i) If $\Gamma \vdash_{\mathsf{PT}} \phi$, then $\Gamma \vDash_{\mathsf{PT}} \phi$.

(ii) If $\Gamma \subseteq \mathscr{L}$ is PT-satisfiable, then Γ is PT-consistent.

Proof. Immediate from 4.3.4, 4.4.2, and Theorem 4.6.1. $\boxtimes$

4.7   Some facts about $\vdash_{\mathsf{PT}}$

Before proving completeness, we need to prove some facts about $\vdash_{\mathsf{PT}}$.

Lemma 4.7.1.   (i) For ϕ a positive Boolean combination of atomic sentences, $\vdash_{\mathsf{PT}} \Diamond\phi$.

(ii) For ϕ, ψ conjunctions of atomic sentences, $\vdash_{\mathsf{PT}} (\phi \,\square\!\!\rightarrow \psi) \rightarrow \square((\phi \land \neg\psi) \leftrightarrow \bot)$.

(iii) For ϕ a conjunction of literals, $\vdash_{\mathsf{PT}} \Diamond\phi \rightarrow ((\phi \,\square\!\!\rightarrow \neg\psi) \leftrightarrow \neg(\phi \,\square\!\!\rightarrow \psi))$.

(iv) For ϕ, ψ conjunctions of atomic sentences, $\vdash_{\mathsf{PT}} (\psi \,\square\!\!\rightarrow \neg\phi) \rightarrow \neg((\phi \lor \psi) \,\square\!\!\rightarrow \phi)$.

130

(v) For ϕ a conjunction of literals and for $\Gamma \subseteq \mathscr{L}$ a set such that if ψ is a conjunction of all positive conjuncts of ϕ and $\neg\alpha$ a conjunct of ϕ, $\psi \boxdot\!\!\to \neg\alpha \in \Gamma$, $\Gamma \vdash_{\mathsf{PT}} \Diamond\phi$.

(vi) For ϕ a conjunction of atomic sentences,

$$\vdash_{\mathsf{PT}} \Diamond\phi \to \left((\phi \boxdot\!\!\to \neg(\psi \boxdot\!\!\to \chi)) \leftrightarrow \neg((\phi \wedge \psi) \boxdot\!\!\to \chi) \right).$$

(vii) For A, B finite sets of atomic sentences, $\vdash_{\mathsf{PT}} \Diamond(\bigwedge A \wedge \neg \bigvee B) \to (\bigwedge A \boxdot\!\!\to \neg \bigvee B)$.

(viii) $\vdash_{\mathsf{PT}} \left(\theta \boxdot\!\!\to \chi \right) \to \left(\left((\phi \vee (\theta \wedge \chi)) \boxdot\!\!\to \psi \right) \leftrightarrow \left(((\phi \vee \theta) \boxdot\!\!\to \chi) \right) \right).$

(ix) For A, B finite sets of atomic sentences,

$$\vdash_{\mathsf{PT}} \Diamond\left(\bigwedge A \wedge \neg \bigvee B \right) \to \left((\phi \vee (\bigwedge A \wedge \neg \bigvee B) \boxdot\!\!\to \psi) \leftrightarrow ((\phi \vee \bigwedge A) \boxdot\!\!\to \psi) \right).$$

Proof. In what follows, we occasionally invoke theorems of S_1 and $\mathsf{S}4$, which usually can be verified with simple model-theoretic arguments, and which, by Theorem 4.5.8 and Theorem 4.5.13, are theorems of PT. We also often leave appeals to truth-functional tautologies, MP, and Theorem 4.4.1.(i) (i.e. the replacement of provable equivalents) implicit.

(i): Note first that there is a sentence $\psi_1 \vee \cdots \vee \psi_n$, for ψ_i conjunctions of atomic sentences, that is truth-functionally equivalent to ϕ. Now, take an arbitrary ψ_i, $1 \leq i \leq n$.

1. $\psi_i \to \phi$ taut.
2. $\Diamond\psi_i \to \Diamond\phi$ 1, $\mathsf{S}4$
3. $\Diamond\psi_i$ A8
4. $\Diamond\phi$ 2, 3

(ii): Since $\left((\phi \wedge \neg\psi) \leftrightarrow \bot \right) \leftrightarrow \left(\phi \to \psi \right)$ is a truth-functional tautology, this follows from A10 by Theorem 4.4.1.(i).

(iii):

131

1. $\big((\phi \boxarrow \neg\psi) \wedge (\phi \boxarrow \psi)\big) \to \big(\phi \boxarrow (\psi \wedge \neg\psi)\big)$ A1

2. $\big((\phi \boxarrow \neg\psi) \wedge (\phi \boxarrow \psi)\big) \to \big(\phi \boxarrow \bot\big)$ 1, Theorem 4.4.1.(i)

3. $\big((\phi \boxarrow \neg\psi) \wedge (\phi \boxarrow \psi)\big) \to \neg\Diamond\phi$ 2, def. $\Diamond$, Theorem 4.4.1.(i)

4. $\Diamond\phi \to \big((\phi \boxarrow \neg\psi) \to \neg(\phi \boxarrow \psi)\big)$ 3

5. $\phi \boxarrow \phi$ A1

6. $\phi \leftrightarrow \big(\phi \leftrightarrow (\psi \vee \neg\psi)\big)$ taut.

7. $\phi \boxarrow (\psi \vee \neg\psi)$ 5, 6

8. $(\phi \boxarrow \psi) \vee (\phi \boxarrow \neg\psi)$ 7, A11

9. $\neg(\phi \boxarrow \psi) \to (\phi \boxarrow \neg\psi)$ 8

10. $\Diamond\phi \to \big(\neg(\phi \boxarrow \psi) \to (\phi \boxarrow \neg\psi)\big)$ 9

11. $\Diamond\phi \to \big((\phi \boxarrow \neg\psi) \leftrightarrow \neg(\phi \boxarrow \psi)\big)$ 4, 11

(iv):

1. $\big((\psi \boxarrow \neg\phi) \wedge \neg(\psi \boxarrow \bot)\big) \to \neg\big((\neg\phi \wedge \psi) \boxarrow \bot\big)$ S$_1$

2. $\neg(\psi \boxarrow \bot)$ Theorem 4.7.1.(i)

3. $\big((\phi \vee \psi) \boxarrow \phi\big) \to \big(\Box((\phi \vee \psi) \to \phi)\big)$ A10

4. $\big((\phi \vee \psi) \boxarrow \phi\big) \to \big(\neg((\phi \vee \psi) \to \phi) \boxarrow \bot\big)$ 4, def. $\Box$

5. $\big((\phi \vee \psi) \boxarrow \phi\big) \to \big((\neg\phi \wedge \psi) \boxarrow \bot\big)$ 4, def. $\Box$

6. $\big(\psi \boxarrow \neg\phi\big) \to \neg\big((\phi \vee \psi) \boxarrow \phi\big)$ 1, 2, 5

(v): Without loss of generality (and again temporarily assuming that q_i, $i \in \omega$, are atomic sentences), let $\psi = p_1 \wedge \cdots \wedge p_n$, $\chi = \neg q_1 \wedge \cdots \wedge \neg q_m$, and $\phi = \psi \wedge \chi$ and let $\Gamma = \{(p_1 \wedge \cdots \wedge p_n) \boxarrow \neg q_1, \ldots, (p_1 \wedge \cdots \wedge p_n) \boxarrow \neg q_m\}$.

1. $\bigwedge \Gamma \to (\psi \boxarrow \chi)$ A1

2. $\big((\psi \boxarrow \chi) \wedge ((\psi \wedge \chi) \boxarrow \bot)\big) \to \big(\psi \boxarrow \bot\big)$ S$_1$

3. $\big(\bigwedge \Gamma \wedge ((\psi \wedge \chi) \boxarrow \bot)\big) \to \big(\psi \boxarrow \bot\big)$ 1, 2

4. $\big(\bigwedge \Gamma \wedge ((\psi \wedge \chi) \boxarrow \bot)\big) \to \neg\Diamond\psi$ 3, def. $\Diamond$

5. $\bigwedge \Gamma \to \neg\big((\psi \wedge \chi) \boxarrow \bot\big)$ 4, A8

6. $\bigwedge \Gamma \to \Diamond\phi$ 5, def. $\Diamond$, def. ϕ

(vi):

1. $\Diamond\phi \to \Big(\big(\phi \mathbin{\square\!\!\rightarrow} \neg(\psi \mathbin{\square\!\!\rightarrow} \chi)\big) \to \neg\big(\phi \mathbin{\square\!\!\rightarrow} (\psi \mathbin{\square\!\!\rightarrow} \chi)\big)\Big)$ Theorem 4.7.1.(iii)

2. $\Diamond\phi \to \Big(\big(\phi \mathbin{\square\!\!\rightarrow} \neg(\psi \mathbin{\square\!\!\rightarrow} \chi)\big) \to \neg\big((\phi \wedge \psi) \mathbin{\square\!\!\rightarrow} \chi\big)\Big)$ 1, A9

3. $\Diamond\phi \to \Big(\neg\big((\phi \wedge \psi) \mathbin{\square\!\!\rightarrow} \chi\big) \to \neg\big(\phi \mathbin{\square\!\!\rightarrow} (\psi \mathbin{\square\!\!\rightarrow} \chi)\big)\Big)$ A9

4. $\Diamond\phi \to \Big(\neg\big((\phi \wedge \psi) \mathbin{\square\!\!\rightarrow} \chi\big) \to \big(\phi \mathbin{\square\!\!\rightarrow} \neg(\psi \mathbin{\square\!\!\rightarrow} \chi)\big)\Big)$ 3, Theorem 4.7.1.(iii)

5. $\Diamond\phi \to \Big(\big(\phi \mathbin{\square\!\!\rightarrow} \neg(\psi \mathbin{\square\!\!\rightarrow} \chi)\big) \leftrightarrow \neg\big((\phi \wedge \psi) \mathbin{\square\!\!\rightarrow} \chi\big)\Big)$ 2, 4

(vii):

1. $(\bigwedge A \mathbin{\square\!\!\rightarrow} \bigvee B) \to \square(\bigwedge A \to \bigvee B)$ A10

2. $(\bigwedge A \mathbin{\square\!\!\rightarrow} \bigvee B) \to \big(\neg(\bigwedge A \to \bigvee B) \mathbin{\square\!\!\rightarrow} \bot\big)$ 1, def. $\square$

3. $(\bigwedge A \mathbin{\square\!\!\rightarrow} \bigvee B) \to \big((\bigwedge A \wedge \neg\bigvee B) \mathbin{\square\!\!\rightarrow} \bot\big)$ 2

4. $\Diamond(\bigwedge A \wedge \neg\bigvee B) \to \neg(\bigwedge A \mathbin{\square\!\!\rightarrow} \bigvee B)$ 3, def. $\Diamond$

5. $\Diamond(\bigwedge A \wedge \neg\bigvee B) \to \Diamond\bigwedge A$ S4

6. $\Diamond(\bigwedge A \wedge \neg\bigvee B) \to (\bigwedge A \mathbin{\square\!\!\rightarrow} \neg\bigvee B)$ 4, 5, Theorem 4.7.1.(iii)

(viii): We give a model-theoretic argument to show that $\Big(\theta \mathbin{\square\!\!\rightarrow} \chi\Big) \to \Big(\big((\phi \vee (\theta \wedge \chi)) \mathbin{\square\!\!\rightarrow} \psi\big) \leftrightarrow \big((\phi \vee \theta) \mathbin{\square\!\!\rightarrow} \chi\big)\Big)$ is a theorem of Burgess' S_1, from which the result immediately follows.

Take an arbitrary world w from an arbitrary model in Burgess' model class $\mathcal{M}_1$ on the frame $\langle \mathfrak{W}, \mathfrak{R}\rangle$. Suppose that $\theta \mathbin{\square\!\!\rightarrow} \chi$ is true at w.

($\Rightarrow$). Suppose that $\big(\phi \vee (\theta \wedge \chi)\big) \mathbin{\square\!\!\rightarrow} \psi$ is true at w. Let v be a $\phi \vee \theta$-world such that $\mathfrak{R}wwv$.

> Case 1. There is a θ-world u such that $\mathfrak{R}wuv$. Then there is a θ-world t such that $\mathfrak{R}wtu$ and such that χ is true at every θ-world s such that $\mathfrak{R}wst$. t is a $\theta \wedge \chi$-world, so it is a $\big(\phi \vee (\theta \wedge \chi)\big)$-world. So there is a $\big(\phi \vee (\theta \wedge \chi)\big)$-world s such that $\mathfrak{R}wst$ and such that ψ is true at every $\big(\phi \vee (\theta \wedge \chi)\big)$-world r such that $\mathfrak{R}wrt$. Let r be a $\phi \vee \theta$-world such that $\mathfrak{R}wrt$. If r is a ϕ world, then it is a $\big(\phi \vee (\theta \wedge \chi)\big)$-world and it is such that $\mathfrak{R}wrt$, in which case it is a ψ-world. If r is not a ϕ-world, it is a θ-world and since it is such that $\mathfrak{R}wrt$, it is a $\theta \wedge \chi$-world. So r is a $\big(\phi \vee (\theta \wedge \chi)\big)$-world such that $\mathfrak{R}wrt$. So ψ is true at every $\phi \vee \theta$-world q such that $\mathfrak{R}wqs$.

133

Case 2. There is not any θ-world u such that $\mathfrak{R}wuv$. Then v is a ϕ world, and so it is a $\big(\phi \vee (\theta \wedge \chi)\big)$-world. So there is a $\big(\phi \vee (\theta \wedge \chi)\big)$-world u such that $\mathfrak{R}wuv$ such that ψ is true at every $\big(\phi \vee (\theta \wedge \chi)\big)$-world t such that $\mathfrak{R}wtu$. θ is not true at u, and so $\theta \wedge \chi$ is not true at u. So ϕ is true at u. So u is a $\big(\phi \vee (\theta \wedge \chi)\big)$-world. So there is a $\big(\phi \vee (\theta \wedge \chi)\big)$-world t such that $\mathfrak{R}wtu$ such that ψ is true at every $\big(\phi \vee (\theta \wedge \chi)\big)$-world s such that $\mathfrak{R}wst$. Let s be a $\phi \vee \theta$-world such that $\mathfrak{R}wst$. Since $\mathfrak{R}wsv$, θ is true at s. So ϕ is true at s. So s is a $\big(\phi \vee (\theta \wedge \chi)\big)$-world such that $\mathfrak{R}wst$. So ψ is true at t. So ψ is true at every $\phi \vee \theta$-world r such that $\mathfrak{R}wrt$. So $(\phi \vee \theta) \mathrel{\square\!\!\rightarrow} \psi$ is true at w.

($\Leftarrow$). Suppose that $(\phi \vee \theta) \mathrel{\square\!\!\rightarrow} \psi$ is true at w. Let v be a $\big(\phi \vee (\theta \wedge \chi)\big)$-world such that $\mathfrak{R}wwv$. Then v is a world at which $\phi \vee \theta$ is true. So there is a world u such that $\mathfrak{R}wuv$ at which $\phi \vee \theta$ is true such that $(\phi \vee \theta) \rightarrow \psi$ is true at every world t such that $\mathfrak{R}wtu$. We want to show that there is a world t such that $\mathfrak{R}wtu$ and such that $\big(\phi \vee (\theta \wedge \chi)\big)$ is true at t and such that $\big(\phi \vee (\theta \wedge \chi)\big) \rightarrow \psi$ is true at every world s such that $\mathfrak{R}wst$. If $\phi \vee (\theta \wedge \chi)$ is true at u, then we are done: just take t to be u. So the only case we need to worry about is the one in which ϕ and $\theta \wedge \chi$ are both false at u but θ is true at u. Then there is a θ-world t such that $\mathfrak{R}wtu$ and such that χ is true at every θ-world s such that $\mathfrak{R}wst$. So t is a $\theta \wedge \chi$-world such that $\mathfrak{R}wtu$ and such that $\big(\phi \vee (\theta \wedge \chi)\big) \rightarrow \psi$ is true at every world s such that $\mathfrak{R}wst$.

(ix): Follows immediately from Theorem 4.7.1.(vii)&(viii). $\qquad\boxtimes$

4.8 Completeness and decidability

Theorem 4.8.1. If $\vDash_{\mathsf{PT}} \phi$, then $\vdash_{\mathsf{PT}} \phi$.

Proof. We first need a few definitions (recall 4.2.2):

Definition 4.8.1.　(i) An *atomic conditional* is a sentence of the form $\bigwedge A \mathrel{\square\!\!\rightarrow} \alpha$, for A a (possibly empty) set of atomic sentences (where $\bigwedge A = \top$ if $A = \emptyset$), and for α an atomic sentence.

(ii) An *atomic theory over a finite set* $\Delta \subseteq \mathscr{L}$ is a set $\Gamma \subseteq \{\psi \in \mathscr{L}_\Delta : \psi$ is an atomic conditional$\}$ subject to the following two constraints:

134

- for $A \subseteq \mathscr{L}_\Delta$ a set of atomic sentences and $\alpha \in A$, $\bigwedge A \boxright \alpha \in \Gamma$.

- for $A, B \subseteq \mathscr{L}_\Delta$ sets of atomic sentences, if $\bigwedge B \boxright \alpha \in \Gamma$ for every $\alpha \in A$ and $\bigwedge A \boxright \beta$, then $\bigwedge B \boxright \beta \in \Gamma$.

(iii) Where Γ is an atomic theory over Δ, *the complete theory of Γ over Δ* is the set $\Gamma^+ = \Gamma \cup \{\bigwedge A \boxright \neg\alpha : \bigwedge A \boxright \alpha \in \mathscr{L}_\Delta$ is an atomic conditional and $\bigwedge A \boxright \alpha \notin \Gamma\}$.

(iv) A *simple conditional* is a sentence of the form $\phi \boxright \psi$ such that ϕ and ψ are conditional-free.

Armed with these definitions, we now show that given a complete theory, we can transform any sentence χ into a sentence θ that is a Boolean combination of atomic conditionals and conditional-free sentences such that the complete theory proves that χ and θ are equivalent. This is part (iv) of the following lemma. Parts (i)–(iii) lead up to this result.

Lemma 4.8.2. Let $\Delta \subseteq \mathscr{L}$, let Γ be an atomic theory over Δ, and let Γ^+ be the complete theory of Γ over Δ.

(i) For any conditional $\phi \boxright \psi \in \mathscr{L}_\Delta$, we can find some $\theta \in \mathscr{L}_\Delta$ that is a Boolean combination of conditionals $\chi_i \boxright \psi$, for χ_i conjunctions of atomic sentences, such that $\Gamma^+ \vdash_{\mathsf{PT}} (\phi \boxright \psi) \leftrightarrow \theta$.

(ii) For any $\chi \in \mathscr{L}_\Delta$, we can find some $\theta \in \mathscr{L}_\Delta$ that is a Boolean combination of simple conditionals and conditional-free sentences such that $\Gamma^+ \vdash_{\mathsf{PT}} \chi \leftrightarrow \theta$.

(iii) For any simple conditional $\phi \boxright \psi \in \mathscr{L}_\Delta$ where ϕ a conjunction of atomic sentences, we can find some $\theta \in \mathscr{L}_\Delta$ that is a Boolean combination of atomic conditionals and conditional-free sentences such that $\Gamma^+ \vdash_{\mathsf{PT}} (\phi \boxright \psi) \leftrightarrow \theta$.

(iv) For any $\chi \in \mathscr{L}_\Delta$, we can find some $\theta \in \mathscr{L}_\Delta$ that is a Boolean combination of atomic conditionals and conditional-free sentences such that $\Gamma^+ \vdash_{\mathsf{PT}} \chi \leftrightarrow \theta$.

Proof. (i). We describe an algorithm for converting $\phi \boxright \psi$ into a sentence θ that is a conjunction of conditionals $\chi_i \boxright \psi$, for χ_i conjunctions of atomic sentences, such that θ is provably PT-equivalent to $\phi \boxright \psi$ over Γ^+. Put ϕ into disjunctive normal form using the atomic sentences that

appear in $\phi \mathrel{\Box\!\!\rightarrow} \psi$. The resulting sentence has the form $(\rho_1 \vee \cdots \vee \rho_n) \mathrel{\Box\!\!\rightarrow} \psi$, for ρ_i conjunctions of literals. We now weed out some of the ρ_i as follows. First, we select those ρ_i that have a negative conjunct $\neg\alpha_l$ and conjuncts are exactly $\alpha_1, \ldots, \alpha_m$ such that $\bigwedge\{\alpha_1, \ldots, \alpha_m\} \mathrel{\Box\!\!\rightarrow} \alpha_l \in \Gamma^+$ or $\top \mathrel{\Box\!\!\rightarrow} \alpha_l \in \Gamma^+$. These state descriptions are necessarily equivalent to $\bot$ over Γ^+, given Theorem 4.7.1.(ii), so we remove them from $\rho_1 \vee \cdots \vee \rho_n$ using A7. If this removes every ρ_i, then $\phi \mathrel{\Box\!\!\rightarrow} \psi$ is equivalent to $\top$ and so we replace the former with the latter. If there are ρ_i remaining, we have $\Gamma^+ \vdash_{\mathsf{PT}} \Diamond\rho_i$ for each of them, by Theorem 4.7.1.(v). We then use Theorem 4.7.1.(ix) to remove all negative conjuncts in each ρ_i. This might leave us with some ρ_i and ρ_j such that every conjunct of ρ_i appears in ρ_j. Since in this case we have that $\rho_i \leftrightarrow (\rho_i \vee \rho_j)$ is a truth-functional tautology, we remove ρ_j, by Theorem 4.4.1.(i). What remains is a disjunction $\rho_1 \vee \cdots \vee \rho_m$, for ρ_i conjunctions of atomic sentences. The resulting sentence has the form $(\rho_1 \vee \cdots \vee \rho_l) \mathrel{\Box\!\!\rightarrow} \psi$ and is such that for any ρ_i, ρ_j, $i \neq j$, $\Gamma^+ \vdash_{\mathsf{PT}} \rho_i \mathrel{\Box\!\!\rightarrow} \neg\rho_j$. For if $\Gamma^+ \vdash_{\mathsf{PT}} \rho_i \mathrel{\Box\!\!\rightarrow} \rho_j$, then $\rho_i \mathrel{\Box\!\!\rightarrow} \alpha \in \Gamma^+$, for all conjuncts α of ρ_j, in which case we would have already removed ρ_i or ρ_j. From this it follows that $\Gamma^+ \vdash_{\mathsf{PT}} \neg((\rho_i \vee \rho_j) \mathrel{\Box\!\!\rightarrow} \rho_i) \wedge \neg((\rho_i \vee \rho_j) \mathrel{\Box\!\!\rightarrow} \rho_j)$, thanks to Theorem 4.7.1.(iv). We then use A12 and A4 to obtain a conjunction of all $\rho_i \mathrel{\Box\!\!\rightarrow} \psi$, for the remaining ρ_i.

(ii). We describe an algorithm for converting χ into a sentence θ that is a Boolean combination of simple conditionals and conditional-free sentences such that θ is provably PT-equivalent to χ over Γ^+. Note that a *non-simple conditional* is a sentence of the form $\phi \mathrel{\Box\!\!\rightarrow} \psi$ such that ψ is not conditional-free. Also, we define the *Boolean-complexity* of a sentence χ such that the Boolean-complexity of χ is 0 if χ is atomic or a conditional, and the Boolean-complexity of χ is $n + 1$ if $\chi = \neg\theta$ or $\chi = \theta \wedge \xi$ and the Boolean-complexity of θ and ξ is at most n. Next, we define the $\mathrel{\Box\!\!\rightarrow}$-*complexity* of a sentence χ such that the $\mathrel{\Box\!\!\rightarrow}$-complexity of χ is 0 if χ does not contain any occurrences of $\mathrel{\Box\!\!\rightarrow}$, and the $\mathrel{\Box\!\!\rightarrow}$-complexity of χ is $n+1$ if $\chi = \theta \mathrel{\Box\!\!\rightarrow} \xi$, for ξ a Boolean combination of sentences $\zeta_1, \ldots, \zeta_m$, where the $\mathrel{\Box\!\!\rightarrow}$-complexity of ζ_i is at most n. Finally, using transfinite ordinals, we let the *complexity* of a conditional $\chi = \theta \mathrel{\Box\!\!\rightarrow} \xi$ be $\big(\omega \times \omega \times (\text{the } \mathrel{\Box\!\!\rightarrow}\text{-complexity}$

of $\chi)\big) + \left(\omega \times \begin{cases} 0 & \text{if } \theta = \bigwedge A, \text{ for } A \text{ a (possibly empty) set of atomic sentences} \\ 1 & \text{otherwise} \end{cases}\right) + (\text{the}$

Boolean-complexity of $\chi\big)$.

1. χ either does or does not contain non-simple conditional subsentences.

   - If it does not, halt.

   - If it does, look at the set Λ that contains the most $\Box\!\!\rightarrow$-complex non-simple conditional subsentences of χ. Next, look at the the set $\Omega \subseteq \Lambda$ that contains the sentences with the most Boolean-complex consequents. Arbitrarily select a sentence $\phi \Box\!\!\rightarrow \psi \in \Omega$ and go to step 2.

2. ϕ either is or is not a conjunction of atomic sentences.

   - If it is not, use (i) to turn $\phi \Box\!\!\rightarrow \psi$ into $(\rho_1 \Box\!\!\rightarrow \psi) \wedge \cdots \wedge (\rho_m \Box\!\!\rightarrow \psi)$, for ρ_i conjunctions of atomic sentences. Go to step 1.

   - If it is, then look at ψ. ψ is either of the form $\neg\xi, \xi \wedge \zeta$, or $\xi \Box\!\!\rightarrow \zeta$.

     - If ψ is of the form $\neg\xi$, then ξ is either of the form $\neg\mu, \mu \wedge \nu$, or $\mu \Box\!\!\rightarrow \nu$.

       * If ξ is of the form $\neg\mu$, replace $\neg\neg\mu$ with μ. Go to step 1.

       * If ξ is of the form $\mu \wedge \nu$, use A11 to replace $\phi \Box\!\!\rightarrow \neg(\mu \wedge \nu)$ with $(\phi \Box\!\!\rightarrow \neg\mu) \vee (\phi \Box\!\!\rightarrow \neg\nu)$. Go to step 1.

       * If ξ is of the form $\mu \Box\!\!\rightarrow \nu$, use A8 and Theorem 4.7.1.(vi) to replace $\phi \Box\!\!\rightarrow \neg(\mu \Box\!\!\rightarrow \nu)$ with $\neg\big((\phi \wedge \mu) \Box\!\!\rightarrow \nu\big)$. Go to step 1.

     - If ψ is of the form $\xi \wedge \zeta$, use A1 and A2 to replace $\phi \Box\!\!\rightarrow (\xi \wedge \zeta)$ with $(\phi \Box\!\!\rightarrow \xi) \wedge (\phi \Box\!\!\rightarrow \zeta)$. Go to step 1.

     - If ψ is of the form $\xi \Box\!\!\rightarrow \zeta$, use A9 to replace $\phi \Box\!\!\rightarrow (\xi \Box\!\!\rightarrow \zeta)$ with $(\phi \wedge \xi) \Box\!\!\rightarrow \zeta$. Go to step 1.

This algorithm halts because, at every stage, we either reduce the number of conditional subsentences with maximum complexity, or else we reduce the complexity of the unique maximally complex conditional subsentence.

   (iii). We describe an algorithm for converting a simple conditional $\phi \Box\!\!\rightarrow \psi$, for ϕ a conjunction of atomic sentences, into a Boolean combination of atomic conditionals θ that is provably PT-equivalent to $\phi \Box\!\!\rightarrow \psi$ over Γ^+.

1. If ψ is atomic, halt. If not, go to step 2.

2. ψ is either of the form $\chi \wedge \zeta$ or of the form $\neg \chi$.

   - If ψ is of the form $\chi \wedge \zeta$, use A1, A2, and Theorem 4.4.1.(i) to turn $\phi \mathbin{\square\!\!\rightarrow} \psi$ into $(\phi \mathbin{\square\!\!\rightarrow} \chi) \wedge (\phi \mathbin{\square\!\!\rightarrow} \zeta)$. Go to step 1.

   - If ψ is of the form $\neg \chi$, use A8, Theorem 4.7.1.(iii), and Theorem 4.4.1.(i) to turn $\phi \mathbin{\square\!\!\rightarrow} \psi$ into $\neg (\phi \mathbin{\square\!\!\rightarrow} \chi)$. Go to step 1.

(iv). We can put the above three algorithms together to obtain an algorithm for converting χ into a sentence θ that is a Boolean combination of atomic conditionals and conditional-free sentences such that θ is provably PT-equivalent to χ over Γ^+. First, use the algorithm described in the proof of (ii) to convert χ into a sentence θ_1 that is Boolean combination of simple conditionals and conditional free sentences such that θ_1 is provably PT-equivalent to χ over Γ^+. Then use the algorithm described in the proof of (i) to convert each conditional $\phi_i \mathbin{\square\!\!\rightarrow} \psi_i$ in θ_1 into a a sentence θ_{2_i} that is a conjunction of conditionals $\chi_{i_j} \mathbin{\square\!\!\rightarrow} \psi$, for χ_{i_j} conjunctions of atomic sentences, such that θ_{2_i} is provably PT-equivalent to $\phi_i \mathbin{\square\!\!\rightarrow} \psi_i$ over Γ^+. Finally, use the algorithm described in the proof of (iii) to convert each $\chi_{i_j} \mathbin{\square\!\!\rightarrow} \psi$ in each θ_{2_i} into a sentence $\theta_{3_{i_j}}$ that is a Boolean combination of atomic conditionals and conditional-free sentences such that $\theta_{3_{i_j}}$ is provably PT-equivalent to $\chi_{i_j} \mathbin{\square\!\!\rightarrow} \psi$ over Γ^+. $\boxtimes$

Remark 4.8.3. An atomic theory can intuitively be thought of as a theory that tells us which sets are absolutely algorithmically decidable and which sets are reducible to which. We may either think of an atomic theory as putting constraints on the interpretation functions which tell us what sets of natural numbers are said to be algorithmically decidable according to the atomic sentences of our formal language. Or we may first fix an interpretation function, which then allows us to evaluate (the members of) the atomic theories as true or false.

Lemma 4.8.4. For Γ an atomic theory over a finite $\Delta \subseteq \mathscr{L}$ and for $\psi \in \mathscr{L}_\Delta$, if $\Gamma^+ \vDash_{\mathsf{PT}} \psi$, then $\Gamma^+ \vdash_{\mathsf{PT}} \psi$.

Proof. Suppose $\Gamma^+ \nvdash_{\mathsf{PT}} \psi$. We will construct a PT-model of Γ^+ where ψ is not valid. For $A \subseteq \mathscr{L}_\Delta$ a set of atomic sentences, let $\overline{A} = \{\alpha \in \mathscr{L}_\Delta : \bigwedge A \mathbin{\square\!\!\rightarrow} \alpha \in \Gamma^+\}$. For our set of worlds, we let

$\mathfrak{W} = \{\overline{A} : A \subseteq \mathscr{L}_\Delta$ is a set of atomic sentences$\}$, and we let $w_@ = \overline{\emptyset}$. $\mathfrak{W}$ is partially ordered by the inclusion relation $\subseteq$. In fact, $\langle \mathfrak{W}, \subseteq \rangle$ is an upper semilattice, and thus a PT-frame. To see this, note that the second constraint in 4.8.1.(ii) ensures that for $A, B \subseteq \mathscr{L}_\Delta$ sets of atomic sentences and $\delta \in \mathscr{L}_\Delta$ atomic, if $\bigwedge A \boxarrow \delta \in \Gamma$ or $\bigwedge B \boxarrow \delta \in \Gamma$, then $\bigwedge(A \cup B) \boxarrow \gamma \in \Gamma$. This guarantees that $\overline{A \cup B}$ is the least upper bound of $\overline{A}$ and $\overline{B}$.

To get our model on the PT-frame $\langle \mathfrak{W}, \subseteq \rangle$, we define $\mathfrak{I}$ such that for atomic $\alpha \in \mathscr{L}_\Delta$, $\mathfrak{I}(\alpha) = \{w \in \mathfrak{W} : \alpha \in w\}$. Now, let $\mathfrak{M} = \langle \mathfrak{W}, w_@, \subseteq, \mathfrak{I} \rangle$. To see that $\mathfrak{M}$ is a PT-model, note first that $\mathfrak{W}$ is finite and so we do not need to worry about infinitely descending $\subseteq$-chains. Next, note that the first two clauses of 4.8.1.(ii) guarantee that 4.3.1.(ii).(a) holds and the second clause of 4.8.1.(ii) guarantees that 4.3.1.(ii).(b) holds. what is more, Γ^+ is valid in $\mathfrak{M}$, since the construction guarantees that for any $\bigwedge A \boxarrow \phi \in \Gamma^+$ and any $w \in \mathfrak{W}$, the minimal $\bigwedge A$-world relative to w is a ϕ-world.

Next, note that $\mathfrak{M}$ validates $\langle \Gamma^+, \psi \rangle$ iff $\Gamma^+ \vdash_{\mathsf{PT}} \psi$. (This is the so-called "truth lemma.") Note that every member of Γ^+ is automatically valid in $\mathfrak{M}$, given the way we have constructed $\mathfrak{M}$ from Γ^+. So we need to show that ψ is valid in $\mathfrak{M}$ iff $\Gamma^+ \vdash_{\mathsf{PT}} \psi$. We first show by induction on the Boolean-complexity that if ψ is a Boolean combination of atomic sentences and atomic conditionals, then if ψ is valid in $\mathfrak{M}$, then $\Gamma^+ \vdash_{\mathsf{PT}} \psi$, and if ψ is not valid in $\mathfrak{M}$, then $\Gamma^+ \vdash_{\mathsf{PT}} \neg\psi$. For the base case where ψ has Boolean-complexity 0, ψ is either an atomic sentence or an atomic conditional. Then if ψ is valid in $\mathfrak{M}$, then $\psi \in \Gamma^+$, and so $\Gamma^+ \vdash_{\mathsf{PT}} \psi$. If $\psi = \alpha$, for α atomic and α is not valid in $\mathfrak{M}$, then $\top \boxarrow \neg\alpha \in \Gamma^+$, by the way we have constructed Γ^+, and so $\Gamma^+ \vdash_{\mathsf{PT}} \neg\alpha$, by A6. And if ψ is an atomic conditional $\bigwedge A \boxarrow \alpha$ and $\bigwedge A \boxarrow \alpha$ is not valid in $\mathfrak{M}$, then $\bigwedge A \boxarrow \neg\alpha \in \Gamma^+$, and so $\Gamma^+ \vdash_{\mathsf{PT}} \neg(\bigwedge A \boxarrow \alpha)$, by Theorem 4.7.1.(i)&(iii). For the induction step, suppose that ψ has Boolean-complexity n and that if ψ is valid in $\mathfrak{M}$, then $\Gamma^+ \vdash_{\mathsf{PT}} \psi$, and if ψ is not valid in $\mathfrak{M}$, then $\Gamma^+ \vdash_{\mathsf{PT}} \neg\psi$. Then if ψ has Boolean-complexity $n + 1$, then either $\psi = \neg\chi$ or $\psi = \theta \wedge \xi$, where χ, θ, and ξ have Boolean-complexity of at most n. Then if $\neg\chi$ is valid in $\mathfrak{M}$, then χ is not valid in $\mathfrak{M}$, and so $\Gamma^+ \vdash_{\mathsf{PT}} \neg\chi$, as desired. And if $\neg\chi$ is not valid in $\mathfrak{M}$, then χ is valid in $\mathfrak{M}$, and so $\Gamma^+ \vdash_{\mathsf{PT}} \chi$, and so $\Gamma^+ \vdash_{\mathsf{PT}} \neg\neg\chi$, as desired. And if $\theta \wedge \xi$ is valid in $\mathfrak{M}$, then θ and ξ are both valid in $\mathfrak{M}$, in which case we have $\Gamma^+ \vdash_{\mathsf{PT}} \theta$ and $\Gamma^+ \vdash_{\mathsf{PT}} \xi$, and so $\Gamma^+ \vdash_{\mathsf{PT}} \theta \wedge \xi$, as desired. And, finally, if $\theta \wedge \xi$ is not valid in $\mathfrak{M}$, then θ or ξ is not valid in $\mathfrak{M}$, in which case we have either

$\Gamma^+ \vdash_{\mathsf{PT}} \neg\theta$ or $\Gamma^+ \vdash_{\mathsf{PT}} \neg\xi$, and so either way we have $\Gamma^+ \vdash_{\mathsf{PT}} \neg(\theta \wedge \xi)$, as desired.

We have shown that if ψ is a Boolean combination of atomic sentences and atomic conditionals, then if ψ is valid in $\mathfrak{M}$, then $\Gamma^+ \vdash_{\mathsf{PT}} \psi$, and if ψ is not valid in $\mathfrak{M}$, then $\Gamma^+ \vdash_{\mathsf{PT}} \neg\psi$, and so $\Gamma^+ \nvdash_{\mathsf{PT}} \psi$. So if ψ is a Boolean combination of atomic sentences and atomic conditionals, then ψ is valid in $\mathfrak{M}$ iff $\Gamma^+ \vdash_{\mathsf{PT}} \psi$. Now let ψ be an arbitrary sentence. Then by Theorem 4.8.2.(iv), we can convert ψ into a sentence χ that is Boolean combination of atomic conditionals and conditional-free sentences such that χ is provably PT-equivalent to ψ over Γ^+. By soundness (Theorem 4.6.1), $\mathfrak{M}$ validates ψ iff $\mathfrak{M}$ validates χ. Also, we just saw that $\mathfrak{M}$ validates χ iff $\Gamma^+ \vdash_{\mathsf{PT}} \chi$. And since χ is provably PT-equivalent to ψ over Γ^+, $\Gamma^+ \vdash_{\mathsf{PT}} \chi$ iff $\Gamma^+ \vdash_{\mathsf{PT}} \psi$. So, $\mathfrak{M}$ validates ψ iff $\Gamma^+ \vdash_{\mathsf{PT}} \psi$. $\boxtimes$

To complete the proof, we note the following:

Lemma 4.8.5. Suppose that $\nvdash_{\mathsf{PT}} \phi$. Then there is a complete theory Γ^+ over $\mathscr{L}_{\{\phi\}}$ such that $\Gamma^+ \nvdash_{\mathsf{PT}} \phi$.

Proof. Suppose that $\psi_1, \dots, \psi_m$ are the members of $\{\theta \mathbin{\square\!\!\rightarrow} \alpha \in \mathscr{L}_{\{\phi\}} : \theta$ is a conjunction of atomic sentences without repetitions and α is atomic$\}$ and suppose that $\chi_1, \dots, \chi_m$ are the members of $\{\theta \mathbin{\square\!\!\rightarrow} \neg\alpha \in \mathscr{L}_{\{\phi\}} : \theta$ is a conjunction of atomic sentences without repetitions and α is atomic$\}$. Now, let $\Gamma_0 = \emptyset$, and given Γ_n, let $\Gamma_{n+1} = \Gamma_n \cup \{\psi_{n+1}\}$ if $\Gamma_n \cup \{\psi_{n+1}\} \nvdash_{\mathsf{PT}} \phi$ and let $\Gamma_{n+1} = \Gamma_n \cup \{\chi_{n+1}\}$ otherwise.

Note that for any $n \in \omega$, if $\Gamma_n \nvdash_{\mathsf{PT}} \phi$, then $\Gamma_{n+1} \nvdash_{\mathsf{PT}} \phi$. Suppose that $\Gamma_{n+1} \vdash_{\mathsf{PT}} \phi$. Then $\Gamma_n \cup \{\psi_{n+1}\} \vdash_{\mathsf{PT}} \phi$ and $\Gamma_n \cup \{\chi_{n+1}\} \vdash_{\mathsf{PT}} \phi$, and so for some $\theta, \alpha \in \mathscr{L}_\Delta$, $\Gamma_n \cup \{\theta \mathbin{\square\!\!\rightarrow} \alpha\} \vdash_{\mathsf{PT}} \phi$ and $\Gamma_n \cup \{\theta \mathbin{\square\!\!\rightarrow} \neg\alpha\} \vdash_{\mathsf{PT}} \phi$. Then $\Gamma_n \cup \{(\theta \mathbin{\square\!\!\rightarrow} \alpha) \vee (\theta \mathbin{\square\!\!\rightarrow} \neg\alpha)\} \vdash_{\mathsf{PT}} \phi$. But by A11, we also have $\vdash_{\mathsf{PT}} (\theta \mathbin{\square\!\!\rightarrow} \alpha) \vee (\theta \mathbin{\square\!\!\rightarrow} \neg\alpha)$. So, $\Gamma_n \vdash_{\mathsf{PT}} \phi$. This tells us that if we let $\Gamma^+ = \Gamma_m$, then Γ^+ is a complete theory over $\mathscr{L}_\phi$ such that $\Gamma^+ \nvdash_{\mathsf{PT}} \phi$. $\boxtimes$

Now suppose $\vdash_{\mathsf{PT}} \phi$. Let Γ^+ be a complete theory over $\mathscr{L}_{\{\phi\}}$ such that $\Gamma^+ \nvdash_{\mathsf{PT}} \phi$. Theorem 4.8.4 tells us that there is a PT-model that validates Γ^+ but that does not validate ϕ. So there is a PT-model that does not validate ϕ. So $\nvDash_{\mathsf{PT}} \phi$. $\boxtimes$

Corollary 4.8.6. PT is decidable.

Proof. We saw in the proof of Theorem 4.8.1 that each relevant model for some restricted language $\mathcal{L}_{\{\phi\}}$ has as its set of worlds objects that amount to state descriptions. So there will only be finitely many such models, and so to decide whether $\vdash_{\mathsf{PT}} \phi$, we only need to check finitely many finite models. $\boxtimes$

4.9 Intended models

Definition 4.9.1.   (i) A *T-interpretation* is a function f that maps the atomic sentences of $\mathcal{L}$ to $\wp(\omega)$.

  (ii) For f a T-interpretation, the *intended* PT-*model* $\mathfrak{M}_i$ on f is the tuple $\langle \mathbf{I}, \{\mathbf{0}\}, \subseteq, \mathfrak{I} \rangle$, where $\mathfrak{I}$ is defined such that for atomic $\alpha \in \mathcal{L}$, $\mathfrak{I}(\alpha) = \{\mathbf{i} \in \mathbf{I} : \text{for some } \mathbf{d} \in \mathbf{i}, f(\alpha) \in \mathbf{d}\}$.

  (iii) $\phi \in \mathcal{L}$ is PT$_i$-*valid* ($\vDash_{\mathsf{PT}_i} \phi$) iff, for every T-interpretation f, ϕ is valid in the intended PT-model on f.

It is immediately apparent that every intended PT-model is a PT-model and so that if $\vDash_{\mathsf{PT}} \phi$, then $\vDash_{\mathsf{PT}_i} \phi$. By Theorem 4.6.1, we then have that if $\vdash_{\mathsf{PT}} \phi$, then $\vDash_{\mathsf{PT}_i} \phi$.

Theorem 4.9.1. If $\vDash_{\mathsf{PT}_i} \phi$, then $\vdash_{\mathsf{PT}} \phi$.

Proof. Suppose that $\nvdash_{\mathsf{PT}} \phi$. The proof of Theorem 4.8.1 shows that there is a finite PT-model $\mathfrak{M} = \langle \mathfrak{W}, w_@, \leq, \mathfrak{I} \rangle$ such that ϕ is not valid in $\mathfrak{M}$. From $\mathfrak{M}$ we can recover the upper semilattice $\langle \mathfrak{W}, \leq \rangle$ with minimal element $w_@$. Next, note that, by Theorem 3.6 of Chapter II of Lerman Lerman (1980), there is a countably infinite sequence of independent Turing degrees, meaning that no member of the sequence is reducible to a finite join of any members of the sequence. Call this the *Lerman sequence*. We want to use the Lerman sequence together with the fact that $\mathfrak{M}$ is finite to obtain an intended PT-model $\mathfrak{M}_i$ such that for all $\psi \in \mathcal{L}_{\{\phi\}}$, ψ is valid in $\mathfrak{M}$ iff ψ is valid in $\mathfrak{M}_i$.

We obtain $\mathfrak{M}_i$ by fixing a particular T-interpretation f. We first pick a representative set of natural numbers of each member of the Lerman sequence. Associate a set of natural numbers with each member of the set of worlds $\mathfrak{W}$ of $\mathfrak{M}$, as follows:

- Associate a member of the computable degree $\mathbf{0}$ with the minimal world $w_@$ of $\mathfrak{W}$.

- If $w \in \mathfrak{W}$ is the $\leq$-join of $v_1 \in \mathfrak{W}, \ldots, v_n \in \mathfrak{W}$ such that $v_1 < w$ and $\ldots$ and $v_n < w$ and there is no $u \in \mathfrak{W}$ such that $v_1 < u < w$ or $\ldots$ or $v_n < u < w$, and if $N_1, \ldots, N_n$ are the sets associated with $v_1, \ldots, v_n$, respectively, then let the set associated with w be $N_1 \oplus \cdots \oplus N_n$.

- If $v \in \mathfrak{W}$ is not the $\leq$-join of any $v_1 \in \mathfrak{W}, \ldots, v_n \in \mathfrak{W}$ but v is such that there is some $w \in \mathfrak{W}, w < v$ such that there is no $u \in \mathfrak{W}$ such that $w < u < v$, and N_i is the set associated with w, then take the set N_j that is the first member of the Lerman sequence that has not been used yet. Let v be associated with $N_i \oplus N_j$.

Once each member of $\mathfrak{W}$ has a set of natural numbers associated with it, let the T-interpretation f be such that for atomic $\alpha \in \mathscr{L}_{\{\phi\}}$, $f(\alpha) = \{N \in \omega : N$ is the set of natural numbers associated with the $\leq$-least $w \in \mathfrak{W}$ such that $[\![\alpha]\!]_{\mathfrak{M},w} = 1\}$.

Constructing the T-interpretation f in this way guarantees that the intended PT-model $\mathfrak{M}_i$ on f is such that for all atomic conditionals $\psi \in \mathscr{L}_{\{\phi\}}$, ψ is valid in $\mathfrak{M}$ iff ψ is valid in $\mathfrak{M}_i$. The same holds if ψ is a Boolean combination of atomic conditionals and conditional-free sentences. And if ψ is not a a Boolean combination of atomic conditionals and conditional-free sentences, then we can use Theorem 4.8.5 to find a complete theory Γ^+ over $\mathscr{L}_{\{\phi\}}$ such that $\Gamma^+ \nvdash_{\mathsf{PT}} \psi$ and Theorem 4.8.2.(iv) to find a $\chi \in \mathscr{L}_{\{\phi\}}$ that is a Boolean combination of atomic conditionals and conditional-free sentences such that $\Gamma^+ \vdash_{\mathsf{PT}} \psi \leftrightarrow \chi$. We then know from Theorem 4.8.1 how to construct a finite PT-model in which all members of Γ^+ are valid but χ is not. Using this new model, we can repeat the previous construction to obtain a new T-interpretation and an intended PT-model on it in which χ and thus also ψ is not valid.

It follows that if $\nvDash_{\mathsf{PT}} \phi$, then there is an intended PT-model in which ϕ is not valid, and so $\nvDash_{\mathsf{PT}_i} \phi$. $\boxtimes$

Corollary 4.9.2. Let $\vDash_{\mathsf{PT}_{i_{c.e.}}}$ be defined as in 4.9.1 except that we restrict the ideals to the ideals on the computably enumerable degrees. Then if $\vDash_{\mathsf{PT}_{i_{c.e.}}} \phi$, then $\vdash_{\mathsf{PT}} \phi$.

Proof. Thomason (1971) shows that there is a countably infinite sequence of independent computably enumerable degrees. Thus, the proof of Theorem 4.9.1 goes through unchanged. $\boxtimes$

4.10 On the full propositional language

Whether the present conditional logic of Turing reducibility for the full propositional language that allows for conditionals in the antecedents of conditionals can be completely axiomatized is an open question. We close by explaining why the previous strategy of proving completeness would not carry over to the case of the full propositional language.

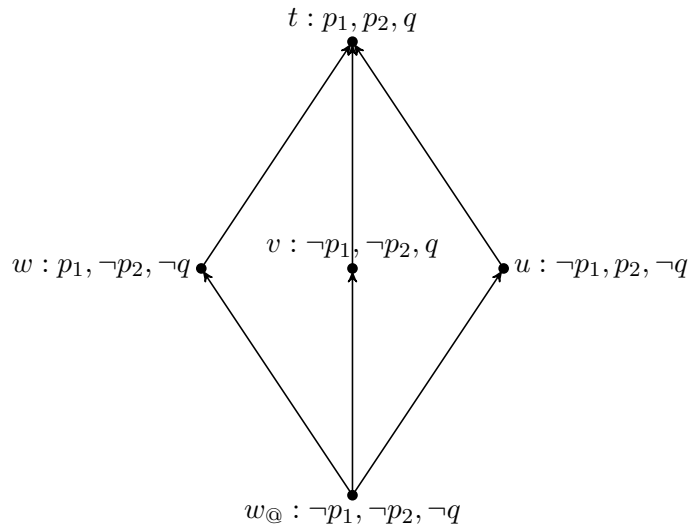In the full propositional language, we will have models such as the following:
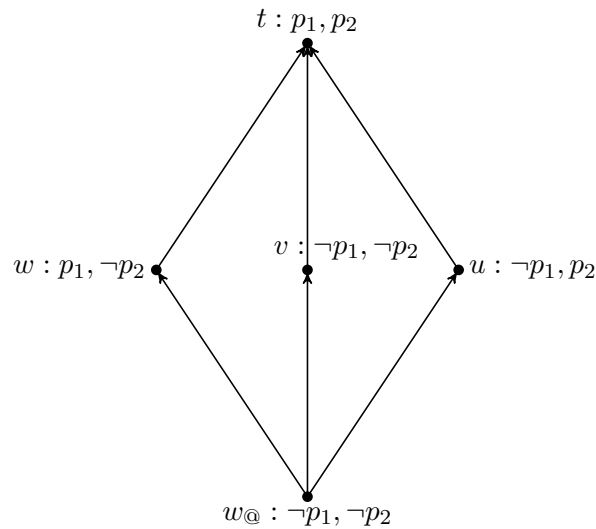
Here, $(p_1 \boxright p_2) \boxright p_2$ is false at $w_@$ because $(p_1 \boxright p_2)$ is true at v but p_2 is false at v. But now note that v and $w_@$ are identical state descriptions over $\{p_1, p_2\}$. Since the method we used to prove completeness for the restricted language amounted to treating worlds as state descriptions, using that method would have the above model collapse into the following model:

$$t : p_1, p_2$$

$$w : p_1, \neg p_2 \qquad u : \neg p_1, p_2$$

$$w_@ : \neg p_1, \neg p_2$$
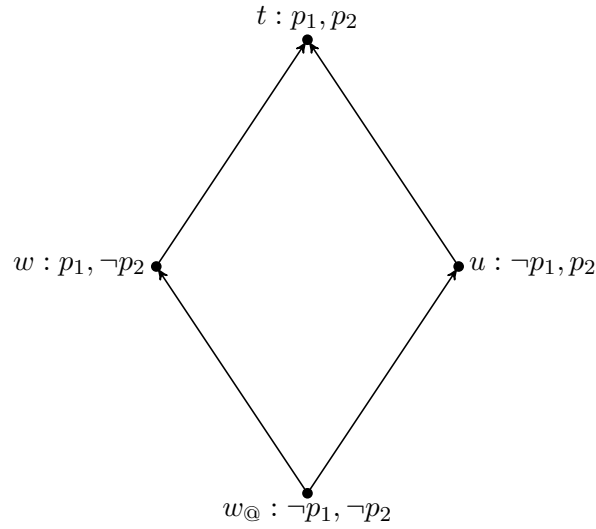
But in this model, $(p_1 \;\Box\!\!\rightarrow\; p_2) \;\Box\!\!\rightarrow\; p_2$ is true at $w_@$. So we cannot use the previous strategy of proving completeness.

This means that, using only p_1 and p_2, our restricted language cannot express certain claims about Turing reducibility which the full propositional language can express. To see this, consider the following model, which is a version of the first one but with the atomic letter q added:

$$t : p_1, p_2, q$$

$$w : p_1, \neg p_2, \neg q \qquad v : \neg p_1, \neg p_2, q \qquad u : \neg p_1, p_2, \neg q$$

$$w_@ : \neg p_1, \neg p_2, \neg q$$

If p_1, p_2, and q are interpreted to express the claims that the sets A_1, A_2, and B are algorithmically decidable respectively, this model represents the conjunction of the following claims about Turing reducibility: $A_1 \nleq_T A_2$, $A_2 \nleq_T A_1$, $A_1 \nleq_T B$, $B \nleq_T A_1$, $A_2 \nleq_T B$, $B \nleq_T A_2$, and $A_2 \leq_T$

$A_1 \oplus B$. We can of course express this in our restricted language using the letters p_1, p_2, and q as follows:

$$\neg\big(p_2 \boxright p_1\big) \wedge \neg\big(p_1 \boxright p_2\big) \wedge \neg\big(q \boxright p_1\big) \wedge \neg\big(p_1 \boxright q\big) \wedge \neg\big(q \boxright p_2\big) \wedge \neg\big(p_2 \boxright q\big) \wedge \big((p_1 \wedge q) \boxright p_2\big)$$

But as we just saw, our restricted language cannot express this using only p_1 and p_2. The full propositional language, on the other hand, can express this as follows:

$$\neg\big(p_2 \boxright p_1\big) \wedge \neg\big(p_1 \boxright p_2\big) \wedge \neg\big((p_1 \boxright p_2) \boxright p_2\big)$$

Bibliography

Almotahari, M. (2017). 'Not' again! Another essay on the metaphysics of material objects. *Philosophy and Phenomenological Research*, 94(3):711–737.

Alxatib, S. and Pelletier, F. J. (2011). The psychology of vagueness: Borderline cases and contradictions. *Mind and Language*, 26(3):287–326.

Anderson, A. R. and Belnap Jr., N. (1961). First degree entailments. *Mathematische Annalen*, 149(4):302–319.

Armour-Garb, B. P. and Beall, J. (2005a). *Deflationary Truth*. Open Court, Chicago.

Armour-Garb, B. P. and Beall, J. (2005b). Deflationism: The basics. In Armour-Garb and Beall (2005a), pages 1–31.

Armour-Garb, B. P. and Priest, G. (2005). Analetheism: A Pyrrhic victory. *Analysis*, 65(2):167–173.

Asenjo, F. G. (1966). A calculus of antinomies. *Notre Dame Journal of Formal Logic*, 7(1):103–105.

Austin, J. L. (1962). *How to Do Things with Words*. Harvard University Press, Cambridge, Mass.

Avron, A. (1991). Natural 3-valued logics—characterization and proof theory. *Journal of Symbolic Logic*, 56(1):276–294.

Ball, D. and Rabern, B., editors (2017). *The Science of Meaning*. Oxford University Press, Oxford.

Baras, D. (MS). How close are impossible worlds? Unpublished manuscript.

Barcan, R. C. (1946). A functional calculus of first order based on strict implication. *Journal of Symbolic Logic*, 11(1):1–16.

Batens, D. (1990). Against global paraconsistency. *Studies in Soviet Thought*, 39(3):209–229.

Beall, J. (2007a). Prolegomenon to future revenge. In Beall (2007b), pages 1–29.

Beall, J. (2007b). *Revenge of the Liar. New Essays on the Paradox*. Oxford University Press, Oxford.

Beall, J. (2009). *Spandrels of Truth*. Oxford University Press, Oxford.

Beall, J. (2011). Multiple-conclusion LP and default classicality. *Review of Symbolic Logic*, 4(2):326–336.

Beall, J. (2013a). Shrieking against gluts: The solution to the 'just true' problem. *Analysis*, 73(3):438–445.

Beall, J. (2013b). A simple approach towards recapturing consistent theories in paraconsistent settings. *Review of Symbolic Logic*, 6(4):755–764.

Beall, J. (2014). Finding tolerance without gluts. *Mind*, 123(491):791–811.

Beall, J. (2015a). Free of detachment: Logic, rationality, and gluts. *Noûs*, 49(2):410–423.

Beall, J. (2015b). Non-detachable validity and deflationism. In Caret and Hjortland (2015), pages 276–288.

Beall, J. (2017). There is no logical negation: True, false, both and neither. *Australasian Journal of Logic*, 14(1):1–29.

Beaney, M., editor (1997). *The Frege Reader*. Blackwell, Oxford.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford University Press, Oxford.

Berger, U., Diener, H., Schuster, P., and Seisenberger, M., editors (2012). *Logic, Construction, Computation*. Ontos, Heusenstamm.

Bertlett, S. J. and Suber, P., editors (1987). *Self-Reference. Reflections on Reflexivity*. Martinus Nijhoff Publishers, Dordrecht.

Berto, F. (2013). Impossible worlds. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Center for the Study of Language and Information, Stanford University, Stanford, winter 2013 edition.

Berto, F. (2014). Absolute contradiction, dialetheism, and revenge. *Review of Symbolic Logic*, 7(2):193–207.

Berto, F., French, R., Priest, G., and Ripley, D. (2017). Williamson on counterpossibles. *Journal of Philosophical Logic*.

Bonacina, M. P. and Stickel, M. E., editors (2013). *Automated Reasoning and Mathematics. Essays in Memory of William W. McCune*. Springer, Heidelberg.

Boolos, G. (1993). *The Logic of Provability*. Cambridge University Press, Cambridge.

Boolos, G. S., Burgess, J. P., and Jeffrey, R. C. (2007). *Computability and Logic*. Cambridge University Press, Cambridge, 5th edition.

Boutilier, C. (1990). Conditional logics of normality as modal systems. In *AAAI-90: Proceedings of the 8th National Conference on Artificial Intelligence*, pages 594–599, Cambridge, Mass. American Association for Artificial Intelligence, MIT Press.

Brandom, R. B. (1994). *Making It Explicit. Reasoning, Representing, and Discursive Commitment*. Harvard University Press, Cambridge, Mass.

Briggs, R. (2012). Interventionist counterfactuals. *Philosophical Studies*, 160(1):139–166.

Brochhagen, T., Roelofsen, F., and Theiler, N., editors (2015). *Proceedings of the 20th Amsterdam Colloquium*. Institute for Logic, Language and Computation, Amsterdam.

Brogaard, B. and Salerno, J. (2013). Remarks on counterpossibles. *Synthese*, 190(4):639–660.

Bromand, J. (2002). Why paraconsistent logic can only tell half the truth. *Mind*, 111(444):741–749.

Burgess, J. P. (1981). Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic*, 22(1):76–84.

Caret, C. R. and Hjortland, O. T., editors (2015). *Foundations of Logical Consequence*. Oxford University Press, Oxford.

Carroll, L. (1895). What the tortoise said to Achilles. *Mind*, 4(14):278–280.

Carston, R. (1996). Metalinguistic negation and echoic use. *Journal of Pragmatics*, 25(3):309–330.

Carston, R. (2002). *Thoughts and Utterances. The Pragmatics of Explicit Communication*. Blackwell, Oxford.

Chisholm, R. M. (1946). The contrary-to-fact conditional. *Mind*, 55(220):289–307.

Church, A. (1936a). A note on the *Entscheidungsproblem*. *Journal of Symbolic Logic*, 1(1):40–41.

Church, A. (1936b). An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58(2):345–363.

Cleland, C. E. (1993). Is the Church-Turing thesis true? *Minds and Machines*, 3(3):283–312.

Cobreros, P., Égré, P., Ripley, D., and van Rooij, R. (2012a). Tolerance and mixed consequence in the s'valuationist setting. *Studia Logica*, 100(4):855–877.

Cobreros, P., Égré, P., Ripley, D., and van Rooij, R. (2012b). Tolerant, Classical, Strict. *Journal of Philosophical Logic*, 41(2):347–384.

Cobreros, P., Égré, P., Ripley, D., and van Rooij, R. (2013). Reaching transparent truth. *Mind*, 122(488):841–866.

Colyvan, M. (2009). Vagueness and truth. In Dyke (2009), pages 29–40.

Cooper, S. B. (2004). *Computability Theory*. Chapman & Hall/CRC, Boca Raton.

Cooper, S. B., Löwe, B. L., and Sorbi, A., editors (2008). *New Computational Paradigms. Changing Conceptions of What Is Computable*. Springer, New York.

Copeland, B. J., Posy, C. J., and Shagrir, O., editors (2013). *Computability. Turing, Gödel, Church, and Beyond*. MIT Press, Cambridge, Mass.

Cresswell, M. J. and Hughes, G. E. (1996). *A New Introduction to Modal Logic*. Routledge, London.

Davis, M. (1958). *Computability and Unsolvability*. McGraw-Hill, New York.

Davis, M. (2004). The myth of hypercomputation. In Teuscher (2004), pages 195–211.

Davis, W. A. (2016). *Irregular Negatives, Implicatures, and Idioms*. Springer, Dordrecht.

Dietz, R. and Moruzzi, S., editors (2010). *Cuts and Clouds. Vaguenesss, its Nature and its Logic*. Oxford University Press, Oxford.

Dowden, B. H. (1984). Accepting inconsistencies from the paradoxes. *Journal of Philosophical Logic*, 13(2):125–130.

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

Dummett, M. and Crossley, J. N., editors (1965). *Formal Systems and Recursive Functions. Proceedings of the Eighth Logic Colloquium, Oxford, July 1963*. North-Holland, Amsterdam.

Dutilh Novaes, C. (2015). A dialogical, multi-agent account of the normativity of logic. *Dialectica*, 69(4):587–609.

Dutilh Novaes, C. (2016). *Reductio ad absurdum* from a dialogical perspective. *Philosophical Studies*, 173(10):2605–2628.

Dyke, H., editor (2009). *From Truth to Reality. New Essays in Logic and Metaphysics*. Routledge, New York.

Earman, J. (1995). *Bangs, Crunches, Whimpers, and Shrieks. Singularities and Acausality in Relativistic Spacetimes*. Oxford University Press, Oxford.

Egan, A. (2008). Pretense for the complete idiom. *Noûs*, 42(3):381–409.

Egan, A. and Weatherson, B., editors (2011). *Epistemic Modality*. Oxford University Press, Oxford.

Ellis, B., Jackson, F., and Pargetter, R. (1977). An objection to possible-world semantics for counterfactual logics. *Journal of Philosophical Logic*, 6(1):355–357.

Enderton, H. B. (2011). *Computability Theory. An Introduction to Recursion Theory*. Elsevier Science, Amsterdam.

Estrada-González, L. and Olmedo-García, C. (2013). Review of Paul Kabay's *On the Plenitude of Truth. A Defense of Trivialism. Disputatio*, 5(35):93–98.

Etchemendy, J. (1990). *The Concept of Logical Consequence*. Harvard University Press, Cambridge, Mass.

Everett, A. (1994). Absorbing dialetheia? *Mind*, 103(412):413–420.

Everett, A. (1996). A dilemma for Priest's dialethism. *Australasian Journal of Philosophy*, 74(4):657–668.

Feferman, S. (1984). Toward Useful Type-Free Theories. I. *Journal of Symbolic Logic*, 49(1):75–111.

Feferman, S. (2012). Axiomatizing truth: Why and how? In Berger et al. (2012), pages 185–200.

Field, H. H. (1994). Deflationist views of meaning and content. *Mind*, 103(411):249–285.

Field, H. H. (2008). *Saving Truth from Paradox*. Oxford University Press, Oxford.

Fine, K. (1975a). Critical notice of David Lewis' *Counterfactuals. Mind*, 84(1):451–458.

Fine, K. (1975b). Vagueness, truth, and logic. *Synthese*, 30(3):265–300.

Fine, K. (1977). Prior on the construction of possible worlds and instants. In Fine (2005), pages 133–175.

Fine, K. (2005). *Modality and Tense. Philosophical Papers*. Oxford University Press, Oxford.

Fine, K. (2012). Counterfactuals without possible worlds. *Journal of Philosophy*, 109(3):221–246.

von Fintel, K. (2001). Counterfactuals in a dynamic context. In Kenstowicz (2001), pages 123–152.

Fitelson, B. (2013). Gibbard's collapse theorem for the indicative conditional: An axiomatic approach. In Bonacina and Stickel (2013), pages 181–188.

Foolen, A. (1991). Metalinguistic negation and pragmatic ambiguity: Some comments on a proposal by Laurence Horn. *Pragmatics*, 1(2):217–237.

Forbes, G. (1989). *Languages of Possibility. An Essay in Philosophical Logic*. Blackwell, Oxford.

Forbes, G. (1992). Melia on modalism. *Philosophical Studies*, 68(1):57–63.

Fox, D. (2007). Free choice and the theory of scalar implicatures. In Sauerland and Stateva (2007), pages 71–120.

Frege, G. (1918). Negation. In Beaney (1997), pages 346–361.

Frege, G. (1956). The thought: A logical inquiry. *Mind*, 65(259):289–311.

Friedberg, R. M. (1957). Two recursively enumerable sets of incomparable degrees of unsolvability (solution of Post's problem, 1944). *Proceedings of the National Academy of Sciences of the United States of America*, 43(2):236.

Gabbay, D. M. (1999). *What Is Negation?* Kulwer, Dordrecht.

Geach, P. T. (1965). Assertion. *Philosophical Review*, 74(4):449–465.

Geurts, B. (1998). The mechanisms of denial. *Language*, 74(2):274–307.

Gibbard, A. (1980). Two Recent Theories of Conditionals. In Harper et al. (1980), pages 211–247.

Gillies, A. S. (2007). Counterfactual scorekeeping. *Linguistics and Philosophy*, 30(3):329–360.

Goodman, N. (1947). The problem of counterfactual conditionals. *Journal of Philosophy*, 44(5):113–128.

Grice, H. P. (1967). Logic and conversation. In Grice (1989), pages 1–143.

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press, Cambridge, Mass.

Hájek, P., Valdś-Villanueva, L., and Westerståhl, D., editors (2005). *Logic, Methodology and Philosophy of Science. Proceedings of the Twelfth International Congress*. Kings' College Publications, London.

Halbach, V. (2011). *Axiomatic Theories of Truth*. Cambridge University Press, Cambridge.

Hale, B. and Hoffmann, A., editors (2010). *Modality. Metaphysics, Logic, and Epistemology*. Oxford University Press, Oxford.

Hamkins, J. D. (2004). Supertask computation. In Löwe et al. (2004), pages 141–158.

Harman, G. (1986). *Change in View. Principles of Reasoning*. MIT Press, Cambridge, Mass.

Harman, G. (2009). Field on the normative role of logic. *Proceedings of the Aristotelian Society*, 109(1pt3):333–335.

Harper, W. L., Stalnaker, R. C., and Pearce, G., editors (1980). *Ifs. Conditionals, Belief, Decision, Chance, and Time*. D. Reidel, Dordrecht.

Heck, Jr., R. G. (1997). *Language, Thought, and Logic. Essays in Honour of Michael Dummett*. Oxford University Press, Oxford.

Heim, I. and Kratzer, A. (1998). *Semantics in Generative Grammar*. Blackwell, Malden, Mass.

Horn, L. R. (1985). Metalinguistic negation and pragmatic ambiguity. *Language*, 61(1):121–174.

Horn, L. R. (1989). *A Natural History of Negation*. University of Chicago Press, Chicago and London.

Horn, L. R. and Ward, G. L., editors (2007). *The Handbook of Pragmatics*. Blackwell, Oxford.

Hughes, M. (2015). Necessary truths are just true: A reply to Rossberg. *Thought*, 3(4):321–331.

Hyde, D. (1997). From heaps and gaps to heaps of gluts. *Mind*, 106(424):641–660.

Iwata, S. (1998). Some extensions of the echoic analysis of metalinguistic negation. *Lingua*, 105(1-2):49–65.

Jenny, M. (2017). Classicality lost: K3 and LP after the fall. *Thought*, 6(1):43–53.

Kenstowicz, M. J., editor (2001). *Ken Hale. A Life in Language*. MIT Press, Cambridge, Mass.

Khoo, J. (2013). A note on Gibbard's proof. *Philosophical Studies*, 166(S):153–164.

Khoo, J. (2015). Modal disagreements. *Inquiry*, 58(5):511–534.

Kleene, S. C. (1950). *Introduction to Metamathematics*. Van Nostrand, New York.

Kleene, S. C. and Post, E. L. (1954). The upper semi-lattice of degrees of recursive unsolvability. *Annals of Mathematics*, 59(3):379–407.

Kment, B. (2014). *Modality and Explanatory Reasoning*. Oxford University Press, Oxford.

Knuth, D. E. (1966). Algorithm and program; information and data. *Communications of the ACM*, 9(9):654.

Kratzer, A. (1986). Conditionals. *Chicago Linguistics Society*, 22(2):1–15.

Kraus, S., Lehmann, D., and Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44(1-2):167–207.

Kripke, S. A. (1965). Semantical analysis of intuitionistic logic I. In Dummett and Crossley (1965), pages 92–130.

Kripke, S. A. (1975). Outline of a theory of truth. *Journal of Philosophy*, 72(19):690–716.

Kripke, S. A. (1980). *Naming and Necessity*. Harvard University Press, Cambridge, Mass.

Kripke, S. A. (2012). The Church-Turing "thesis" as a special corollary of Gödel's completeness theorem. In Copeland et al. (2013), pages 77–104.

Kühnlein, P., Reiser, H., and Zeevat, H., editors (2003). *Perspectives on Dialogue in the New Millenium*. John Benjamins, Amsterdam.

Ladusaw, W. A. (1980). *Polarity Sensitivity as Inherent Scope Relations*. Indiana University Linguistics Club, Bloomington.

Lerman, M. (1980). *Degrees of Unsolvability. Local and Global Theory*. Springer, Berlin.

Lewis, D. K. (1971). Completeness and decidability of three logics of counterfactual conditionals. *Theoria*, 37(1):74–85.

Lewis, D. K. (1973). *Counterfactuals*. Harvard University Press, Cambridge, Mass.

Lewis, D. K. (1981). Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic*, 10(2):217–234.

Lewis, D. K. (1986). *On the Plurality of Worlds*. Blackwell, Oxford.

Lewis, D. K. (1991). *Parts of Classes*. Blackwell, Oxford.

Lewis, K. (2017). Counterfactual discourse in context. *Noûs*.

Littmann, G. and Simmons, K. (2004). A critique of dialetheism. In Priest et al. (2004), pages 314–335.

Liu, J. and Berger, D. L., editors (2014). *Nothingness in Asian Philosophy*. Routledge, London.

Löwe, B., Piwinger, B., and Räsch, T., editors (2004). *Classical and New Paradigms of Computation and Their Complexity Hierarchies. Papers of the Conference "Foundations of the Formal Sciences III"*. Kluwer Academic Publishers, Dordrecht.

MacFarlane, J. (2004). In what sense (if any) is logic normative for thought? Unpublished manuscript, available at `http://johnmacfarlane.net/normativity_of_logic.pdf`, accessed June 3rd, 2017.

Mancosu, P. and Zach, R. (2015). Heinrich Behmann's 1921 lecture on the decision problem and the algebra of logic. *Bulletin of Symbolic Logic*, 21(2):164–187.

Martin, R. L. and Woodruff, P. W. (1975). On representing 'true-in-L' in L. *Philosophia*, 5(3):113–117.

McCall, S. (1966). Connexive implication. *Journal of Philosophical Logic*, 31(3):415–433.

McGee, V. (1985). A counterexample to modus ponens. *Journal of Philosophy*, 82(9):462–471.

McGee, V. (1990). *Truth, Vagueness, and Paradox. An Essay on the Logic of Truth*. Hackett, Indianapolis.

McGee, V. (2005). Two conceptions of truth?—comment. *Philosophical Studies*, 124(1):71–104.

McGee, V. (2006). There are many things. In Thomson and Byrne (2006), pages 93–122.

McGee, V. (2010). Field's logic of truth. *Philosophical Studies*, 147(3):421–432.

Mortensen, C. (2013). Inconsistent Mathematics. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Center for the Study of Language and Information, Stanford University, Stanford, spring 2013 edition.

Moschovakis, Y. N. (1974). *Elementary Induction on Abstract Structures*. North Holland, Amsterdam.

Moss, S. (2012). On the pragmatics of counterfactuals. *Noûs*, 46(3):561–586.

Muchnik, A. A. (1956). On the unsolvability of the problem of reducibility in the theory of algorithms (russian). *Doklady Akademii Nauk SSSR*, 108(2):194–197.

Murzi, J. and Carrara, M. (2015a). Denial and disagreement. *Topoi*, 34(1):109–119.

Murzi, J. and Carrara, M. (2015b). Paradox and logical revision. A short introduction. *Topoi*, 34(1):7–14.

Ninan, D. (2010). Semantics and the objects of assertion. *Linguistics and Philosophy*, 33(5):355–380.

Nolan, D. (1997). Impossible worlds: A modest approach. *Notre Dame Journal of Formal Logic*, 38(4):535–572.

Nouwen, R., van Rooij, R., Sauerland, U., and Schmitz, H.-C. C., editors (2011). *Vagueness in Communication*. Springer, Heidelberg.

Nunberg, G., Sag, I. A., and Wasow, T. (1994). Idioms. *Language*, 70(3):491–538.

Odifreddi, P. (1992). *Classical Recursion Theory*, volume 2. Elsevier, Amsterdam.

Olin, D. (2003). *Paradox*. Acumen, Chesham.

Olszewski, A., Woleński, J., and Janusz, R. (2006). *Church's Thesis After 70 Years*. Ontos, Frankfurt.

Papadimitriou, C. H. (1994). *Computational Complexity*. Addison-Wesley, Reading, Mass.

Parsons, T. (1984). Assertion, denial, and the liar paradox. *Journal of Philosophical Logic*, 13(2):137–152.

Parsons, T. (1990). True contradictions. *Canadian Journal of Philosophy*, 20(3):335–353.

Piccinini, G. (2015). Computation in physical systems. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Center for the Study of Language and Information, Stanford University, Stanford, summer 2015 edition.

Pitts, A. (2011). Exploring a 'pragmatic ambiguity' of negation. *Language*, 87(2):346–368.

Pollock, J. L. (1976). *Subjunctive Reasoning*. Reidel, Dordrecht.

Post, E. L. (1944). Recursively enumerable sets of positive integers and their decision problems. *Bulletin of the American Mathematical Society*, 50(5):284–316.

Price, H. (1983). Sense, assertion, Dummett and denial. *Mind*, 92(366):161–173.

Price, H. (1990). Why 'not'? *Mind*, 99(394):221–238.

Priest, G. (1979). The logic of paradox. *Journal of Philosophical Logic*, 8(1):219–241.

Priest, G. (2002). *Beyond the Limits of Thought*. Oxford University Press, Oxford.

Priest, G. (2006). *In Contradiction. A Study of the Transconsistent*. Oxford University Press, Oxford, 2nd edition.

Priest, G. (2008a). *Doubt Truth to Be a Liar*. Oxford University Press, Oxford.

Priest, G. (2008b). *An Introduction to Non-Classical Logic. From If to Is*. Cambridge University Press, Cambridge, 2nd edition.

Priest, G. (2010). Inclosures, vagueness, and self-reference. *Notre Dame Journal of Formal Logic*, 51(1):69–84.

Priest, G. (2014a). *One. Being an Investigation into the Unity of Reality and of Its Parts, Including the Singular Object which Is Nothingness*. Oxford University Press, Oxford.

Priest, G. (2014b). Speaking of the ineffable. . . . In Liu and Berger (2014), pages 91–103.

Priest, G., Beall, J., and Armour-Garb, B. P., editors (2004). *The Law of Non-Contradiction. New Philosophical Essays*. Oxford University Press, Oxford.

Quine, W. V. (1948). On what there is. *Review of Metaphysics*, 2(5):21–36.

Quine, W. V. (1953). Three grades of modal involvement. In Quine (1966), pages 156–174.

Quine, W. V. (1966). *The Ways of Paradox and Other Essays*. Random House, New York.

Quine, W. v. O. (1960). *Word and object*. MIT Press, Cambridge, Mass.

Rayo, A. (2013). A plea for semantic localism. *Noûs*, 47(3):647–679.

Rescher, N., editor (1968). *Studies in Logical Theory*. Blackwell, Oxford.

Rescorla, M. (2007). Church's thesis and the conceptual analysis of computability. *Notre Dame Journal of Formal Logic*, 48(2):253–280.

Restall, G. (2005). Multiple Conclusions. In Hájek et al. (2005), pages 189–205.

Richard, M. (2008). *When Truth Gives Out*. Oxford University Press, Oxford.

Richard, M. (2009). Indeterminacy and truth-value gaps. In Dietz and Moruzzi (2010), pages 464–481.

Ripley, D. (2011). Contradictions at the borders. In Nouwen et al. (2011), pages 169–188.

Ripley, D. (2012). Conservatively extending classical logic with transparent truth. *Review of Symbolic Logic*, 5(2):354–378.

Ripley, D. (2013a). Paradoxes and failures of cut. *Australasian Journal of Philosophy*, 91(1):139–164.

Ripley, D. (2013b). Revising up: Strengthening classical logic in the face of paradox. *Philosophers' Imprint*, 13(5):1–13.

Ripley, D. (2015). Embedding denial. In Caret and Hjortland (2015), pages 289–309.

Rogers, Jr., H. (1967). *Theory of Recursive Functions and Effective Computability*. McGraw-Hill, New York.

Romero, G. E. (2014). The collapse of supertasks. *Foundations of Science*, 19(2):209–216.

Rossberg, M. (2013). Too good to be "just true". *Thought*, 2(1):1–8.

Rumfitt, I. (2000). "Yes" and "no". *Mind*, 109(436):781–823.

Russell, B. (1905). On denoting. *Mind*, 14(56):479–493.

Sacks, G. E. (1964). The recursively enumerable degrees are dense. *Annals of Mathematics*, 80(2):300–312.

Santorio, P. (2017). Alternatives and truthmakers in conditional semantics. *Journal of Philosophy*.

Sauerland, U. and Stateva, P., editors (2007). *Presupposition and Implicature in Compositional Semantics*. Palgrave Macmillan, New York.

Scharp, K. (2013). *Replacing Truth*. Oxford University Press, Oxford.

Shagrir, O. (2004). Super-tasks, accelerating Turing machines and uncomputability. *Theoretical Computer Science*, 317(1-3):105–114.

Shapiro, S. (1981). Understanding Church's thesis. *Journal of Philosophical Logic*, 10(3):353–365.

Shapiro, S. (1997). *Philosophy of Mathematics. Structure and Ontology*. Oxford University Press, Oxford.

Shapiro, S. (2004). Simple truth, contradiction, and consistency. In Priest et al. (2004), pages 336–354.

Shapiro, S. (2006). Computability, proof, and open-texture. In Olszewski et al. (2006), pages 420–455.

Shoesmith, D. J. and Smiley, T. J. (1978). *Multiple-Conclusion Logic*. Cambridge University Press, Cambridge.

Sieg, W. (2008). Church without dogma: Axioms for computability. In Cooper et al. (2008), pages 139–152.

Simmons, K. (1993). *Universality and the Liar. An Essay on Truth and the Diagonal Argument*. Cambridge University Press, Cambridge.

Smiley, T. (1996). Rejection. *Analysis*, 56(1):1–9.

Smith, P. (2007). *An Introduction to Gödel's Theorems*. Cambridge University Press, Cambridge.

Soare, R. I. (1986). *Recursively Enumerable Sets and Degrees*. Springer, Berlin.

Soare, R. I. (1996). Computability and recursion. *Bulletin of Symbolic Logic*, 2(03):284–321.

Soare, R. I. (2009). Turing oracle machines, online computing, and three displacements in computability theory. *Annals of Pure and Applied Logic*, 160:368–399.

Soare, R. I. (2016). *Turing Computability. Theory and Applications*. Springer, Berlin.

Sperber, D. and Wilson, D. (1995). *Relevance. Communication and Cognition*. Blackwell, Oxford, 2nd edition.

Sprigge, T. L. S. (1970). *Facts, Words and Beliefs*. Humanities Press, New York.

Stainton, R. J. (2006). The pragmatics of non-sentences. In Horn and Ward (2007), pages 266–287.

Stalnaker, R. C. (1968). A theory of conditionals. In Rescher (1968), pages 98–112.

Stalnaker, R. C. (1978). Assertion. In Stalnaker (1999), pages 78–95.

Stalnaker, R. C. (1980). A defense of conditional excluded middle. In Harper et al. (1980), pages 87–104. Reidel, Dordrecht.

Stalnaker, R. C. (1984). *Inquiry*. MIT Press, Cambridge, Mass.

Stalnaker, R. C. (1996). Impossibilities. In Stalnaker (2003), pages 55–67.

Stalnaker, R. C. (1999). *Context and Content. Essays on Intentionality in Speech and Thought.* Oxford University Press, Oxford.

Stalnaker, R. C. (2003). *Ways a World Might Be. Metaphysical and Anti-Metaphysical Essays.* Oxford University Press, Oxford.

Stalnaker, R. C. (2012). *Mere Possibilities. Metaphysical Foundations of Modal Semantics.* Princeton University Press, Princeton, N.J.

Stalnaker, R. C. (2014). *Context.* Oxford University Press, Oxford.

Stalnaker, R. C. and Thomason, R. H. (1970). A semantic analysis of conditional logic. *Theoria*, 36(1):23–42.

Stanley, J. (1997). Rigidity and content. In Heck (1997), pages 131–156.

Swanson, E. (2011). How not to theorize about the language of subjective uncertainty. In Egan and Weatherson (2011), pages 249–269.

Tappenden, J. (1999). Negation, denial, and language change in philosophical logic. In Gabbay (1999), pages 261–298.

Tarski, A. (1933). The concept of truth in formalized languages. In Tarski (1983), pages 152–279.

Tarski, A. (1983). *Logic, Semantics, Metamathematics.* Hackett, Indianapolis, 2nd edition.

Teuscher, C., editor (2004). *Alan Turing. Life and Legacy of a Great Thinker.* Springer, Berlin.

Thomason, S. K. (1971). Sublattices of the recursively enumerable degrees. *Mathematical Logic Quarterly*, 17(1):273–280.

Thomson, J. F. (1954). Tasks and super-tasks. *Analysis*, 15(1):1–13.

Thomson, J. J. and Byrne, A. (2006). *Content and Modality. Themes from the Philosophy of Robert Stalnaker.* Oxford University Press, Oxford.

Todd, W. (1964). Counterfactual conditionals and the presuppositions of induction. *Philosophy of Science*, 31(2):101–110.

Turing, A. M. (1936). On computable numbers, with an application to the *Entscheidungsproblem.* *Proceedings of the London Mathematical Society*, 42(1):230–265.

Turing, A. M. (1939). Systems of logic based on ordinals. *Proceedings of the London Mathematical Society*, 2(1):161–228.

van der Sandt, R. (1988). *Context and Presupposition.* Croom Helm, London.

van der Sandt, R. (2003). Denial and presupposition. In Kühnlein et al. (2003), pages 59–79.

van der Sandt, R. and Maier, E. (2003). Denials in discourse. Unpublished manuscript, available at `http://philpapers.org/archive/VANDID`, accessed June 3rd, 2017.

van der Wouden, T. (1997). *Negative Contexts. Collocation, Polarity, and Multiple Negation.* Routledge, London.

van Fraassen, B. C. (1966). Singular terms, truth-value gaps, and free logic. *Journal of Philosophy*, 63(17):481–495.

Varzi, A. C. (1994). *Universal Semantics.* PhD thesis, University of Toronto.

Varzi, A. C. (1997). Inconsistency without Contradiction. *Notre Dame Journal of Formal Logic*, 38(4):621–639.

Veltman, F. (1985). *Logics for Conditionals.* PhD thesis, Universiteit van Amsterdam.

Weber, Z. (2010). A paraconsistent model of vagueness. *Mind*, 119(476):1025–1045.

Weber, Z., Ripley, D., Priest, G., Hyde, D., and Colyvan, M. (2014). Tolerating gluts. *Mind*, 123(491):813–828.

Welch, P. D. (2004). Post's and other problems of supertasks of higher type. In Löwe et al. (2004), pages 223–239.

Welch, P. D. (2008). Ultimate truth *vis-à-vis* stable truth. *Review of Symbolic Logic*, 1(1):126–142.

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

Williamson, T. (2009). Probability and danger. *The Amherst Lecture in Philosophy*, 4:1–35.

Williamson, T. (2010). Modal logic within counterfactual logic. In Hale and Hoffmann (2010), pages 81–96.

Williamson, T. (2013). *Modal Logic as Metaphysics*. Oxford University Press, Oxford.

Williamson, T. (2015). Counterpossibles. In Brochhagen et al. (2015), pages 30–40.

Wilson, D. and Sperber, D. (2012). *Meaning and Relevance*. Cambridge University Press, Cambridge.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. Routledge and Kegan Paul, London.

Wittgenstein, L. (2009). *Philosophical Investigations*. Blackwell, Oxford.

Woodruff, P. W. (1984). Paradox, truth and logic. Part I: Paradox and truth. *Journal of Philosophical Logic*, 13(2):213–232.

Yablo, S. (2014). *Aboutness*. Princeton University Press, Princeton.

Yalcin, S. (2015). Quantifying in from a Fregean perspective. *Philosophical Review*, 124(2):207–253.

Yalcin, S. (2017). Semantics as model-based science. In Ball and Rabern (2017).

Zardini, E. (2011). Truth without contra(di)ction. *Review of Symbolic Logic*, 4(4):498–453.