

2011-11-29

Truth is a One-Player Game: A Defense of Monaletheism and Classical Logic

Benjamin Burgis

University of Miami, benburgis@gmail.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

Recommended Citation

Burgis, Benjamin, "Truth is a One-Player Game: A Defense of Monaletheism and Classical Logic" (2011). *Open Access Dissertations*. 677.

https://scholarlyrepository.miami.edu/oa_dissertations/677

This Open access is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of Scholarly Repository. For more information, please contact repository.library@miami.edu.

UNIVERSITY OF MIAMI

TRUTH IS A ONE-PLAYER GAME: A DEFENSE OF MONALETHEISM AND
CLASSICAL LOGIC

By

Benjamin A. Burgis

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

December 2011

©2011
Benjamin A. Burgis
All Rights Reserved

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

TRUTH IS A ONE-PLAYER GAME: A DEFENSE OF MONALETHEISM AND
CLASSICAL LOGIC

Benjamin A. Burgis

Approved:

Otávio Bueno, Ph.D.
Professor of Philosophy

Terri A. Scandura, Ph.D.
Dean of the Graduate School

Edward Erwin, Ph.D.
Professor of Philosophy

Risto Hilpinen, Ph.D.
Professor of Philosophy

Hartry Field, Ph.D.
Professor of Philosophy
New York University

BURGIS, BENJAMIN

(Ph.D., Philosophy)

Truth is a One-Player Game: A Defense of Monaletheism and
Classical Logic

(December 2011)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Otávio Bueno.

No. of pages in text. (241)

The Liar Paradox and related semantic antinomies seem to challenge our deepest intuitions about language, truth and logic. Many philosophers believe that to solve them, we must give up either classical logic, or the expressive resources of natural language, or even the “naïve theory of truth” (according to which α and $\text{Tr}\langle\alpha\rangle$ [“it is true that α ”] always entail each other). A particularly extreme form of radical surgery is proposed by figures like Graham Priest, who argues for “dialetheism”—the position that some contradictions are actually true—on the basis of the paradoxes. While Priest’s willingness to dispense with the Law of Non-Contradiction may be unpopular in contemporary analytic philosophy, figures as significant as Saul Kripke and Hartry Field have argued that, in light of the paradoxes, we can only save Non-Contradiction at the expense of the Law of the Excluded Middle, abandoning classical logic in favor of a “paracomplete” alternative in which α and $\neg\alpha$ can simultaneously fail to hold. I believe that we can do better than that, and I argue for a more conservative approach, which retains not only “monaletheism” (the orthodox position that no sentence, either in natural languages or other language, can have more than one truth-value at a time), but the full inferential resources of classical logic.

For Ryan, who told me he'd only pay my bar tab if I agreed to do my Ph.D. at Miami.

Acknowledgements

I'd like to acknowledge the help and encouragement of my excellent advisor, Otávio Bueno. He got me interested in these questions in my first year of graduate work at Miami and he has helped me bounce around my ideas about them, and helped me refine and sharpen those ideas in light of his endless stream of objections and counter-arguments, ever since. I'd like to thank my good friends Ryan Lake, for coining the word in the title, Robin Neiman, for telling me that this would make a more interesting dissertation project than the alternative I'd been considering during my last semester of coursework, and Mark Warren, for many hours of fruitful philosophical conversation over glasses of peaty and delicious single malt whisky. I'd be absurdly remiss if I didn't also thank my wonderful parents, Richard and Kathy Burgis, for inadvertently planting the seeds of my interest in logic by going through a great many math problems with me with a ballpoint pen and various El Azteco restaurant napkins over the course of my childhood. (At the time, I wasn't always grateful.) I'd obviously like to thank my committee members Risto Hilpinen and Ed Erwin, and my outside member Hartry Field, a giant in the field who I felt honored to work with. Finally, I'd be remiss if I didn't thank Graham Priest, whose arguments functioned like water balloons dropped from a great height on my sleeping head, to force me to wake up, confused and alarmed, from my dogmatic slumber.

Table of Contents

Chapter One: Monaletheism and Dialetheism	1
Chapter Two: The Very Possibility of the Debate	17
Chapter Three: Motivations for Dialetheism	41
Chapter Four: Liars and Gluts	78
Chapter Five: Liars and Gaps	94
Chapter Six: Liars and Meaninglessness	110
Chapter Seven: Meaninglessness and Revenge	136
Chapter Eight: There Are No True Contradictions	154
Chapter Nine: Monaletheism and Negation	189
References	238

Chapter One: Dialetheism And Monaletheism

Dialetheism, also called “strong paraconsistency,” is the position in the philosophy of logic, championed chiefly by Graham Priest, according to which there are true statements of the form $(\alpha \wedge \neg\alpha)$. By contrast, less extreme paraconsistent views either reject or are agnostic about the possibility of true contradictions, but still reject the classical principle that “from a contradiction, everything follows” for various other reasons.

If dialetheism is at the far left wing of the spectrum and weaker forms of paraconsistency are somewhere in the middle, I situate myself at the extreme right: There are no true contradictions. As such, the “explosion” of inferences derivable from contradictions in classical logic is (vacuously) truth-preserving, and we have no good reason to reject classical logic in favor of any sort of weaker, inconsistency-tolerant (or “paraconsistent”) logical framework.

My perspective is *classical monism*, the claim that classical logic is currently our best overall theory of the world (that is to say, of which inferences are truth-preserving and hence of *what's true*), relative to the level of abstractness and generality at which formal logics operate. In what follows, I will be defending that view against the dialetheist challenge.

Debates about whether dialetheism is correct are often described as debates about the Law of Non-Contradiction, but that isn't quite right. If the Law of Non-Contradiction is simply a logical formula that tells us that for any conjunction of a statement and its

negation, the negation of that conjunction is true, there is no reason that a dialetheist should have to deny this in order to be a dialetheist, so long as they hold that some such conjunctions are *also true*. It is simply the case that for any α such that $(\alpha \wedge \neg\alpha)$ is a true contradiction, $[(\alpha \wedge \neg\alpha) \wedge \neg(\alpha \wedge \neg\alpha)]$ is *also* a true contradiction.¹

As such, we might do better to think of the orthodox logical position that the dialetheist is challenging not as “the Law of Non-Contradiction,” but as “monaletheism.”² Just as dialetheism is the claim that a statement can have both of the values “true” and “false,” monaletheism is the claim that no statement can have more than one of them at a time.³

Note first of all that monaletheism--like dialetheism--is neutral about whether there can be “truth-value gaps,” statements that are neither true nor false.⁴ (We will return

¹ More to the point, not only *could* the dialetheist take this position on the LNC (after all, they *could* take it on anything!), but important dialetheists *do* take exactly this stand, adopting the LNC as a logical truth in their preferred logics at the same time as they assert that it has true exceptions. Interesting alternate suggestions that might also do the expressive work for which I’m adopting “monaletheism” are ‘the LNC taken as a metaphysical principle’ and ‘the rationality LNC.’ The former is due to Takho (2009), pp. 32-47, and the latter is due to Beall (2009), p. 101. In the former case, I decline to use it (despite a sense that Takho is getting at much the same idea that I am) because Takho means only to rule out Graham Priest’s “metaphysical dialetheism,” whereas I also want to rule out positions like Edwin Mares’ “semantic dialetheism” and Beall’s version of dialetheism (according to which there would be no true contradictions if we hadn’t enhanced our language with a truth predicate as a convenient expressive tool for making certain sorts of generalizations) as well. In the latter case, my hesitation is due to the fact that the ‘Rationality LNC’ is expressed in acceptance/rejection talk, and (as discussed in Chapter Seven, when we to a revenge paradox about rejection), I don’t think that, given the Liar reasoning Beall endorses in other cases, acceptance/rejection talk can do the work in clarifying different positions on truth and paradox that Beall thinks it does.

² Thanks to Ryan Lake for coining this word, in conversation.

³ The issue of whether, as some proponents of the A-Theory of Time argue, propositions change truth-values over time, or, as B-Theorists have it, they have eternal and unchanging truth-values, is a separate issue that need not concern us here.

⁴ Even if some very prominent dialetheists like Priest reject the possibility of gaps that are not gluts, that doesn’t mean that dialetheism itself is incompatible with gap theory. JC Beall, for example, has outlined a ‘speckled’ theory of truth according to which (a) there is a meaningful difference between gaps and gluts, and (b) both kinds of statements exist. He has two kinds of negation, ‘choice’ and ‘exclusion,’ such that Liars constructed using one end up being gappy and those constructed using the other are glutty. See, for example, Beall (2005), pp. 7-22.

to this issue later.) Classical logic is wrong if monaletheism is wrong, but so is, for example, intuitionist logic.

To frame the issue differently, the correctness of monaletheism is a necessary condition for the correctness of classical monism, but it isn't a sufficient one. After all, these days, challenges to classical logic come in all shapes and sizes. Free logicians argue that classical logic is wrong because it's bound up with a false theory of reference, quantum logicians argue that classical logic is wrong because, when it comes to the properties of subatomic particles, conjunction and disjunction don't distribute the way classical logic says they should, and so on.

Even specifically *paraconsistent* challenges to classical logic come in a variety of forms. For example, relevance logicians argue that truth-preservation is insufficient for validity. Even if α cannot be true without β being true, α can, according to the relevance logician, fail to entail β because they do not have a non-logical term in common, or because α didn't really do any work in the classical inference from α to β (consider, for example, the derivability of logical tautologies from any and every premise in classical logic), or because the inference failed to meet some other relevance constraint. (The exact details of such constraints vary from logic to logic.) In such logics, for obvious reasons, explosion fails, but not because it has really existing counter-examples.

Other forms of paraconsistency get much more radical than that. After all, standard presentations of relevance concerns seem to at least assume logical monism, in so far as the line of thought is that classical logic should be rejected because it gets the notion of β following from α *wrong*. Other paraconsistentists (e.g. many of those associated with the "Brazilian School" of paraconsistency) start from a starkly anti-realist

and pluralist picture of logic, whereby logics are merely the formalizations of certain patterns of language or thought, or certain useful reasoning patterns, where “useful” varies from context to context and perhaps also from purpose to purpose. As such, the fact that a weaker sort of consequence relation, whereby contradictions entail some things but not others, serves various useful functions is a good enough reason to change logics (at least for some contexts). For example, standard lists include counterpossible reasoning (“there are no true contradictions, but what if there were?”), reasoning about works of fiction that contain inconsistencies, reasoning about interesting but inconsistent scientific or mathematical theories (e.g. Bohr’s theory of the atom, or the original formulation of the calculus), and artificial intelligence contexts where we want an expert system to draw new conclusion from a computerized database that will, inevitably (given human error and disagreement) contain certain inconsistencies.

Now, I agree in all cases that these are domains in which it’s both possible and useful to (non-trivially) reason about inconsistent situations, but I disagree that we should change logics in order to do so. We already reason counterpossibly all the time. In fact, it’s hard to see how the practice of philosophy would be possible if we couldn’t engage in such reasoning, since in all sorts of key philosophical debates, both sides frequently think the other side not only *is* wrong but *couldn’t* be right. Even so, to debate them, it’s necessary to draw out specific implausible consequences of the opposition’s views, and we do so as if those views had *some* commitments and not others.

However, as Daniel Nolan has convincingly argued, we don’t need to reject classical logic to make sense of this.⁵ We just need to extend it with a (non-trivial) counterpossible conditional connective, and we can even have the rules for this

⁵ See Nolan (1997) for a detailed explication of this view.

connective almost exactly imitate a more orthodox theory of counterfactuals. Instead of saying that counterfactual conditionals are true iff the closest possible worlds at which the antecedent is true are also worlds at which the consequent is true, we can say that they are true iff the closest *worlds* (whether possible or impossible) at which the antecedent is true are also worlds at which the consequent is true. (Presumably, possible worlds are always closer than impossible ones.) There may be difficulties cashing out the right notion of “closeness” here, but it’s not clear that this challenge is any more formidable for impossible worlds than for possible ones.

In the spirit of Nolan’s proposal, I would agree that works of fiction are frequently inconsistent, and that we can and should non-trivially reason about them, but I’d say that far from showing that α and $\neg\alpha$ don’t jointly entail β , this merely shows that $F\alpha$ and $F\neg\alpha$ don’t entail $F\beta$. Just as Sherlock Holmes does not exist, but he exists *in fiction*, explosion is valid (because there are no true contradictions), but it fails *in fiction*, because works of fiction sometimes portray inconsistent situations as being true. A great many things are false, but portrayed as true in works of fiction, and that’s not a conceptual distinction to be passed over lightly.⁶

⁶ Graham Priest has illustrated the possibility of inconsistent fiction nicely with his explicitly inconsistent story *Sylvan’s Box*, in which the belongings of the late Richard (Routley) Sylvan are found to include a box that’s simultaneously empty and non-empty. In this case, pretty clearly, no ad hoc attempt to break the story into maximally consistent chunks and reason about them separately, or to charitably interpret away the inconsistency, or anything of the kind, would be remotely helpful. Clearly, we have an outright inconsistent body of information that we’re quite capable of reasoning about, and relative to which it would be absurd to conclude everything. However, if Priest takes this to establish the stronger claim that any α and any $\neg\alpha$ fail to jointly entail any and every β rather than the weaker claim that $F\alpha$ and $F\neg\alpha$ fail to jointly entail any and every $F\beta$, then he faces the following problem: As we’ll see in Chapter Four, Priest argues (quite convincingly, by my lights) that truth-value gaps are impossible. If there’s no fact that makes α true, then *this* is fact is quite sufficient to make α false, and the claim that it’s neither entails that it’s both in any case. However, fictional worlds seem to be quite obviously incomplete in a way that the real world is not. Take the claim that Sherlock Holmes’ brother Mycroft once wrote a paper anticipating Russell’s Paradox, but that he never published it. This is an obviously meaningful, declarative statement, but what on earth could one ever cite as justifying the claim that this is true in the world of the Holmes stories, or that it is false in that world? The fact that we are unable to decide the point, moreover, seems pretty clearly to

Similarly with interesting historical examples of inconsistent theories that it seemed rational at one time to accept. The notion of “accept” at work here needs to be severely disambiguated. Accept *as true*, or accept *as useful*? Of course, someone may accept, e.g. Bohr’s theory of the atom as *true* while (incorrectly) assuming its

have nothing to do with a lack of epistemic access to the relevant information, but rather with the fact that the relevant information does not exist. As such, it seems to me that the right “internal logic” of our fiction operators should allow for both gaps and gluts to be true *in fiction*, and for there to be a distinction between them. I don’t think it follows from this that Priest is wrong about truth value gaps. Rather, just as it’s false that a detective ever lived on 221B Baker Street but one did *in fiction*, there are no truth-value gaps, but there are in the worlds of various works of fiction. The same goes for contradictions and explosion. If one wants to describe my stance on this matter as a sort of weak logical pluralism, given that I’m using one logic to reason about what’s actually the case another to reason about what’s the case *in fiction*, I don’t have a problem with that, although (a) it’s important to note that “pluralism” in this sense is completely compatible with the extreme version of logical monism that I endorse, according to which validity is truth-preservation, and classical logic is our best theory of truth-preservation, and thus our best theory of *what’s true*, and (b) I’m a bit concerned that talk of reasoning “using a logic” might over-state the role of logic in reasoning. I agree with Harman (1986) that it would be absurd to insist that we have some sort of rational duty to incessantly form beliefs in all of the logical consequences of our current beliefs. People have better things to do with their lives—climb mountains, solve math problems, have sex, drink good whiskey, contemplate the heat death of the universe and so on—than sit around working out consequences to their belief sets, and even if they didn’t, there wouldn’t be enough time in a finite human lifetime to work them all out. Moreover, when our current beliefs have absurd logical consequences, surely there’s a rational norm *against* adding those consequences to one’s belief set and *for* going back and revising the belief that generated them. (The rational norm *against* coming to believe everything because you have inconsistent beliefs, and *for* changing your views when you realize that they’re inconsistent, is a special case of this.) Of course, Harman sometimes talks as if the laws of logic were simply *irrelevant* to the norms of good reasoning, and I don’t agree with that, but specifying plausible bridge principles between the two is harder than it looks. (See MacFarlane (2004), for one philosopher’s attempt to wrestle with this problem.) I more or less agree with Streumer (2007)’s weak principle. To re-phrase it slightly, in an even weaker (and thus, I think, more plausible) way, *if* you have a set of beliefs S, and, having thought about it, you *know* that some further claim C is a logical consequence of S (i.e. that the claims in S can’t all be true without C being true as well), there is a rational norm against simultaneously maintaining your belief in all the claims in S and actively disbelieving (or having a conscious stance of agnosticism about) C. Of course, this is defeasible in the face of various pragmatic considerations. It’s not irrational to try to find your car keys before putting any mental energy into deciding whether to reject your belief that you have hands in light of your stance about skepticism about the external world or vice versa. Note that this does mean that it’s always irrational to knowingly continue to believe inconsistent things. Interestingly enough, some (non-dialetheist) philosophers have rejected this. For example, Penelope Maddy argues that, although contradictions are never true, the Lottery and Preface Paradoxes show that it’s sometimes rational to inconsistent beliefs. (Maddy (2007), p. 295) Against this, I agree with Evnine (2008)’s analysis. Evnine argues that, given the irrationality of knowingly maintaining inconsistent beliefs, what these paradoxes actually show is that we shouldn’t run together the quite separate questions of *probability* and *evidential justification*. (In fact, I think Evnine’s view has a happy consequence for those who, like me, want to believe our best current science even in the face of the pessimistic meta-induction. *Even if* the historical track record makes the refutation of our current theories extremely likely, it’s *still* the case that we’re justified in believing those theories on the basis of the evidence.) One might think this stance on probability gives me an easy avenue of response to Priest (2006a)’s probabilistic argument for his “classical re-capture,” which we’ll consider in Chapter Seven. However, for the sake of presenting the strongest possible argument, I’ll avoid relying on any such deeply controversial claims about probability theory in my reply to Priest.

consistency, and fail to accept *absolutely everything*, but we hardly need to retroactively impute implicit paraconsistency to them in order to make sense of that posture, since someone who doesn't realize the presence of an inconsistency in their belief set is hardly likely to derive arbitrary results from the relevant contradiction. (Indeed, arguably, we're all in this position all the time, in various ways. One plausible way to think about the process of rational belief revision in general is to see it as a never-ending clean-up of various inconsistencies in our belief sets as we constantly discover new ones.) The interesting question is: what does the theorist do *when* they realize that some piece of accepted mathematical or scientific theory is inconsistent? Do they slide over to dialetheism by continuing to accept it *as true* or do they reject it and find a new theory?⁷

Of course, there may be a lag time in which they regard the theory as false (because inconsistent) but in which they haven't found a suitable replacement yet. During this lag time, they may regard the theory as, although not strictly speaking true, perhaps as being as close to the truth as it's possible to come at the moment, and something which, due to its predictive power and so in, it's instrumentally useful to continue to treat as if it were true. We can, if you'd like, even formalize this with a usefulness operator U , where $U\alpha$ is read as " α is as close to the truth as we're currently capable of getting, and it's instrumentally useful to treat α as if it were the case in certain sorts of contexts." As with fiction, $U\alpha$ and $U\neg\alpha$ won't generate just any and every $U\beta$, for the simple reason

⁷ Mark Colyvan has argued that, at least in one historical case, this would have actually been the correct move. (See Colyvan (2008), pp. 24-35.) Given Quinean indispensability considerations and the historic reliance of Newtonian physics on an inconsistent formulation of the calculus, it was once rational for people to accept the existence of inconsistent mathematical objects, although it isn't any more. Although I accept Quine's indispensability argument for mathematical realism, its historical application here seems importantly non-obvious to me. Indispensability considerations only give us a good reason to believe in the objects necessary to make sense of theories that we have good reason to believe are *true*, and the verdict of history in this case has come down decisively on the side of those who took the inconsistencies implicit in the original formulation of the calculus as excellent evidence that the theory (as stated) was *false*.

that $U\alpha$ and $U\neg\alpha$ can be simultaneously true, so not all inferences from them are truth-preserving. As with the fiction case, far from the introduction of this operator being a matter of *ad hoc* formal tinkering to get around the necessity of changing logics, there's an absolutely crucial conceptual distinction between what someone is claiming when they assert α and what they are claiming when they assert $U\alpha$, and it's one worth calling attention to with formalization.⁸

Of course, the artificial intelligence case sketched above might be a case where it's not only valuable to reason non-trivially about inconsistencies, but it's useful to do so *with paraconsistent logic*. After all, in any massive database of information, with data being inputted from all sorts of different expert sources, inconsistencies will be inevitable, not because the world is inconsistent, but because of the inevitability of human error and disagreement. In any such case, at least one of the data points involved in the conflict will be *false*, but the computer has no way of adjudicating between the competing claims. Moreover, the whole point of programming a computer to reason about what follows from all this information is that no *human* is in a position to sort through all the information to decide which bits are wrong, for the simple reason that there is too much of it, so no human can intervene to sort it all out.

⁸ Again, in seeing whether $U\alpha$ implies some $U\beta$, it might be useful at certain junctures to see whether the relevant α entails the relevant β according to some paraconsistent logic. If one wants to call this logical pluralism, that's fine with me, relative to the two constraints noted before—that it's a kind of pluralism consistent with my (extremely strong version of) monism, and that I take talk of “reasoning using a logic” to be somewhat exaggerated in any case. That said, in the limited sense that the laws of logic do impinge norms of good reasoning, I wouldn't deny that, when reasoning “within the scope of the operator,” a non-classical logic can sometimes play a roughly analogous role. (Certainly, this much is true when reasoning about impossible worlds for the sake of deciding which counterpossible conditionals are true.) When we extend classical logic with any operator O such that $O\alpha$ and $O\neg\alpha$ can both be true, the rules for circumstances for $O\alpha$ entailing $O\beta$ may reasonably be taken straight from the rules about α entailing β according to some paraconsistent logic, and when we extend classical logic with an operator O^* such that $O^*\alpha$ and $O^*\neg\alpha$ can simultaneously fail, it may be useful to copy the rules for $O^*\alpha$ entailing $O^*\beta$ from the rules for α entailing β according to some gappy or “paracomplete” logic.

If we simply program the AI to reserve judgment in cases where data points conflict, it will be worthless to us as a source of further information. What if one data point involves a claim incompatible with thousands of others?

Of course, if the computer takes all of the data points to be true and sees what (non-trivially) follows, some of the information it is working from will be false, so falsehoods might creep into its conclusions. This, however, would most likely be true even for a perfectly consistent database, and there's a sense in which the worry misses the point. We can gain useful information from a system without that system being infallible.

That said, the fact that it's sometimes useful to program computers with what is, strictly speaking, a false logical theory (paraconsistency) is no more of a threat to classical monism than the fact that computer systems used to predict weather patterns are equipped with an approximately Newtonian understanding of physics is a threat to our "monism" about the correctness of the Special Theory of Relativity. Sometimes false theories serve useful purposes.

Of course, if we take what I've been saying about these cases as not only an explanation of why the classical monist need not be troubled by these examples, but as an argument *against* the opposition, the sort of anti-realist, pluralist paraconsistentist who rejects explosion on the basis of these considerations (even though they don't think explosion has counter-examples) would be well within their rights to find all of my arguments spectacularly unconvincing. This is because they start from a gestalt picture of *what sort of thing logics are* that is in stark opposition to mine. Since they don't necessarily regard logics as being about truth or truth-preservation, but merely as being formal captures of certain useful patterns of reasoning, the distinctions to which I've been

appealing—between being accurate and being *merely instrumentally useful*, between actually true and true-according-to-the-relevant-context, and so on—would simply fall flat for them.

When I say that classical logic is, at the level of abstractness and generality that formal languages operate on, our best overall theory of the world, I'm assuming a traditional view of the sort of thing logics are that's inherited from Frege. (Who was, appropriately enough, one of the founding fathers of the system that I follow general practice in labeling as "classical logic," although some non-classical logicians may feel not unreasonably non-plussed by this usage, given "classical" logic's deep incompatibility with the older, Aristotelian logic.) As he emphasized in his polemics against psychologism, Frege regarded the laws of logic as not just the laws of thought, but as "the science of the most general laws of truth."⁹ (This, by the way, nicely accounts for the normative force of logical inferences. If rationality is, as much contemporary naturalized epistemology has it, essentially an instrumental matter of reasoning in ways that will get us to the goal of truth, and if logic is essentially the science of truth-preservation, then we can see why it's rational to accept the logical consequences of views that one takes to be true.¹⁰) On this traditional Fregean view, there are *logical truths* (every sentence is either true or false, none of them are both, if the world makes some disjunction true and one of its disjuncts false, it makes the other one true, and so on), and these truths are informative not just about language or thought or certain useful patterns of reasoning, but about external reality.

⁹ Frege (1979), p. 128

¹⁰ The traditional objection that, if validity is truth-preservation, we can never add to our stock of information through inferring logically valid conclusions from it, only goes through if we are all omniscient about matters of truth-preservation.

Logic is concerned with the laws of truth, not with the laws of holding something to be true, not with the question of how men think, but with the question of how they must think if they are not to miss the truth.¹¹

Of course, in endorsing this strand of Frege's view of logic, I don't mean to endorse the full philosophical details of the way he cashed out these general claims.¹² Certainly, in using the Quinean formulation "best overall theory of the world," I mean to be signal my acceptance of the claim (which there might be reason to doubt that Frege would have accepted¹³) that our knowledge of these "general laws of truth" (or equivalently, given deflationary considerations about truth, "general laws of what is the case") is fallible and constantly open to revision in the light of new evidence or new arguments. In fact, the Quinean picture of our knowledge of these matters seems exactly right to me. Any given piece of evidence can always be made compatible with any of our existing beliefs, if we are willing to revise away enough other beliefs to clear the way.¹⁴

The issue, in any given case, is whether or not it is rational to do so, and this is a complicated and nuanced question that cannot be answered in any completely general

¹¹ Frege (1979), p. 149

¹² I don't, in endorsing what Frege says in *these specific passages*, mean to endorse any particular larger element of Frege's picture. I certainly don't believe, for example, that logic concerns itself with any sort of abstract "third realm." (I think logical truths ultimately rest on perfectly ordinary facts about tables, chairs, dogs, cats, electrons, protons and so on. Also, of course *if* any non-physical objects exist—for example, abstract mathematical objects—then they too play their role in making logical truths true. Logical truths are, in other words, simply ordinary truths at a high level of generality.) The precise interpretation of Frege's views on some of these issues is controversial even among Frege scholars in any case, and I have no intention of wading into such historical controversies here.

¹³ Whether or not Frege's notion of analyticity would have put him at odds with my views about the revisability of logic is another historical issue that I don't plan to engage with here. In fact, the whole question of analyticity is one that can be neatly side-stepped for our purposes. While I'm personally unsympathetic to the claim that there's any interesting sense in which "analytic truths" can be distinguished from "synthetic" ones, some theorists endorse such a distinction at the same time as agreeing with more or less everything I say here about our knowledge of the laws of logic. For example, Graham Priest (1979) simultaneously endorses the analytic/synthetic distinction *and* the "web of belief" model. Whether or not this adds up to a coherent position is an issue for another time.

¹⁴ Obviously claims like "everything is open to revision," "overall theories of the world should be revised in some way when they meet conflicting evidence," "considerations like simplicity and nonad hocness should play a role in theory change" and other statements with which one might express this Quinean epistemic picture aren't exceptions to the general picture. All of these claims are fallibly believed and open to revision.

way. For example, rather than it being the case that we should *always* revise our empirical theories to fit with what we take the laws of logic to be, or that we should *always* revise our beliefs about what the laws of logic are to fit with challenging new empirical evidence, the question of which solution is most plausible in a given context is a difficult determination that has to be fought out on a case-by-case basis. Considerations such as simplicity and non-adhocness will come into play. In general, however, we should be very *wary* and *conservative* about revising our ideas about the general laws of truth, because those tend to sit terribly close to the center of our web of belief, and tinkering with them can have drastic repercussions throughout the rest of the web. (Indeed, this is precisely the sort of argument I'll be making against dialetheism in Chapter Eight.) That said, if our logical beliefs are going to be anything more than unquestioned dogmas believed out of dumb habit, serious challenges to them have to be taken seriously. Sometimes our web of belief *needs* to be shaken up in drastic ways, because it was constructed along false and refuted lines. For example, the belief set of any given participant in the debates with which I concern myself in the present work has relatively little in common with the belief set of a heresy-squelching official of the medieval church, and that's a very good thing.

Now, arguing for the overall Frege/Quine picture of logical truth and truth-preservation against the anti-realist, pluralist picture held by the sort of non-dialetheist paraconsistentist we've been discussing, and against the relevance logician's concerns about the sufficiency of truth-preservation, would be a book-length project in and of itself. (And, to be clear, it would certainly be a worthwhile one. Someone should write

that book.) My project here is a narrower and more humble one. I just want to defend classical orthodoxy against the specific challenge posed by the dialetheist.

The dialetheist challenge arises from *within* the picture of the nature of logic and our epistemic access to it that I'm assuming. Dialetheists like Graham Priest don't deny that validity is best understood as truth-preservation. They accept that analysis, but argue that explosion has counter-examples, and that it is thus not truth-preserving. Far from seeing logic-change as a contextual matter dictated by pragmatic considerations, they agree that adopting a paraconsistent account of logical consequence involves revising our web of belief (although they deny the drastic repercussions). They simply regard such a revision as correct and justified.¹⁵

However, some towering figures in the history of recent "analytic" philosophy, such as David Lewis and (ironically enough, given the preceding discussion) Quine, have questioned whether even that narrow debate is possible or meaningful. My (anecdotal) impression is that such doubts about the debate are absolutely endemic among their contemporary colleagues.

One worry is the kind that tends to be expressed by furrowing one's brow, adjusting one's glasses, and earnestly asking, "but...I don't understand...what could be more basic than *the Law of Non-Contradiction*?" A closely related worry, perhaps even a different way of expressing the same underlying thought, is the concern that it's

¹⁵ On Priest's acceptance of the web of belief model, see Priest (1979). On his acceptance of the truth-preservation model of validity, see section 11.11 of Priest (2006b). (I do note that, in concession to a point made by Hartry Field, Priest has recently revised his position and decided that validity can't be truth-preservation in the most obvious sense, which is to say in a sense expressed by conditionals. He does still regard validity as being, in some important sense, about truth-preservation.) On the issue of drastic repercussions—i.e. Priest's "classical re-capture" and why I think that it fails—see Chapter Seven, below. Of course, other dialetheists may disagree with Priest on some or even all of these matters, but the point is that dialetheism is at least *compatible* with the Frege/Quine picture, and that I am begging no questions against it by assuming that picture while I argue for monaletheism.

impossible to score dialectical points against dialetheists, because they always have the option of embracing the relevant contradiction. (“OK, you make a good point about α , and maybe I do have to accept it, but *I also accept* $\neg\alpha$! Ha-ha! I still win!”) More seriously, some might worry that treating the dialetheist position that some but not all contradictions are true seriously enough to engage with intricate arguments for and against it means disregarding a fairly obvious logical truth, which is that if some contradictions are true, everything is true and reasoning is impossible. Finally, and perhaps most disturbingly, when we move from the sphere of vague worries to rigorous objections, there’s Quine’s claim that the classical monaletheist (who believes that there are no true contradictions, and that explosion is valid) and the paraconsistent dialetheist (who believes that some contradictions are true, and that they entail some things but not others) are simply talking past one another, since they don’t mean the same thing by their use of the negation operator.

Chapter Two will be devoted to taking on these worries one by one, first dealing with the bundle of worries related to the *basicness* of logical truth, then taking a hard look at the explosion proof and how it’s possible to accept its validity but still take the dialetheist challenge seriously, and finally addressing Quine’s concern about negation. Having (I hope) established that it is possible to argue with dialetheists, in Chapter Three, we will move on to the actual business of doing so, addressing a few important secondary arguments for true contradictions. In Chapter Four, we will move on to the central argument for dialetheism, which is the argument from the Liar and related semantic paradoxes. We will stay there all the way through Chapter Seven, devoting four chapters to showing why standard solutions fall flat in this context, exploring (and ultimately

rejecting) the “truth-value gap” proposal, and then adapting the most plausible elements of gap theory to a new solution that retains classical logic (and indeed classical monism) and fails to fall prey to standard problems. In Chapter Eight, we’ll move on to advancing positive arguments *against* dialetheism, showing that the dialetheist approach to the Liar paradox (biting the bullet) causes insuperable difficulties when we come to Curry’s Paradox, and showing that unweaving monaletheism from our web of belief causes all sorts of important and basic things to undesirably unravel. Finally, in Chapter Nine, we’ll advance an overall monaletheist (and gap-intolerant) view of negation, and take stock of where that leaves us.

At the outset, it’s worth clarifying certain parameters of success for our project. After all, the most important motivation for dialetheism comes from the argument from the Liar and related semantic paradoxes, and there have already been a great many book-length attempts to provide consistent solutions to those paradoxes. These solutions, however, tend to suffer from one or more of the following series of problems:

- (a) They fail to take seriously the idea that the paradoxes constitute *arguments for dialetheism*, and as such fail to shoulder their rational obligation to avoid begging the question.
- (b) They ignore all arguments for dialetheism *other* than the arguments from the semantic paradoxes. They ignore, for example, the arguments from the paradoxes of motion and change, from inconsistent obligations and from naïve set theory. Thus, even if they *did* succeed in consistently solving the *semantic* paradoxes, they would still fail to block the total case for dialetheism.

- (c) Even as solutions to the semantic paradoxes, they ultimately fail to save consistency, because they are vulnerable to various strengthened and revenge paradoxes, whereby the very concepts used in the solution are used to forge new and more virulent paradoxes.
- (d) They purchase consistency only at the expense of expressive or inferential power. In other words, they either require us to abandon the full resources of natural languages in favor of new, artificial languages or regimented fragments of natural languages in which paradoxical sentences cannot be expressed, but neither can at least some completely innocuous, meaningful and non-paradoxical sentences, or they require us to change logics, rejecting classical logic in favor of some gappy, “paracomplete” or otherwise non-classical (and, invariably, inferentially weaker) alternative.

If I can avoid all four of these pitfalls, and address the motivations for dialetheism (primarily, but not *only*, the semantic paradoxes) without begging the question, or falling prey to revenge paradoxes, or jettisoning expressive resources, or abandoning classical logic, this work will have thereby justified its existence even in a landscape already cluttered with competing solutions to the semantic paradoxes. We will return to this check-list at the end of Chapter Nine. Meanwhile, to even embark on this project, we need to establish the possibility of the debate.

Chapter Two: The Very Possibility Of The Debate

Some philosophers claim not to understand how it is possible to argue about whether we should change our ideas about logic. What, they ask, could be more fundamental than logical truths, on the basis of which we can justify our belief in those truths? Moreover, when we argue about logic, aren't we necessarily using logic (to see which arguments are good and which are bad), and thus assuming a position on the disputed question? A few of these worries are packaged together nicely in David Lewis' response to an invitation to contribute an essay to an anthology debating the Law of Non-Contradiction.

I'm sorry; I decline... My feeling is that since this debate instantly reaches deadlock, there's really nothing much to say about it. To conduct a debate, one needs common ground; principles in dispute cannot of course fairly be used as common ground; and in this case, the principles *not* in dispute are so very much less certain than non-contradiction itself that it matters little whether or not a successful defense of non-contradiction could be based on them.¹⁶

A preliminary point is that this line of argument should give no aid and comfort whatsoever to the classical logician. If the alleged impossibility of debating logic somehow gives him permission to remain dogmatically attached to classical logic, it should give the paraconsistent logician precisely the same permission to remain dogmatically attached to her logic of choice. All we can do, if the argument against the possibility of the debate goes through, is stand behind our respective barricades and glare at each other.

¹⁶ Lewis (2004), p. 176

Fortunately, I do not believe that this is all we can do. On the first point, I would argue that the “what could be more basic or certain than logical truths?” question illegitimately smuggles extremely dubious foundationalist epistemic assumptions into the philosophy of logic.¹⁷ Why, after all, should we assume that our beliefs must be justified by reference to more fundamental beliefs? From Descartes onwards, this epistemic program has faced what so far look like insuperable difficulties. The narrower your class of foundational, allegedly self-evident beliefs, the fewer interesting beliefs of the kind that we all ordinarily take to be cases of knowledge can be derived from them. (Of course, you can always bite the bullets, stamping your foot in each case and insisting that the various *prima facie* cases of knowledge that turn out not to be knowledge on your theory really aren’t, because they aren’t derivable from the class of foundational beliefs, but the more of that you do, the less convincing your account of knowledge is going to be to anyone who doesn’t already agree with you.) Conversely, the broader that class of allegedly foundational beliefs, the less plausible it becomes that they really are all basic and self-evident.¹⁸

As to the second and more troubling question, I think it is useful to look at how the debate proceeds in practice for clues for how it is possible for it to proceed. The

¹⁷ It could be objected that the claim that all beliefs must be justified on the basis of more fundamental beliefs isn’t unique to foundationalism, but rather represents the overlapping consensus of foundationalism and infinitism, but the objection I’m considering really does seem to assume foundationalism rather than infinitism. After all, the idea in standard dismissals of the debate about basic logical principles is not that we’re not justified in believing logical truths because they can’t be justified by anything more basic, but rather that things don’t get more basic than logical truth, and that as such, trying to have an argument about whether we should accept them or not is a misguided enterprise. In other words, logical truths sit at the justificatory foundation at the bottom of the stairs.

¹⁸ Some foundationalist views shy away from giving basic beliefs any label quite as emphatic as “self-evident,” but it’s still the case that they take the basic beliefs to be (a) justified, but (b) not on the basis of the requirements imposed on all other beliefs. Thus, I think, even for weaker versions of foundationalism, we can say that the basic beliefs have a “special epistemic status that makes it clearly right to believe them,” and the same criticism goes through—the broader the class of beliefs for which one makes this claim, the less plausible the claim becomes.

answer is that proponents of two rival logics can argue with each other using only the *shared fragment* of their systems. When someone uses an argument that is only valid in one or another of the rival logics at issue, their opponent can legitimately accuse them of begging the question. In practice, this is exactly what happens.

Moreover, this isn't nearly as restrictive as it sounds. In practice, in real cases--looking at classical logic, free logics, relevance logics, paraconsistent logics (some of which are also relevance logics and some of which are not), quantum logic, intuitionist logic and so on--the great majority of the inferences that are valid in one system and are likely to be appealed to in the course of philosophical argumentation are valid in any of other systems as well. Nor should this be surprising. A simple and appealing explanation is that we all start out with a shared stock of pre-theoretical intuitions about basic concepts like negation and validity, and about which inferences go through and which do not. Different logics are, at bottom, attempts to formally codify these intuitions. Just as our best moral theory is the one that captures as many of our moral intuitions as possible and explains away the non-captured ones as well as possible, so too for our best logic and our deepest intuitions about what follows from what.

In classical logic, anything and everything follows from any contradiction. To make things concrete, take the following example. If we start by assuming that the Russell Set ($R = \{x: x \notin x\}$) is both a member of itself and not a member of itself, we can, in a few easy steps, derive the result that John McCain is a 400-year-old vampire. After all, if the Russell Set is both a member of itself and not a member of itself, then by conjunction-elimination we derive the consequence that the Russell Set is a member of itself. From there, by Disjunction-Addition, we conclude that either the Russell Set is a

member of itself or John McCain is a 400-year-old vampire. From a second application of conjunction-elimination to the first premise, we conclude that the Russell Set is *not* a member of itself. From the previous two results and Disjunctive Syllogism, we get the startling result that John McCain is indeed a 400-year-old vampire.

Formally,

1. $\alpha \wedge \neg\alpha$ A
2. α 1, Conjunction-Elimination
3. $\alpha \vee \beta$ 2, Disjunction-Addition
4. $\neg\alpha$ 1, Conjunction-Elimination
5. β 3,4, Disjunctive Sylloigsm

Now, this proof is a sufficiently familiar one that, in one form or another, most introductory logic students are exposed to in their first course, but this only makes it more odd that, when we think about it carefully, it is terribly unclear what sort of proof it is supposed to be. It is, obviously, not supposed to be the sort of proof that functions to convince anyone of its conclusion, since the typical logic instructor showing it to her students does not herself believe that any statement of the form $(\alpha \wedge \neg\alpha)$ is true, and she believes still less that any and every possible β is true. It is often referred to as a *reductio* proof, but strictly speaking, it is nothing of the sort. A *reductio* proof that the first premise is not true would be spectacularly short and uninteresting:

1. $\alpha \wedge \neg\alpha$ A for RAA
2. $\neg(\alpha \wedge \neg\alpha)$ 1, RAA

Perhaps it is a conditional proof, to get us the result that $(\alpha \wedge \neg\alpha) \rightarrow \beta$. If so, again, it is a bit like using a sledgehammer to crack a nut. The falsehood of $(\alpha \wedge \neg\alpha)$ is alone sufficient to guarantee the truth of that conditional in classical logic.

We are running out of formal options. Turning to the informal ones, we could speculate that it is supposed to show that there are not and could not be any true statements of the form $(\alpha \wedge \neg\alpha)$ because, if any such statement were true, everything would be true, and at least some things pretty clearly fail to be true. Or, more weakly than this counterpossible claim, it could be seen as showing that anyone who believed that some such statement was true would be rationally compelled to believe absolutely everything. The problem is that if it is supposed to be an argument for either of these things, it is a terribly unconvincing one. As an argument against the possibility of true contradictions, it begs the question by employing a rule (Disjunctive Syllogism) that is only universally truth-preserving if there are no true contradictions. After all, if for some α and some β , α were both true and false, whereas β was just false, then $(\alpha \vee \beta)$ would be true, $\neg\alpha$ would be true and β would nevertheless fail to be true. (This is formalized in the semantics of Graham Priest's "Logic of Paradox" (LP) and similar systems.) For the same reason, someone who believed some statement $(\alpha \wedge \neg\alpha)$ was true would not be rationally compelled to conclude any and every random β , because they would have no rational reason to believe that Disjunctive Syllogism was a valid rule of inference.

Reflection on this point shows that, even from a classical monist perspective, "from a contradiction, anything follows" is no more interesting a logical principle than "from Hitler winning World War II, anything follows," which could be proven as follows. "If Hitler won World War II, then β " is true for any and every β , because all

conditionals with false antecedents come out as true in classical logic. Now, assume that Hitler won World War II. By Modus Ponens, we are able to conclude any and every random β .

The obvious objection is that we have switched positions midway through the argument on the subject of whether Hitler won World War II. Quite so. However, this is not a bit different in principle from the way we have switched positions on the possibility of true contradictions when we wrote down $(\alpha \wedge \neg\alpha)$ as our first premise in the explosion proof, and then proceeded to use Disjunctive Syllogism as if it were a valid rule of inference, which it would not be if the first premise was true.¹⁹

As such, while the chain of inferences connecting the Russell Set's membership conditions to John McCain's age and supernatural status is perfectly valid as far as it goes (if classical logic is right), the answer to what the proof tells us one way or the other about any remotely philosophically interesting subject whatsoever is, I think, "not a damned thing." It's valid *if* monaletheism is correct, and by helping ourselves to explosion in order to argue against dialetheism, we would be putting the cart before the horse and egregiously begging the question. Dialetheism cannot be batted away with the blunt instrument of triviality-avoidance. It needs to be engaged on the arguments.

Nor does the worry that the dialetheist can always embrace the relevant contradiction whenever they're shown to be wrong about anything carry much weight. Just because, from the dialetheist's perspective, true contradictions are logically possible, it hardly follows that any particular contradiction is particularly plausible even from that

¹⁹ Of course, the Hitler argument, while valid, is not sound, since the second premise is false. This does not, however, get us to a relevant differences between the cases, since (according to the classical logician) *no argument from a contradiction to everything*, and indeed no argument from any contradiction to any conclusion whatsoever, is *ever sound* (although all such arguments are valid), because contradictions are never true.

perspective. Consistency is not the only theoretical virtue. Moreover, any philosopher, arguing for any position on any matter whatsoever, is always “free” in some sense to get around all sorts of objections by embracing ad hoc (but logically possible) implausibilities. (Faced with a Galileo, one may always add some epicycles to one’s geocentric model instead of giving up and accepting the heliocentric alternative.) Doing that isn’t a particularly good way to earn converts to your position, and the more of it you have to do, the less plausible your original position becomes. In practice, dialetheists like Priest (or, say, JC Beall or Edwin Mares) haven’t acted in anything like the way some have worried that they could, but if they *were* reduced to the practice of “embracing the relevant contradiction” every time someone scored a point against them, this would be a great dialectical *defeat* for them.

Of course, even if the objections just considered did go through, all they would establish would be that arguing with dialetheists is fruitless. In a famous passage in *The Philosophy Of Logic*, Quine goes a step further and argues that it isn’t even possible.

To turn to a popular extravaganza, what is someone were to reject the law of non-contradiction and so accept an occasional sentence and its negation as both true? An answer one hears is that this would vitiate all science. Any conjunction of the form ‘p.~p’ logically implies every sentence whatever; therefore acceptance of one sentence and its negation as true would commit us to accepting every sentence as true, and thus as forfeiting all distinction between true and false.

In answer to this answer, one hears that such full-width trivialisation could perhaps be staved off by making compensatory adjustments to block this indiscriminate deducibility of all sentences from an inconsistency. Perhaps, it is suggested, we can so rig our new logic that it will isolate its contradictions and contain them.

My view of the dialogue is that neither party knows what he is talking about. They think they are talking about negation, “~”, ‘not’; but surely the notion ceased to be recognisable as negation when they took to regarding some conjunctions of the form ‘p.~p’ as true, and stopped regarding such sentences as implying all others. Here, evidently, is the deviant logician’s predicament: when he tries to deny the doctrine, he only changes the subject.²⁰

²⁰ Quine (1970), p. 81

Now, the language of “rigging” seems hopelessly unfair to the paraconsistentist, at least the dialetheist paraconsistentist who Quine is discussing, the one who regards “some conjunctions of the form ‘ $p \sim p$ ’ as true...” As we saw in the previous section, the failure of someone who believes in the existence of true contradictions to accept Disjunctive Syllogism, and hence explosion, is not be a matter of ad hoc “rigging,” but a conceptually obvious, principled consequence of their view.

Perhaps, to be charitable to Quine, we can read “rigging” as a sloppy word choice, and assume that the connotations of ad hoc arbitrariness are not intended. Even so, he could not be making a more severe objection to paraconsistentism. By analogy, Quine’s objection to paraconsistentism is less like the standard-issue atheist objections to the rationality of religious belief (that there is no evidence that God exists, or that the theist cannot solve the Problem of Evil) than it is like the logical positivists’ claim that statements of religious belief are literally nonsensical and meaningless.²¹ One could imagine the following dialogue between an adamant Quinean (let’s call him Bill) and a paraconsistent logician (let’s call him Bob):

Bob: I reject the principle of *ex contradictione quodlibet*.

Bill: No you don’t.

Bob: Excuse me?

Bill: You think you do, but you don’t.

Bob: Well, I accept certain claims and their negations, but I don’t accept

everything.

²¹ Of course, Quine is ‘just’ accusing the paraconsistentist of changing the subject, not accusing her of talking nonsense, so the analogy is not exact, but the point is that in each case, the theist or the paraconsistentist respectively aren’t taken to even be *contradicting* the positivist or the classical logician, and the possibility of their doing so is simply ruled out.

Bill: You're using the word "negation" to mean something different from what it means in the context of the claim you think you're rejecting.

Bob: How do you know I don't mean what you mean?

Bill: If you meant what I meant by "negation," you wouldn't disagree with me about *ex contradictione quodlibet*.

Given the severity of Quine's critique, one would expect that, whatever *else* they disagreed on, all paraconsistent theorists would at least agree that Quine is wrong about negation. They mean the same thing by " \neg " that Quine (and everyone else) does. They simply disagree with him about whether every instance of $(\alpha \wedge \neg\alpha)$ entails every β .

Oddly enough, this picture of the dialectical landscape is far from accurate.

In their paper "On The Usefulness Of Paraconsistent Logic", Newton da Costa, Jean-Yves Béziau and Otavio Bueno say that,

In paraconsistent logic, we will denote by the symbol \neg , and call it negation, a connective that is not the same as classical negation. There are those who criticize such an abuse of language. Let us note, however, that it is difficult to claim that there is only one negation, let us say classical negation (which would exactly model the negation of natural language or mathematics). In the literature, the word negation and the corresponding symbols have long been used to denote different concepts...²²

Da Costa, Béziau and Bueno are not alone among paraconsistent logicians on this point. For example Diderik Batens argues that paraconsistent logic is appropriate in some contexts and classical logic is appropriate in others because "[t]he metalinguistic reach of classical negation is beyond the reach of paraconsistency..."²³ Someone who asserts the

²² Da Costa, Béziau and Bueno (2005), p. 467-8 (Note that in this passage, \sim is being used for 'classical negation' and \neg is being used for 'paraconsistent negation.')

²³ Batens (1990), p. 227

negation of something in the paraconsistent sense of negation “does not rule out the sentence that is negated and is intended not to rule this out.”²⁴

In discussing Batens’ claim, Koji Tanaka claims that

The Batens objection may appear intuitive, since \sim and \neg are *prima facie* incompatible with each other...²⁵

In a footnote to this passage, Tanaka clarifies that they may be compatible in the sense that both sorts of negation can occur in the same logic, but of course, even this point *presupposes* that there are indeed two different types of negation. He goes on to respond to Batens by suggesting that,

...there is no reason why paraconsistent logicians cannot introduce into the language of the logical system an absurdity constant (or f), and let it be governed by the rule $\perp \vdash \beta$ for all β . In order to express the exclusion of α by $\neg\alpha$, we may further introduce the rule $\{\alpha, \neg\alpha\} \vdash \perp$. In this way, asserting α and $\neg\alpha$ leads to triviality... Therefore, the fact that α rules out $\neg\alpha$ can be expressed paraconsistently... [This] undermines Batens’ claim that paraconsistent logics are too weak to be applicable in some context.²⁶

This dispute within the paraconsistent family about whether paraconsistent logic is applicable to all or merely some contexts is beside the point for our present purposes. Batens thinks that he still needs classical negation for some purposes, while Tanaka believes paraconsistent negation can express the same information in those contexts if his use of it is supplemented in the right way. The important point in all of this is that Batens and Tanaka are both presupposing the correctness of Quine’s claim that once they stopped believing in the validity (or at least the validity in all contexts) of *ex*

²⁴ Ibid, p. 223

²⁵ Tanaka (2003), p. 36

²⁶ Ibid, p. 37

contradictione quodlibet, their use of the term “negation” ceased to be recognizable as negation in the old sense.²⁷

In the quoted sources, neither Quine nor Batens nor Tanaka feels the need to give us any sort of *argument* to justify the claim that paraconsistent and classical logicians mean different things by “not.” Perhaps they all take the claim to be obvious. Similarly, da Costa, Béziau and Bueno don’t offer any sort of argument or even examples in support of their claim that natural language and mathematics contain multiple sense of “not,” one of which is captured by “paraconsistent negation” and another one of which is captured by “classical negation.” Again, perhaps they think it is obvious.

Since it is less clear in the case of da Costa, Béziau and Bueno *when* they take natural-language and mathematical uses of “not” to be meant in the paraconsistent sense and when they take it to be meant in the classical sense, much less how we are supposed to tell from a third-person perspective which is meant in a given context, let’s start with what we can call the “Quine-Batens thesis,” that anyone who rejects *ex contradictione quodlibet* is using the word “negation” in a non-standard way, and see if (given that neither Quine nor Batens, nor Takana in his reply to Batens, provide reasons for the thesis) we can find a relatively plausible reason *why* they might think this.²⁸

²⁷ One of the many strange things about this claim is that, by parity of reasoning, it seems to follow that the contemporary classical symbolic logician doesn’t mean the same thing by ‘not’ as Aristotle or later Aristotelian categorical logicians meant by it for thousands of years. After all, not all categorical syllogisms with inconsistent premises deliver their conclusions. ‘All cats are animals, some cats are not animals, and therefore all dogs are robots’ is not a valid syllogism in Aristotelian logic.

²⁸ One could argue that the question of which concept instances of ‘not’ refer to is radically undetermined in consistent contexts, since the rules governing each operator are the same in those contexts, and that they only diverge in inconsistent contexts. (This would dovetail with the version of pluralism defended in Bueno, French and da Costa (1998) and Bueno and Shalkowski (2009), which is, to put it mildly, *much more pluralistic* than the more familiar sort of pluralism popularized by Beall and Restall (2006), and this sort of underdetermination plays a key role.) I see at least two important problems with this move. (1) Are we talking just about a purely epistemic issue of the undetermination of theory by evidence, or is it in some sense more than epistemic? Is there, in other words, a stable fact of the matter? If the former, what could possibly determine what the (radically epistemically unavailable) fact of the

One obvious thought is that they might regard the *meaning* of negation as given by which rules one takes to be valid about what can be inferred from negated statements, the way that someone might claim that in an alternate version of chess where the knight moved in a z pattern rather than an l pattern, it was not a knight. (This view, which we can call the Identity Thesis, would be roughly the equivalent in philosophy of logic to behaviorism in the philosophy of mind. The question of what the operator *is* and the question of what it *does* are, according to this view, literally the same question.) Given the assumption of the Identity Thesis, the claim that if Bob and Bill accept different rules about what follows from a statement α and its negation, regardless of the content of α , they must mean different things by “negation,” becomes trivial.

This just pushes back the question. *Why* should we identify the meaning of the term with the rules about it one takes to be valid? After all, pre-theoretically, it *seems* like the classical logician and the paraconsistentist are disagreeing with each other about

matter was about whether negation was meant paraconsistently or classically in consistent contexts? If the latter, this is a very radical position—two distinct concepts are sometimes literally the same concept, rather than just having the same consequences—and it’s unclear how it can be developed for the weak paraconsistentist without collapsing into dialetheism. Even for the dialetheist, if the Identity Thesis forces them into accepting true contradictions about identity for concepts, this seems like a steep price to pay. What about Priest’s maxim about not multiplying contradiction beyond necessity? (2) How do we differentiate consistent from inconsistent contexts? The sorts of paraconsistentists we’re dealing with for the present discussion generally aren’t dialetheists, so inconsistent context can’t mean “a context where we’re dealing with contradictions that are actually true,” or else *all contexts are consistent*. If an inconsistent context means “a context in which people sometimes say or thing inconsistent things,” then *all contexts are inconsistent*. If an inconsistent context means “a context in which it’s rational to accept inconsistencies,” then (especially given the extremely strong link between logical consequence and norms of good reasoning without which weaker-than-dialetheic paraconsistency makes no sense), classical and paraconsistent negation coming apart in inconsistent contexts would mean that, in cases where it’s rational to accept “A and not-A” and “not” is meant paraconsistently, it’s only rational to accept the paraconsistent consequences of the contradiction, whereas in cases where it’s rational to accept “A and not-A” and “not” is meant classically, it’s rational to accept the full *classical* consequences of the contradiction. But wait! The full classical consequences of the contradiction include *absolutely everything without exception*, and presumably it’s *never* rational to believe absolutely everything without exception. (At any rate, the intuition that it’s never rational to do so is presumably a key motivation for the development of paraconsistent logics!) As such, if this is how we differentiate consistent and inconsistent contexts, then negation *always* behaves paraconsistently and never classically, so there’s no reason to postulate two different concepts at work in ordinary language in the first place.

whether “the Russell Set is a member of itself” and “the Russell Set is not a member of itself” jointly entail “Hitler won World War II.” Even more severely, when the dialetheist paraconsistentist says that the Liar Sentence is both true and not true, it certainly *seems* like they are disagreeing with the classical logician who doesn’t believe that any sentence is both true and not true.

However, if the Identity Thesis (and hence Quine-Batens Thesis) is correct, neither the dialetheist nor the nondialetheist paraconsistent logician is disagreeing with the classical logician about any of this. To see why, let’s use “not” to indicate the classical sense of negation and “shnot” to mean negation in the new sense. The classical logician says that no statement of the form “A and not A” is true, and that statements of that form entail all other statements.²⁹ The dialetheist, on the other hand, believes that some statements of the form “A and shnot A” are true, and both the dialetheist and the nondialetheist paraconsistentist believe that such statements entail only some (and not all) other statements. They are not disagreeing with the classical logician about *anything*. Rather, they and the classical logician are simply talking past one another.

²⁹ One could object that, even if, on the level just mentioned, they aren’t disagreeing about anything, it’s still the case that they’re disagreeing on the level of formal rules—explosion is valid in one system, but not in other. Still, given the assumption that the negation operator is being used to denote different concepts in the respective systems, this doesn’t amount to a real disagreement. The concept denoted by the negation operator in classical logic is simply absent from paraconsistent logic and vice versa. No one regards S5 modal logic, with its expansive rules for stacking modal operators, as a *revision* of classical logic as it existed before, but as an extension of it, since if (un-extended) classical logic as it existed before the development of S5 as being in conflict with S5, since the former didn’t have any rules governing the behavior of modal operators incompatible with the rules governing modal operators in the latter or vice versa, since there *were no modal operators*, and hence no rules governing their behavior, in the former. If the Identity Thesis is correct, the same is true in this case—there are no rules governing classical negation in paraconsistent logic and no rules governing paraconsistent negation in classical logic, and hence the rules can’t contradict each other. If one argues (see the quote from Tanaka, below) that we *can* express classical negation into paraconsistent logic, then either (a) once it’s been introduced, the rules governing its behavior are compatible with the rules governing its behavior in classical logic, or (b) the rules governing its behavior are incompatible with the rules governing its behavior in classical logic. Given (a), there is still no disagreement between the classical and the paraconsistent logician. Given (b), *and the Identity Thesis*, the new negation operator that’s been introduced into paraconsistent logic alongside the old paraconsistent negation operator simply *can’t* be classical negation, and there’s *still* no disagreement between the classical and the paraconsistent logician.

This may be bad enough, but it gets worse. If the Identity Thesis is right, then there isn't *one* paraconsistent negation, but quite a few. After all, there are many different ways to block the route from contradictions to triviality, and accordingly many different paraconsistent logics. If even one rule pertaining to negation is different from one system to the next, then not only do paraconsistentists not mean the same thing by “not” when they deny the truth of the claim “the truth of any conjunction of the form ‘A and not A’ entails the truth of everything without exception” as classical logicians mean when they assert that inference (or, rather, an inference of the same syntactic form), but each individual school of paraconsistency is rejecting a different principle than all the others.

The point can be generalized. If we take the Identity Thesis seriously, no one, ever, has ever or will ever disagree with anyone else about the validity of even one rule involving negation, because this is impossible by definition. If they take different rules about negation to be valid, they are not talking about the same thing.

It gets even worse. If we take the Identity Thesis seriously, one ever has, ever will or ever could revise their own views about which rules pertaining to negation are valid, at least in the ordinary sense of “revision,” where one holds one view on one subject at time T1 and later, at time T2, repudiates their T1 view and adopts a new view *about the same subject*. Take a case where we might think this happens. Some logician Bluto is a truth-value gap theorist at Time T1. As part of his gappy logic, he rejects the validity of the inference from $\neg\neg\alpha$ to α , on the grounds that α and $\neg\alpha$ simultaneously fail to be the case in contexts where the Law of the Excluded Middle breaks down. Then, after reading Graham Priest's book *In Contradiction*, he decides to abandon his belief in truth-value gaps, on Priest's grounds that in any situation where there is no fact that makes P true, the

fact that there is no fact that makes P true is sufficient to make $\neg P$ true. Having seen the light, Bluto now cheerfully accepts that $\neg\neg P$ entails P . Intuitively, we want very badly to say that Bluto held one view about negation when he was a gap-theorist, and that he now holds a different view about *precisely the same phenomenon*, but according to the Identity Thesis, this simply can't be true. We have to say, instead, that Bluto held one view about "negation*" and now holds a different view about "negation**." Negation* and negation** cannot be the same thing, because each of their meanings are identical to the rules governing their use, and those rules are different.

This bullet is much too large to bite without fear of lead poisoning. Moreover, I know of no positive argument, advanced by anyone, *for* the Identity Thesis. Anyone who does believe it apparently takes it as an unargued given. At any rate, at least in its full literal form, it's unclear who really believes it.

To be sure, many logicians sometimes *talk* as if they accepted the Identity Thesis, for example when they say they are going to "define negation" for a system and then they proceed to lay out the basic rules involving negation which are considered valid in that system. It would be a bit much, however, to conclude from this that logicians who talk in this way actually believe the Identity Thesis (a question most of them have probably never consciously considered one way or the other), much less that their belief in it is correct or justified by good reasons. Given the extreme counter-intuitive consequence of taking it seriously, it seems better to say that anyone who does seriously advocate the Identity Thesis is mistaken, but that in the case of any particular logician who uses "defining negation for a system" talk, we should charitably interpret this as a slightly

sloppy shorthand for “defining the rules for negation in the system” or “illustrating the claims about negation embodied in the system” or some such.

Now, a weaker claim than the Identity Thesis might be that, although the question of what someone *means* by negation is distinct from the question of what rules about negation they take to be valid, the latter could be taken as *evidence* about the former. We can call this weaker claim the Evidence Thesis, and, given a sufficiently mild interpretation, it is extremely plausible. If two people take radically different rules to govern the behavior of negation, then this can certainly be (limited, defeasible and non-decisive) evidence that they mean something different by the word “negation.” For example, if Bernard asserts that $\neg\alpha$ is true iff α is false, while Bruno denies this and asserts that $\alpha\neg\beta$ (which he reads as “alpha not beta”) is true just so long as α and β are both true, we might be within our rights to suspect that Bernard means “negation” in much the same sense that the rest of us mean it, but that Bruno is using the word “negation” to describe the concept most of us refer to as “conjunction.”

The Quine-Batens Thesis is not an obvious entailment of the Evidence Thesis, as it is of the Identity Thesis, but the Evidence Thesis could be the basis for an initially plausible argument for the Quine-Batens Thesis. After all, paraconsistent and classical logicians do have a significant disagreement about what follows from statements of the form “A and not A,” and this could be evidence that those on the paraconsistent side of the fence mean something different by “not” (or, perhaps even “and”) than those on the classical side. Moreover, since we are talking about evidence rather than identity, we can take this serious disagreement as weightier evidence than the relatively minor disagreements within the paraconsistentist family, and, taking the former to be sufficient

evidence of a difference in meaning and the latter to be insufficient evidence of this, avoid some of the more extreme counter-intuitive consequences we derived from the Identity Thesis.³⁰ Believing the Quine-Batens Thesis on the basis of the Evidence Thesis, rather than the Identity Thesis, need not commit us to the absurd conclusion that no one ever disagrees with anyone about which rules governing the behavior of negation are valid, or that no one can ever revise their own views on this score.³¹

The counter-intuitiveness of the Quine-Batens Thesis, however, stands undiminished regardless of *why* we endorse it. When the dialetheist paraconsistentist says that some statements of the form $(\alpha \wedge \neg\alpha)$ are true, and the classical logician denies this claim, it *seems* like they are disagreeing with each other. When any sort of paraconsistentist denies *ex contradictione quodlibet*, it seems like they are disagreeing with the very same principle that the classical logician believes in, but that cannot be so if “paraconsistent negation” and “classical negation” are two different things rather than one concept which different sides in the dispute make different claims about.³²

³⁰ Of course, one might argue the opposite claim, that weak paraconsistentists and classical logicians mean one thing by negation, and strong paraconsistentists (i.e. dialetheists) mean another. The line of thought being considered would be that, although two logicians can agree on what negation means and disagree about whether contradictions entail everything, disagreement on the even more fundamental question of whether the truth of a statement and the truth of its negation can ever overlap is decisive evidence that the two parties in the disagreement must be talking about different subjects. It’s not obvious to me that this is any *less* plausible than the claim that *all* paraconsistentists mean one thing by negation and that classical logicians mean another, or that there’s anything to recommend one line of thought over the other about where the “change of meaning” line should be drawn. In any case, I find both stands equally implausible, and take it to be obvious that both the debate about true contradictions and the debate about explosion are both debates between two sides that *disagree with each other*, and aren’t merely talking past each other, and that it is possible on either of these issues for someone to change their minds.

³¹ That said, the counter-intuitive conclusion that no one disagrees about explosion, or changes their views about whether explosion holds, remains in full force.

³² Of course, while many non-dialetheist paraconsistentists seem to be ready to accept the Quine-Batens Thesis, I know of no dialetheists who do so. (Priest argues that classical logic embodies a false theory of negation, as we’ll see below, but that’s not quite the same thing.) Moreover, there seems like an excellent reason why no dialetheist *could* comfortably accept it. Let’s say that da Costa, Bueno and Béziau are right about natural language containing multiple negations (one of which, presumably, is “negation in the paraconsistent sense” and another of which is “negation in the classical sense”). Now, I have absolutely no idea how, on this conjecture, we are supposed to tell how any given natural-language instance of “not” is

Given that paraconsistentists tend to agree with defenders of classical orthodoxy on all sorts of things about negation—that $\neg\alpha$ is true iff α is false, that $\neg\neg\alpha$ entails α and so on³³—we can accept the Evidence Thesis, but disagree that disagreement over *ex contradictione quodlibet* is sufficient evidence that the disputants aren't referring to the same concept when each uses the word "not." Even da Costa, Bueno and Béziau, after denying that that paraconsistent and classical logicians mean the same thing by "negation" and claiming that ordinary language and mathematics contain uses of "not" that denote various different concepts, qualify their claim in an important way.

Of course, a unitary operator must have some basic properties to be called a 'negation.' For instance, no one will call the necessity operator a negation.³⁴

Well, if even da Costa, Bueno and Béziau, who take uses of the word "negation" by logicians of different stripes to denote entirely different concepts, recoil at calling *some* unitary operators negations because they don't have enough properties in common with the various negations referred to in the literature, this seems like powerful evidence that we all have a pretty good intuitive sense of when a use of the word "not" or the

meant, particularly given that the average speaker of a natural language presumably doesn't have the vocabulary or technical knowledge to let us know which sense they have in mind. (The situation is even worse with regard to instances of "not" tokened by speakers and writers who are now dead, and still worse in the case of those who spoke or wrote before the existence of any modern logical theory, classical, paraconsistent or otherwise.) Assume, however, that we have a good way of solving these problems. Now, since the dialetheist wants to accept a contradiction about the Strengthened Liar sentence—"this sentence is not true"—and use paraconsistent logic to contain it, presumably he would want to read the 'not' in the paraconsistent sense. (Even without a good way of solving these problems if the dialetheist in question is, like da Costa and Bueno, a logical pluralist who takes it to be the case that multiple logics are adequate for the same language, he could still simply say that any logic "adequate" for non-trivially reasoning about paradoxical sentences must be paraconsistent.) Fine. It seems to me, however, that this approach would face an insurmountable revenge paradox, from what we can call the Classical Liar. "This sentence is not true, in the classical sense of 'not.'" Moreover, notice that even if one abandoned the claim that ordinary language contains multiple senses of 'not' in favor of the claim that 'not' in ordinary language is *always* meant paraconsistently, but still maintained that ' \neg ' means different things in paraconsistent and classical logic, the Classical Liar Paradox still stands.

³³ Of course, there are paraconsistent logicians who disagree with one or another of these principles, but the important point for the moment is that many prominent paraconsistent logicians, such as Graham Priest, agree with both of the principles just mentioned, and with many other claims about negation made by classical logicians as well.

³⁴ Da Costa, Béziau and Bueno (2005), p. 468

symbol “ \neg ” is being used to talk about negation and when it has been re-purposed to talk about something else. Why not take that intuitive sense seriously, and, while embracing the Evidence Thesis, say that those who (in the da Costa-Bueno-Béziau terminology) are talking about something that “counts as negation” are all denoting the same concept (even if they are disagreeing about its behavior), whereas those who are talking about necessity (even if they call it negation) are not? Disagreeing on some principles, even extremely important principles, is simply not sufficient to indicate that we aren’t talking about the same thing. (If it were, no theory about any the nature of any concept could ever be *extremely misguided*, and intuitively it seems like some such theories are.) The existence of both libertarians and compatibilists doesn’t normally cause anyone to talk about multiple “free wills,” the existence of both reliabilists and internalists, coherentists and foundationalists, does not normally cause anyone to talk about multiple “knowledges,” and the existence of competing logical theories which say different things about negation should not cause us to say that there are multiple “negations.”

The reason that it *seems* like Bob and Bill are talking about the very same principle and holding different positions on it when one asserts *ex contradictione quodlibet* and the other denies it is that they really are disagreeing. The reason that it seems like dialetheists and classical logicians disagree about whether the Liar sentence is simultaneously true and not true is that they really do disagree.

Of course, even if dialetheists and other paraconsistent logicians are talking about the same thing as their opponents when they the word “not” and the symbol “ \neg ,” which is all we need to get the debate between monalethiesm and dialetheism off the ground, they certainly hold different views about it. Most obviously, dialetheists hold the very

non-standard view about negation that the truth of a statement and the truth of its negation can overlap. In the next chapter, we'll start seriously grappling with arguments for and against that view. For now, though, let's take a moment to consider an interesting, closely related possibility that we neglected in the previous discussion. Perhaps classical logic embodies a false theory of negation, such that, even if the Frege/Quine picture of logical truth is correct, and even if explosion has no counter-examples, we should *still* reject classical logic in favor of paraconsistent logic.

Graham Priest has developed just such a theory of different accounts of negation embodied in paraconsistent and classical logic, a theory that is completely independent of his commitment to dialetheism, such that even if we can avoid dialetheism, if he's right about negation, we would still be rationally compelled to adopt a paraconsistent account of logical consequence.

It is worth quoting Priest's description at some length, so there can be no doubt about precisely what he is claiming.

One may distinguish between three accounts of the relationship between negation, contradiction and content. (1) A cancellation account. According to this, $\neg\alpha$ cancels the content of α . Hence, a contradiction has *no* content. In particular then, supposing that an inference is valid when the content of the premises contains that of the conclusion, a contradiction entails nothing--or nothing with any content; it may entail another contradiction. (2) A complementation account. According to this, $\neg\alpha$ has whatever content α does not have. Hence $\alpha \wedge \neg\alpha$ has *total* content, and entails everything. (3) An intermediate account, where the content of $\neg\alpha$ is a function of the content of α , but neither of the previous kinds. According to this account, $\alpha \wedge \neg\alpha$ has, in general, *partial* content, neither null nor total. Hence, contradictions entail some things but not others.

An account of kind (3) is given in relevant and paraconsistent logics. An account of kind (2) is packed into orthodox modern logics, such as "classical" and intuitionist logic. An account of kind (1) is clearly

distinct from either of these. It appears to have been an influential account in Ancient and early Medieval Logic.³⁵

In another passage, Priest re-iterates the distinction, referring back to the discussion just quoted:

...I distinguish between three sorts of account of negation: those according to which contradictions entail:

1. nothing;
2. something (but not everything);
3. everything.

The classical and intuitionist accounts are of kind 3....³⁶

In these two passages, Priest has given us both a simple and clear-cut diagnosis of the roots of explosion, and a very clean division of the conceptual possibilities. We have three different conceptions of negation, and, purely as a consequence of the views of negation, three different conceptions of the content of contradictions. If, as in the cancellationist view, asserting a negation has the effect of simply erasing the negated proposition from one's belief set, then asserting a contradiction adds no content to that belief set. If, as in the complementationist view, asserting the negation of a proposition entails everything not entailed by the original proposition, then, since a conjunction contains everything either conjunct contains, asserting a contradiction adds *all* content to one's belief set. To get the intuitively appealing result that contradictions have some content (but only some), you need the intermediate account which Priest claims for paraconsistent logic.

Most importantly for our present purposes, Priest is very clear in his contention that classical logic assumes the view of negation labeled (2) in the first passage and 3 in the second passage. He claims in the second passage that explosion is a result of the

³⁵ Priest (2006b), p. 31

³⁶ Priest (2006b), p. 75

“account of negation” assumed by classical logic. In the first passage, he tells us what that account is. “A complementation account...is packed into orthodox modern logics, such as ‘classical’ and intuitionist logic.” This means, remember, that “ $\neg\alpha$ has whatever content α does not have.” Moreover, unless α containing the content that β is sufficient for α entailing β , it’s deeply unclear what conceptual connection Priest could see between explosion and the *account of negation* embodied in classical logic.

Now, I agree that the intermediate account of negation is far more plausible than either cancellationism or complementationism. Even if my previous belief set never included α , I have surely *changed my beliefs* when I come to believe $\neg\alpha$. After all, it previously contained no beliefs about whether α was the case, and now it does! Similarly, it is just not true that any random β must be entailed by either α or $\neg\alpha$ for any α . The claim that apples are fruits does not entail that some cats are human beings, and *neither does its negation*. Paraconsistent logic surely relies on an intermediate account, and if Priest is right in claiming that classical logic assumes complementationism, then he has succeeded in producing a powerful argument that paraconsistent logic does a better job of capturing our intuitions about negation than classical logic.

Fortunately for the classical logician, Priest is exactly wrong in claiming that classical logic assumes (or is even compatible with) a complementationist account of negation. This can be demonstrated by examining in detail what a complementationist logic would look like. For the sake of clarity, I will call classical logic CL and the complementationist alternative TCL, for “truly complementationist logic.”

In TCL, for any α and any β , if α didn’t entail β , then $\neg\alpha$ would. (To re-emphasize, remember that Priest’s classification of accounts of negation embodied in

logics, and of what follows from contradictions given those accounts, *only makes sense if you assume that containing information is a sufficient condition for logically entailing it.*)

Given complementation, if “the Red Sox will not win the 2011 World Series” did not entail that “the Earth will be invaded by aliens in 2012,” then “the Red Sox will not win the 2011 World Series” *would* entail this disturbing result. The sorts of examples used to show students in Critical Reasoning courses that the Fallacy of Denying the Antecedent is in fact a fallacy would actually be classified as valid arguments in TCL. After all, given complementation, and given that α and $\neg\alpha$ were both consistent claims and hence both had partial content, and that β followed from α , $\neg\beta$ really *would* follow from $\neg\alpha$.

It gets worse. If negated propositions entail whatever the claims being negated do not, and if (as they must, given that first assumption) contradictions entail *everything*, then it follows inexorably from those two assumptions that tautologies entail *nothing*. Thus for example, Double Negation would fail for them. After all, $\neg\neg\neg(\alpha\wedge\neg\alpha)$ would *not* be entailed by $\neg(\alpha\wedge\neg\alpha)$, since $\neg\neg\neg(\alpha\wedge\neg\alpha)$ is already entailed by $(\alpha\wedge\neg\alpha)$. It would not be part of the complement of what is entailed by $(\alpha\wedge\neg\alpha)$, since that complement would be *empty*.

Conjunction-elimination would also fail for such α 's. $(\neg(\alpha\wedge\neg\alpha) \wedge \neg(\beta\wedge\neg\beta))$ cannot entail $\neg(\alpha \wedge \neg\alpha)$, since $\neg(\alpha \wedge \neg\alpha)$ is entailed by its contradictory, $\neg(\neg(\alpha\wedge\neg\alpha)\wedge\neg(\beta\wedge\neg\beta))$. Similarly for disjunction addition: $(\neg(\alpha\wedge\neg\alpha) \vee \beta)$ could not be entailed by $\neg(\alpha\wedge\neg\alpha)$, since it is entailed by $(\alpha\wedge\neg\alpha)$.

Even the Law of Identity would fail for logical tautologies in TCL. If $\neg(\alpha \wedge \neg\alpha)$ can entail *nothing*, it cannot entail itself.

Worse yet for the claim that *classical* logic is complementationist, although the hypothesis that classical logic assumes complementation explains the principle that “from a contradiction, anything follows,” it actually gets in the way of the parallel classical principle that “from anything, a tautology follows.” If the information that some tautology is the case is contained by α , it *cannot* be contained by $\neg\alpha$.

Given the number of rules of CL that have already been seen to fail in TCL, and vice versa, at this point it would be safe to conclude that TCL is a *very* different creature than CL. Contradictions explode in both CL and TCL, but they clearly explode for different reasons. As such, CL simply is not complementationist. It is just as intermediate-ist in its treatment of negation as its paraconsistent rivals.

The reason that explosion is considered to be valid in classical logic has nothing to do with the account of negation it embodies. The real explanation is much simpler. *Monaetheism* is packed into classical logic. Given the assumption that the failure of any given contradiction to be true is guaranteed by its logical form, it makes perfect sense to say that the inference from any contradiction to any conclusion is (vacuously and uninterestingly) truth-preserving, just as an inference from any premise *to* a statement whose truth is guaranteed by its logical form is guaranteed to be (equally vacuously and uninterestingly) truth-preserving. In order to see whether the assumption *actually* underlying explosion is a reasonable one, we must now turn to a much harder and more complicated subject, which is the debate between monaetheism and dialetheism.

Chapter Three: Motivations for Dialetheism

The best-known argument for dialetheism is the argument from the Liar Paradox. Indeed, some critics seem to regard the Liar as *the* motivation for dialetheism, or at least the only one worth responding to. For example, in *Saving Truth From Paradox*, Hartry Field vaguely speaks of “other motivations,” only to dismiss them.

Paraconsistent dialetheism has been promoted in recent years, by Graham Priest and others, as the best way of dealing with the semantic paradoxes and related paradoxes such as Russell’s. (Priest cites other motivations as well, but few have found these persuasive and I will not discuss them.)³⁷

Whatever one thinks of Field’s procedure for deciding which of Priest’s arguments are worth addressing, this at least sounds like he’s prepared to take Russell’s Paradox seriously. A few chapters later, he corrects this misapprehension, telling us that “the possibility that we should take a dialethic attitude toward set theory itself” is “a possibility that Priest takes seriously but I suspect few of his readers will, and...in any case I am unwilling to...”³⁸

Considering that there are several chapters on dialetheism in *Saving Truth From Paradox*, and that, as such, it may be the closest thing that currently exists to a book-length response to the case for dialetheism that Priest lays out in books like *In Contradiction* and *Doubt Truth To Be A Liar*, Field’s neglect of all of Priest’s non-Liar-related arguments for dialetheism represents a serious hole in the literature. In what

³⁷ Field (2008), p. 361. To be clear, Field is willing to take a non-set-theoretic *variation* of Russell’s paradox seriously—the Russell Property, which is the property of not instantiating itself. My attitude is the opposite of Field’s. I take the indispensability argument for mathematical realism to be compelling—see below—and hence, since it seems to establish that sets *exist*, I take monaletheism to be under serious *prima facie* threat from Priest’s argument for set-theoretic inconsistencies. On the other hand, I know of no similarly plausible motivation for realism about properties-as-abstract-objects, and hence no reason to take seriously the claim that the Russell Property *exists*.

³⁸ Field (2008), pp. 384-385.

follows, we'll take a step in the direction of plugging that hole by addressing what seem to be (in ascending order) Priest's three strongest non-Liar-related arguments; the argument from the "paradoxes of motion and change," the argument from inconsistent obligations and the argument from the paradoxes of naïve set theory. This last item will take us very close to the subject of the next few chapters, the Liar, given the strong structural parallels between the difficulties for monaletheism posed by the Russell Set and those posed by the Liar Sentence.

I agree with Field that the argument from the Liar Paradox is the strongest weapon in Priest's arsenal, and the one that it is most difficult to counter. I disagree, however, with Field's implication that the other arguments are nothings. They are somethings, and if we want our rejection of dialetheism to be the reasoned outcome of serious deliberation, rather than a rash dismissal of an unfamiliar idea simply *because* it runs counter to our familiar patterns of thought, we will need to say something about these arguments.

I. The Paradoxes of Motion and Change

In the tradition of figures like Engels, Hegel, Zeno and Heraclitus—that is to say, of at least a couple of minor players on both sides of *Monty Python's* Greeks vs. Germans soccer game—Priest believes that change is inextricably entwined with contradiction. Unlike any of those other people, Priest has provided a rigorous formalization of his claim.

He believes that the process of change over time is best captured by a tense logic. This should not of course be any of the standard tense logics, in which contradictions are as explosive as they are in any other classic logic, but Priest’s own “Logic of Paradox” (LP) suitably extended to include tense operators. P means “it was the case that” and F means “it will be the case that.” (We’ve also got H, which is “it was always the case that,” understood as $\neg P\neg$, and G, which is “it will always be the case that,” understood as $\neg F\neg$, but those extra ones don’t come up for our purposes here.) As in the un-tensed version of LP, we don’t think of truth values as functions with either 1 (T) or 0 (F) is assigned to each propositions, but as a relation in which a proposition can be related just to 1 (T), just to 0 (F) or to both 1 and 0 (TF).³⁹

With this machinery in place, Priest can express what he calls Zeno’s Principle:

$$(\alpha \wedge P\sim\alpha) \vdash [(\alpha \wedge \neg\alpha) \vee P(\alpha \wedge \sim\alpha)]$$

Translated back into English, this would be “the conjunction of alpha and it was the case that not-alpha entails that the conjunction of alpha and not-alpha either was or is the case.” In other words, every time anything changes in any way—for example, every time someone is awake but in the past was asleep, every time a door is opened that in the past was closed, etc., etc., etc.—this state of affairs is absolute, deductive proof of the presence of a true contradiction.⁴⁰

³⁹ In his article ‘What’s so bad about contradictions?’, Priest includes a fourth option, which is that propositions can be related to neither 1 nor 0. In *In Contradiction*, however, he argues strenuously against the possibility of such ‘gaps.’ In any case, since (as we’ll see below) he makes no appeal to gaps here, we can safely exclude that possibility from our discussion of his theory of change.

⁴⁰ In his initial discussion (Priest (2006a), pp. 160-162), Priest merely argues that *some* changes are “ Δ -type changes” (i.e. changes that involve contradictions), but the considerations that he advances would apply to any and all changes that involve ‘instants of change,’ and he makes it abundantly clear in his critique of the ‘cinematic theory of change’ that he doesn’t think that a succession of states in which only α or only $\neg\alpha$ is the case would add up to real *change* from α to $\neg\alpha$, so it’s abundantly clear that his initially modest goal of seeing whether *some* changes are contradictory is rapidly superseded by the more ambitious one of showing

So as not to miss the true radicalism of the suggestion here, it is important to emphasize that we are talking about real contradictions, in which objects have and lack precisely the same property in precisely the same sense, and not the sort of easy “contradictions” that could be generated with vague predicates⁴¹. (E.g. a car with a greenish-blue paint job is not “green, but also blue and therefore not-green,” even for Priest.) Take his stock example of a person entering into a room, where at the instant of change—as his center of mass is passing through the exact center of the doorway—he both is and isn’t in the room. Priest doesn’t mean that he “sort of is and sort of isn’t” in the room, he means that he is and, in precisely the same sense, he isn’t. *That’s* interesting.

Also interesting is the fact that Francis Jeffrey Pelletier, in a review in *The Bulletin of Symbolic Logic*, interprets Priest as talking about vagueness.⁴² He thinks that Priest’s alleged contradiction occurs at the “moment” when one’s center of mass is precisely at the center of the doorway, not as it that exact center of gravity *is passing* the center of the doorway.⁴³ Some of Priest’s formulations are open to this interpretation, but it misses not only his repeated protestations that he is *not* talking about vagueness, but the general

that *all* changes are contradictory. When, in the following section (Priest (2006a), pp. 162-165), he introduces the tense logical machinery and advances ‘Zeno’s Principle’, there is no remaining room for doubt about this subject. Priest is flatly asserting that change is impossible without contradiction.

⁴¹ Of course, there are philosophically interesting proposals on the table about vagueness-generated contradictions—we’ll meet some of those in Chapter Nine—but Priest here is making a much different sort of proposal.

⁴² Pelletier (2006), pp. 132-133. Along the same lines, Thomas Takho, in his otherwise extremely interesting article *The Law Of Non-Contradiction As A Metaphysical Principle*, takes Priest’s example as an “obvious example of vagueness in language.” (Takho (2009), p. 40) In both cases, I think critics fail to engage with the real force of Priest’s argument when they fail to take seriously his insistence that he *isn’t* talking about vagueness.

⁴³ Here are Priest’s exact words: “As I walk through the door, am I in the room or out of (not in it)? To emphasize that this is not a problem of vagueness, suppose we identify my position with my exact centre of gravity, and the door with the vertical plane passing through its center of gravity.” (Priest (2006a), p. 161) Given his explicit assurance that he isn’t talking about vagueness, and his harsh words about the account of motion as *merely* being about an object occupying different positions at different points in time (see n. 30, below), this cannot be plausibly interpreted as a matter of vagueness about whether ‘occupying the position at the exact center of the doorway’ counts as ‘occupying a position in the room’ or ‘occupying a position that’s not in the room.’

spirit of Priest's argument. For Priest, the state of being at the exact center of the doorway would be one of many static states, as opposed to the dynamic state of change *between* states.

In any case, if this was what Priest meant, it seems odd that he would have thought that his argument would show that "I am in the room" is both true and false, rather than just that "I am in the room" and "I am in the hallway outside the room" are not jointly exhaustive of the possibilities. The statement "I am in the room" would be false (but not also true), and the true (but not false) proposition would involve a predicate like exact in-between-ness. No contradictions need be involved. This is, pretty clearly, not what Priest is getting at.

Given this clarification, it's worth spending a moment really considering the enormity of what is being claimed. Unlike the classical candidates for true contradictions derived from semantic and set-theoretic paradoxes, this is not something that can be thought of as strange spandrel of our mechanisms for referring to the world. Nor is this solely about some odd area, like quantum mechanics, where strange things may seem plausible. We are, rather, talking about inconsistent facts about tables, chairs, stars, planets, dogs and cats.

The second thing to notice is that, although Priest does not seem to realize it, "Zeno's Principle," expressed in this way, simply assumes a controversial view of the ontology of time, the "A-Theory," according to which there is an ontologically privileged present moment. Some A-Theorists—usually called presentists—believe that *only* the present moment exists, while others subscribe to what's been called "moving spotlight eternalism," according to which past and future states of affairs exist but have objectively

existing “MacTaggart properties” of pastness and futurity.⁴⁴ For example, the American Civil War was once future, then it was present and it is now in the past. Priest seems to be an A-Theorist of the second, more moderate kind.⁴⁵

This distinction, however, need not come into play here. The important point is that, for a B-Theorist—someone who denies stance-independent objective reality to the “McTaggart properties” of pastness, presentness and futurity—any logic that captures the nature of time could neither require nor admit of tense operators. For the B-Theorist, after all, there is no fact of the matter about whether any given fact “was the case” or “will be the case.” Writing this in 2007, I am within my rights to refer to 2006 in the past tense, just as someone writing in 2006 is within their rights to refer to 2007 in the future tense, and neither of us is wrong. That is because the “was the case”/“will be the case” distinction is relative to when “now” is, and for the B-Theorist “now” is just an indexical. If I say that “Miami is here” and someone in Hong Kong says that “Hong Kong is here,” we are not having a disagreement.

⁴⁴ For a classic defense of full-on ‘presentism,’ see Prior (1967). For a defense of the moving spotlight eternalism view, see Smith (1993) Smith’s critique of presentism in Smith (2003) is also interesting. He makes the point that, although the founding fathers of tense logic were full-on presentists, it’s hard to see what could make $P\alpha$ true if the past no longer exists, or $F\alpha$ true if the future does not yet exist. Indeed, when you consider that it *takes time* for light to bounce off visible surfaces and onto the retinas of our eyes, causing visual images, or for soundwaves to hit our ears, etc., meaning that technically everything we perceive is (slightly) in the past, one of the odder consequences of presentism is that it entails that *none of our experiences are veridical!* Everything that we perceive is non-existent! Of course, Smith’s more moderate version of the A-Theory has some extremely odd consequences of its own. For example, if the past and the future exist (but have objective, stance-independent properties of pastness and futurity), then the question of whether we are currently in the past, the present or the future, which sounds like a nonsense question, is a very real skeptical worry. It is far from clear how someone with Smith’s views could answer it. (Thanks to Brian Mondy and Nick Stang, in conversation, for helping to bring out this objection.)

⁴⁵ His frequent collaborator Richard Routley was a full-on presentist, and made heavy use of presentism in his case for ‘noneism’ (the doctrine that non-existent objects have properties and can be referred to) in *Exploring Meinong’s Jungle*. (Assuming presentism makes it a lot easier to argue for noneism, given the obvious possibility of referring to past and future objects. As always, one person’s Modus Ponens is another’s Modus Tollens.) Priest, however, parts ways with Routley in his own polemic for noneism, *Towards Non-Being: The Logic and Metaphysics Of Intentionality*. He lists the assumption that only the present exists as a feature of Routley’s version of noneism, but not of his own. This is hardly conclusive, but it does seem to suggest that Priest subscribes to the more moderate version of the A-Theory.

For a B-Theorist, there can be “earlier than” and “later than” a certain point in time, but no raw “was” or “will be” any more than there is a “taller than” without a point of comparison. The B-Theorist could no more make sense of a tense logic than the rest of us could make sense of a “height logic” with tallness and shortness operators.

This severity of Priest’s commitment to the tensed theory of time becomes even more apparent when he applies his analysis of change to the notion of “the flow of time.” For Priest, one moment literally changes into another, and at the instant of change, it is both. He thinks this captures the intuitive notion that events go from being future to present to past, and that a virtue of his theory is that it “makes sense of” the “fact” that time is importantly un-space-like.

Priest’s A-Theoretic commitments may seem too obvious to be worth going on about, and perhaps they would be, except for the odd fact that Priest repeatedly claims that his analysis of change is neutral in the debate over the reality of tense.⁴⁶ In any case, his analysis of change is most definitely not neutral between the A-Theory and the B-Theory. B-Theorists emphatically don’t accept that time is importantly un-space-like. That’s the nub of their disagreement with A-Theorists.

This is deeply problematic, but for the moment, as we unpack Priest’s argument, let’s suppose that the A-Theory is correct and see what follows. Let’s further suppose that time is dense and continuous, meaning that it is infinitely divisible into smaller and smaller intervals, and that for every pair of zero-duration instants, there is a third instant between them. In order to get a mental handle on this picture, an analogy to geometry might be helpful. An interval is like a segment of a line, and instant is like a point on the

⁴⁶ Worse yet, he slips up and refers to tensed statements as “indexical,” which only makes sense from the perspective of the tenseless theory that he so clearly rejects. Even brilliant logicians, it seems, can be sloppy in their formulations about areas of philosophy which they do not ordinarily work in.

line that takes up no space. Intervals can last for one sixteenth of a second, or for a hundred billion years, or whatever, but instants are exact locations within those intervals...what you would get, so to speak, if you could somehow hit the pause button on a DVD player controlling the universe as a whole.

Now, although the assumption that time is dense and continuous is in fact frequently made in discussions about the ontology of time, some philosophers have argued that it is inconsistent with certain results in quantum mechanics, which might indicate that time is discrete after all.⁴⁷ A discussion of whether they are right about this would lead us well astray of the main line of argument here, so let's just note that it is still controversial whether the quantum results really do have this conceptual consequence, that at least some philosophers argue that this alleged consequence sneaks in indefensible verificationist baggage, and that here as elsewhere, the quantum results conflict with the relevant relativistic results in any case. Whoever's right about that, let's grant at least for the sake of argument that Priest is right about the dense and continuous nature of time.

With this conceptual machinery in place, we can ask the central question motivating his analysis of change. At the instant in which an object O changes from having property P to lacking it, which one does it have? There would seem to be four options.

- (1) O is P.
- (2) O is not P.
- (3) O is neither P nor not P.
- (4) O is both P & not P.

⁴⁷ See the first two dialogues of Smith and Oaklander (1995) for a detailed discussion of these points.

In this case, (3) would seem to entail (4) in any case, so it cannot be seen as an *alternative* to (4). There may be cases in which we should decline to either attribute a property or the negation of that property to some object—where, for example, some sort of category mistake is being made—but this at least does not one of them. When it comes to a sentence attributing a particular and clearly possible physical property to a particular and actually existing physical object that is clearly capable of having that property, this is a truth-evaluable proposition if anything is. If “Ben is in the room,” is not true, then “Ben is not in the room” is false. Thus if we negate both sentences, we’ve affirmed the sentence, “Ben is both in the room and not in the room.”

This leaves us with (1), (2) or a single inconsistent option (3)/(4), which for the sake of convenience, we’ll simply refer to as (4). So, which is it? If (1) is true, then the change has not yet occurred. Priest seems well within his rights to deny that *this* is the instant of change. If it is (2), then in what sense is this the instant in which the change has occurred? It seems to be, rather, some instant after the change. As such, we are left with (4).⁴⁸

The obvious place for the opponent of true contradictions to object to this reasoning is the original assumption that there is some individual instant at which change

⁴⁸ JC Beall, who believes in gluts but believes that none of them occur in the truth-predicate-free “base language,” provides an excellent response to Priest’s account of change in Beall (2009). The argument that I take to be Priest’s central arguments for this account is one that Beall ignores (and vice versa). (Beall focuses on Priest’s “Zeno’s Arrow” argument that, if, at each instantaneous point in an arrow’s motion, it makes no progress, given that a “sum of nothings... is nothing,” unless we postulate contradictory positions in our account of motion, it would be impossible for the arrow to move.) I take Beall to have responded to this argument so well that there’s no point in duplicating his effort, so I’m happy to take what I say about this subject and what Beall says to complement each other nicely as an overall response to Priest on change and (mostly) leave it at that. (Beall’s discussion comes in Beall (2009), pp. 127-129.) The only thing I’ll add to what Beall says about the argument I’m *not* addressing is that, if interpreted (charitably) the way that Beall does, it does seem to commit a composition fallacy or something close to it, but that if interpreted literally (a “sum of nothings” meaning an actual arithmetical sum), it simply gets the math wrong. There *is no sum* of all the instantaneous points in some interval, because the mathematical operation of addition simply doesn’t apply to the members of uncountably infinite sets. (See Smith and Oaklander (1995), p. 7, for a discussion of this point.)

occurs. Why not say that “change” is a description of the variation of properties of an object among the instants at which that object exists? For a certain interval, O is P and in the next interval O is not-P. That is what “change” means, and at every individual instant, O is one or the other but not both.

Priest is of course aware of this option, which he identifies with Russell and calls the “orthodox” or “cinematic” account of change. He believes this does not capture the notion of “change,” and also indicates that there is something magical about going straight from having one property to having another property without some sort of transitional state.

...because if there is no such instant, there is no time at which the system is *changing*. Thus, in a sense, there is no change in the world at all, just a series of states patched together. The universe would appear to be more like a sequence of photographic stills, shown consecutively, than something in a genuine state of change or flux....the cinematic theory of change is highly counter-intuitive.⁴⁹

At one point, he takes this to the extreme of saying that if God had merely created a series of successive states spread out over time, he would not have created change.⁵⁰

It seems to me that this argument fails on three counts.

First, it conflicts with our best science. Priest’s entire framework for dealing with these questions—and certainly, his claimed advantage that it “explains” how time “flows” and is “importantly un-space-like”—systematically assumes the A-Theory of time, and Einstein’s Special Theory of Relativity (STR) seems to entail the B-Theory. As philosophers like Hilary Putnam, D.C. Mellor and JCC. Smart (and, in their own way,

⁴⁹ Priest (2006a), p. 162

⁵⁰ After quoting Russell’s comments about motion, Priest writes, “What we have here...is the cinematic account of change, where the change in question is motion. And this particular case of the account is no more plausible than the general form. A sequence of states, even a dense and continuous one, indistinguishable from corresponding rest-states, does not seem to be a state of motion. If God were to take temporal slices of an object at rest in different places and string them together in a continuous fashion, he would not make the object move.” (Priest (2006a), pp. 173-174)

popularizing physicists like Brian Green) have argued, it seems impossible to make sense of the idea of an ontologically privileged present moment without absolute simultaneity. It seems to be integral to the notion of “the present moment” that if Event A is part of the present and it is simultaneous with Event B, then Event B is part of the present as well. The problem, of course, is that if the STR is correct, simultaneity is relative to reference frames, so Event B can be simultaneous with Event B (but not with event C) according to one reference frame, Event B can be simultaneous with Event C (but not with Event A) according to another one, and so on, and before you know it, a total hash has been made of the compartmentalization of reality into past, present and future. Taking the STR seriously seems to pretty much demand a “static,” “space-like” understanding of the fourth, temporal dimension of space-time.

Secondly, Priest’s theory of change threatens to undermine his “classical re-capture.” Thus, *even if* dialetheism is correct, we shouldn’t be too eager to embrace contradictions about change. We’ll come back to the classical re-capture issue in some detail in Chapter Eight, so we can content ourselves with the briefest of sketches here. The idea is approximately this:

As we’ve seen, if there were true contradictions, Disjunctive Syllogism would be invalid. Nor would it be the only law of classical logic that wouldn’t survive the jump to the “realm of the transconsistent.” (Perhaps most obviously, Reductio Ad Absurdum would fail as well. Less obviously, for example, so would Modus Tollens.⁵¹) Now, *most* of the laws considered valid in Classical Logic (CL) are also considered to be valid in

⁵¹ Consider the following example. “If there is a sentence S such that S is ‘this sentence is not true’, then Sentence S is true. It is not the case that S is true. Therefore, it is not the case that there is a sentence S such that S is ‘this sentence is not true.’” Modus Tollens is to implication what Reductio Ad Absurdum is to entailment.

Priest's logic LP, but even so, the ones that don't make the cut have really tremendous intuitive plausibility. To pick a famous example, even a dog that follows a scent and, when it comes to a fork in the path, sniffs one fork, fails to pick up the scent, and then goes ahead and starts running down the second fork without bothering to check it first, seems to be instinctively reasoning according to Disjunctive Syllogism. The idea that it is doing something irrational beggars the imagination.

Priest's solution to this problem is that, since the statistical frequency of true contradictions is very low, all else being equal, the epistemic probability of any particular contradiction being true is correspondingly very low. As such, even though they aren't deductively valid, rules like Disjunctive Syllogism still at least have probabilistic force, so someone reasoning according to them in domains where no inconsistencies have been discovered is doing something rationally defensible. In Chapter Eight, we'll examine the vexed question of how exactly Priest can be so sure that, if there are true contradictions, their frequency is very low. For now, the important point is that for the classical re-capture to work, there can't be very many true contradictions.⁵²

⁵² Or, at least, such is Priest's central argument for the classical re-capture in Priest (2006a). In the original 1987 edition, he added a second argument, from a principle about rational belief, but he has since rescinded that, for reasons we'll encounter in Chapter 7. Of course, one might think that the frequency of true contradictions is irrelevant, if the dialetheist can simply specify domains *entirely* free of contradictions, such that we can confidently use classical logic to reason about those, being (rationally) confident that those classical rules that can lead us from true premises to entirely untrue conclusions in inconsistent domains won't get us into trouble. Contradictions, one might think, are like kangaroos. When driving in certain sparsely populated areas of Los Angeles, it's rational to watch out for kangaroos. When driving around South Miami Beach, this isn't a rational worry. It's *logically possible* that a kangaroo could somehow pop up on Alton Road just after one crosses the bridge from Miami proper to Miami Beach, but it's rational to be pretty confident that it won't happen. Similarly, we know where contradictions live—in self-referential truth talk, in set theory, in states of change and so on—and we should duly restrict ourselves to paraconsistent reasoning in those domains, never knowing when we could run over a contradiction, but everywhere else, we needn't worry. My response would be that we can tell a fairly obvious causal story about why kangaroos are unlikely to turn up in Miami Beach not obviously available for domains of inquiry in which true contradictions have allegedly been discovered, and that, on the assumption that dialetheism is correct, contradictions aren't like kangaroos. They're like cockroaches. Cockroaches like to live in dark, dank corners and crevices where they won't easily be seen. You can have cockroaches in your house without discovering them for a long time. (By analogy, if dialetheism is right, certain contradictions

Now, remember that tense logic only makes sense if propositions are constantly changing their truth-values, so that at one moment α is true and $F\alpha$ is false, and at the next α is false and $P\alpha$ is true, and so on. (If propositions have eternal, unchanging truth-values, you don't need tense logic.) Given that, as Heraclitus so poetically reminded us, more or less everything is changing, more or less all the time, if every assertion that some Object O has property P, considered at the moment when it changes from having P to lacking it, is both true and false, that seems to leave us with an uncomfortably large number of true contradictions. I'll leave the question of whether it's a large *enough* number to decisively block the classical re-capture or not open, but for anyone who's hopes for reconciling dialetheism with observed common sense lie with the classical re-capture, it's certainly far too large a number for comfort.

The last of my three objections is the simplest, and I think, the most damaging. Priest rejects the Russellian theory of change because it wouldn't satisfy his intuitions

have always been true, but for *thousands of years*, western philosophy was utterly unaware of them! If anything, the cockroach comparison *understates* the ability of contradictions to stay hidden from prying eyes.) Now, imagine three houses lined up in a row. If cockroaches are discovered in the house on the left and the house on the right, should the owner of the house in the middle start taking precautionary measures like stocking up on bug spray and making sure not to leave food out where cockroaches could get to it? Surely, the answer depends on what the relevant differences are between the house in the middle and the houses to its left and right. If the house on the left and the house on the right are both owned by inattentive slobes who are constantly leaving trails of crumbs across the floors, who rarely clean and who frequently leave the back doors of their homes open out of laziness, whereas the house in the middle has a far more conscientious sort of owner, then the precautionary measures just mentioned may be irrationally over-cautious. If, on the other hand, there's no obvious difference—no relevant trait that the houses on the left and the right have in common and that the house in the middle does not share—then the mere fact that contradictions have not been discovered in the house on the middle *yet* should give its owner no grounds whatsoever for confidence that his house is likely to be cockroach-free, and the precautionary measures would be well-advised. If there are true contradictions in the abstract mathematical realm, the semantic realm, the realm of *physical objects undergoing changes*, the realm of legal (and perhaps even moral) rights and obligations and so on, it's not at all obvious what these 'houses' all have in common that's different in principle from those 'houses' in which no true contradictions have yet been discovered. Moreover, even this last comment somewhat charitably assumes an obvious way to carve up reasoning. After all, if *change* necessarily involves contradictions, one might think the domain of "things subject to change" in general is infested with contradiction. What, however, *doesn't* undergo changes? One might answer that the realm of abstract mathematical objects, if such exist, is changeless, but even if so, *this is also a domain in which Priest claims to have discovered contradictions!*

about change. Things seem to change, and the “cinematic” theory simple doesn’t satisfy the intuitive sense of what real change would have to be. Fine. On what basis do we know that “real change” is occurring? If we poetically assume theism to frame this more sharply, how do we know that God created change?⁵³ If He had only created a succession of instantaneous static states, what would be different? If he whimsically decided to eliminate change from His universe, starting today, and to only maintain the static states, how could we tell? (The resonances with Benacerraff’s critique of mathematical Platonism are, of course, intentional.) Instantaneous states are unobservable by their very nature. For us to be able to see things, hear things and so on, we need to have access to them for intervals, so that beams of light can impact our retinas, sound waves can impact our ear drums, and so on.

Now, *sometimes*, we have excellent reason to suppose that certain unobservable things are radically different in their properties than observable things. (Quantum non-locality is an obvious example.) In these cases, however, those excellent reasons come from the role that the behavior of unobservables play in an overall theory that leads to testable results on the observable level. Nothing similar is true here. In fact, the lack of observable-level consequences is key to making this view compatible with Priest’s overall picture of things, since not even Priest believes that the observable world is inconsistent, and in fact he has argued at great length that we have no good reason to postulate observable physical contradictions.⁵⁴ So, without observable evidence for inconsistent states of change, all we have is a vague and lingering intuitive sense that change is real and that Russellian change wouldn’t count.

⁵³ This is, of course, only a rhetorical flourish. Atheism is, as far as I know, one philosophical view on which Priest and I are in complete agreement.

⁵⁴ See Priest (2006), pp. 62-64

But wait. At bottom, the state of “having an intuition” is presumably a certain kind of neurological event resulting from cultural conditioning, eccentric personal history, deep evolutionary programming or some complex mixture of the three. At best, our intuitions about change reflect the way we’ve evolved to see the world. Of course, any given trait could be a Gouldian spandrel, but even *if* our shared intuitions about change are in some sense adaptive, that’s not much comfort. The way that the world “pushes back” against us in natural selection is a very blunt, approximate matter that plays out on the macro-, observable level, and it seems terribly hard to see how evolutionary psychology could explain why our vague, lingering intuitions about change would track the true state of unobservable instantaneous states of change.

In fact, if such states were contradictory, I see absolutely no reason to suppose that our intuitions would track this fact. Even if we conceded on separate grounds that it was possible that states of change worked that way (i.e. that dialetheism was true), to have a good reason to believe it, we’d need some kind of confirmation (even in a high-level holistic sense, with very distant empirical checkpoints) at the observable level. We don’t have the slightest reason at that level to accept it, so we shouldn’t.

Let’s move on.

II. $(O\alpha \wedge O\neg\alpha)? (O\alpha \wedge \neg O\alpha)?$

We don’t have any particularly good reason to believe that states of change are inconsistent, but we actually might have good reasons to believe that obligations are, in a

certain sense, inconsistent.⁵⁵ This may even be true of moral obligations, but it is undoubtedly true of legal obligations. Priest gives a plausible example in *In*

Contradiction.

Take a very simple case. I contract with party X to be present at a certain spot at a certain time. Separately, I contract with party Y not to be present at that spot at that time. Both contracts are validated in the usual way, by witnessing, etc. I may do this with or without ill intention. It may be my intention to deceive one of the parties. On the other hand, I may just be absent-minded. In such circumstances I am legally obliged both to be and not to be at this spot at this time. (And if it is suggested that this is not a case of inconsistent obligation simpliciter, since it is obliged to X to be at that spot and obliged to Y not to be, just take X and Y to be the same person.)

How can one be sure that I am committed to inconsistent obligations in the situation described? The answer is simple. If, after the event, I am sued by the party of whichever contract I did not comply with, the court will hold me in breach of obligation and award damages appropriately. Having committed myself to do something different is no defence. This is an important test, so let me spell it out clearly. A sufficient condition for my being legally obliged to do something is that the court of the appropriate jurisdiction (and in the last instance the ultimate appellate court) would hold me in default were I not to do it.⁵⁶

Now, in this situation ($O\alpha \wedge O\neg\alpha$) seems to be undoubtedly true. Before getting too deeply into this, though, we should clarify a background issue. It is common to casually read the obligation operator O as “it is obligatory that,” but that doesn’t quite work. α could, after all, be any sentence whatsoever, and for many sentences affixing “it is obligatory that” to the beginning of the sentence would render the result not just false but utterly nonsensical. Rather, we should be careful to read it as something very roughly like “if certain relevant obligations were to be fulfilled, α would be the case.” Priest suggests a formulation about “all extant obligations,” but this also seems quite

⁵⁵ In conversation, during his 2006 visit to the University of Miami, Priest claimed that the argument from inconsistent legal obligations was the best argument for dialetheism. I have no idea whether this represents his considered view or not, but it certainly seems sufficient to warrant the inclusion a discussion of it here.

⁵⁶ Priest (2006a), p. 182

problematic.⁵⁷ After all, consider a situation where Jack is lunging at Jill with a knife and Jane is in a position to stop him. Let α be “Jane stops Jack from killing Jill” and β be “Jack stops in his tracks and refrains from killing Jill.” $O\alpha$ and $O\beta$ both seem quite correct to me, but $O(\alpha \wedge \beta)$ seems to confuse the issue. As such, I’m not ready to follow Priest (and many other deontic logicians) in assenting to the principle that $O\alpha \wedge O\beta \vdash O(\alpha \wedge \beta)$, or in general to the principle that obligation is transmitted over entailment.

That said, for the purposes of the present discussion, nothing much will hang on this, since I see no reason to deny that inconsistent obligations of the kind $O(\alpha \wedge \neg\alpha)$ could exist. It could be that if we were to fulfill certain relevant obligations (even *the very same* obligation, as in Priest’s case of multiple contract), the result would be that we would have brought about a true contradiction. So much the worse for us, if true contradictions are impossible, but that may be the frustrating reality of our deontic situation.⁵⁸

Of course, in the case of *moral* obligations, we may have a good reason to deny that $O(\alpha \wedge \neg\alpha)$ is possible (given that bringing about the relevant $\alpha \wedge \neg\alpha$ is impossible, which we need to be very careful about, so as not to beg the question against the dialetheist), which is the “ought implies can” principle.⁵⁹ Now, there is a vast literature

⁵⁷ Priest (2006a), p. 189

⁵⁸ In fact, remember that not even Priest believes that there are any contradictions in the ‘observable world’ (see above) and the sort of true contradiction that we’d have to bring about in the sort of case we’re talking about would be on the observable level, so *even if dialetheism is correct*, this may still be impossible.

⁵⁹ On the issue of moral inconsistency, Priest is slightly coy. He says at one point that his discussion of legal conflicts should “apply equal well” to other sorts of norms, and he includes moral norms in that list, saying that “many writer have recently suggested not all *prima facie* moral conflicts can be resolved, and I think the truth lies with them.” (Priest (2006a), pp. 197-198) On the crucial subject of the argument to true contradictions, however, all he says is that “the legal analogy suggests that there may be moral dialetheias too.” (Priest (2006a), p. 199)

on this subject, and many critics worry that the principle can't withstand Frankfurt-type cases, but on a basic intuitive level, there does seem to be something monstrously unjust about blaming people for things that were not under their control. Like many who share this intuition, but take Frankfurt-type cases seriously, I'm deeply sympathetic to the project of reconstructing "ought implies can" on a plausible compatibilist reading of "can." That said, we have an excellent reason to side-step the whole issue for the purposes of this discussion, because when O is read as pertaining to *legal* obligations rather than moral ones, there's no plausible reason to suppose that any sort of "ought implies can" principle will help us. It may be monstrously unjust to hold people responsible for doing things they had no control over, but it's hard to believe that anyone with even a passing familiarity with the history of the last century, much less the previous several thousand years, would deny that it's possible for governments to be monstrously unjust or to pass monstrously unjust laws.⁶⁰

The interesting issue, then, is not whether $(O\alpha \wedge O\neg\alpha)$, or even $O(\alpha \wedge \neg\alpha)$ could ever be true, since it seems clear that such things can be true, but whether any of that gets us to the dialetheia $(O\alpha \wedge \neg O\alpha)$. Of course, in some deontic logics, $O\neg\alpha \vdash \neg O\alpha$, but the only obvious reason why this should be true is the background assumption that $O\alpha$ and $O\neg\alpha$ are never simultaneously true, which we've already rejected. (Compare to the fiction case: no one who believes that fiction contains inconsistencies would be tempted to assent to $F\neg\alpha \vdash \neg F\alpha$. Far from a matter of rigging the rules in an ad hoc way

⁶⁰ Of course, one could object that not all inconsistent legal obligations are cases of injustice. In Priest's contract case, for example, the multiply-obligated individual has no one to blame but themselves. Whatever one thinks of this, my point is a narrow one: even *if* we assent to the strongest possible case for 'ought implies can,' really milking the relevant moral intuition, that won't help us out of inconsistent legal obligations.

to avoid dialetheias, this is a simple principled matter of a total absence of any motivation for believing that entailment to be truth-preserving.) In any case, Priest rejects $O\neg\alpha$ $\vdash \neg O\alpha$, and indicates that sometimes inconsistent obligations entail inconsistencies about obligation, but not always. Sadly, he gives no indication of which situations are which.

I can think of one possible route from $(O\alpha \wedge O\neg\alpha)$ to an outright contradiction using standard, plausibly well-motivated deontic principles, although interestingly enough, Priest doesn't mention it. Here it is: $O\alpha \vdash P\alpha$ and $O\alpha \dashv\vdash \neg P\neg\alpha$ are both valid in standard deontic logics, and both seem intuitively attractive. Start with $(O\alpha \wedge \neg O\alpha)$. By conjunction-elimination, we get $O\alpha$. By the second rule just mentioned, we get $\neg P\neg\alpha$. By a second instance of conjunction-elimination, we get $O\neg\alpha$. Plugging in the first of the two rules mentioned above, we get $P\neg\alpha$. Now, plugging in conjunction, we've got the outright contradiction $(P\neg\alpha \wedge \neg P\neg\alpha)$.

Whether to accept either or both of the rules we've just mentioned, given our acceptance of inconsistent obligations, is a tricky issue. Remember, for the purposes of this discussion, we're leaving the issue of inconsistent *moral* obligations to one side, but granting that in any case, there are definitely inconsistent *legal* obligations. On the legal reading of O and P, there is no obvious reason why we should accept the inference from $O\alpha$ to $P\alpha$. Just because some legal system will hold you accountable for failing to bring

about α does *not* mean that that legal system will allow you to bring it about. Legal systems can be unjust.⁶¹

On the other hand, in the moral case, the inference seems to be an excellent one. In fact, there seems to be a certain sense in which the concept of moral permissibility is contained within the concept of moral obligation. (Think about ordinary language instances in which an obligation is asserted. “Is it OK if I do it?” “Oh, it’s more than OK. You *should* do it.”) Moral obligation is moral permissibility plus.

Now, given this picture *and* a really strong version of the assumption that our obligations fit together consistently, the inference from $O\alpha$ to $\neg P\neg\alpha$ makes sense as well. We can think of it this way. Being merely permitted to bring about α is to be in a deontic relationship of .5 to it, and being obliged to bring it about is to be in a deontic relationship of 1 to it. (Obligation is permissibility plus.) Now, from the mere extra assumption that our obligations are consistent, it merely follows that the sum of our deontic relationships to α and $\neg\alpha$ are never greater than 1.5, but the normal picture goes a step beyond this, to assume that if one’s deontic relationship to α is 1, it can’t even be the case that one’s deontic relationship to $\neg\alpha$ is .5. Assuming for the sake of simplicity that the only possible levels of deontic relationship are 1, .5 and 0, the normal picture has it that the sum of one’s deontic relationships to α and $\neg\alpha$ is always 1...either you are permitted to do either one, or you are obligated to do one and not even permitted to do the other.

⁶¹ As with Priest’s very sensible test for when you have a legal obligation, quoted above, being prevented from doing something by order of a legal system seems sufficient for $\neg P\alpha$.

If, however, for the sake of argument, we grant Priest that there are genuine moral dilemmas, then this picture falls apart. By analogy, imagine that all of the plates in your house are in one of two sizes, and that the larger size is twice as large as the smaller size. Now, your microwave is only large enough to fit one of the larger plates in it, so if either you or your housemate is heating something up on one of the large plates, the other can't sneak in their own small plate to be heated up at the same time. A basic principle of spatial reasoning here will be that, from "the microwave is currently holding one of the larger plates" you can infer that "it is not the case that the microwave is holding one of the smaller plates." (Assume that "holding" doesn't mean "stacked on top of each other.") Now, say that the microwave breaks, and you go to the store and get a new, larger microwave that can fit two of the larger plates at the same time. In this new context, you have absolutely no residual reason to accept the old inference from the microwave holding one of the larger plates to it not holding one of the smaller plates. If it can fit one of the larger plates *and* another one of the same, why not think it can fit one of the larger and one of the smaller? By analogy, if we grant that the sum of one's deontic relationships to α and $\neg\alpha$ can add up to 2, why retain the old inference that if one's deontic relationship to α is 1, one's deontic relationship to $\neg\alpha$ can't even be .5?

To summarize all of this, the following seems to be the situation. Either there can be genuine moral dilemmas or there can't (we haven't taken a position on this issue), but there can certainly be genuine legal dilemmas. The inference under discussion from inconsistent legal obligations to inconsistencies about legal permissions fails because $O\alpha \vdash P\neg\alpha$ fails in legal contexts. If there are no genuine moral dilemmas, the issue does not

arise, but if there are, then we have good reason to think that $O\alpha \vdash \neg P\neg\alpha$ fails. So far, so good.

Unfortunately, that doesn't quite get us out of the woods as far as the possibility of legal dialetheias. Priest has a plausible-sounding case that goes directly to $(P\alpha \wedge \neg P\alpha)$.

Suppose that there is a certain country which has a constitutional parliamentary system of government. And suppose that its constitution contains the following clauses:

In a parliamentary election:

- (1) no person of the female sex shall have the right to vote;
- (2) all property holders shall have the right to vote.

We may also suppose that it is part of common law that women may not legally possess property. As enlightenment creeps over the country, this part of common law is revised to allow women to hold property. We may suppose that a *de facto* right is eventually recognized as a *de jure* one. Inevitably, sooner or later, a woman, who we will call 'Jan', turns up at a polling booth for a parliamentary election claiming the right to vote on the ground that she is a property holder. A test case ensues. Patently, the law is inconsistent. Jan, it would seem, both does and does not have the right to vote in this election.⁶²

Nothing about this seems wrong to me except for the last sentence. Any sort of sophisticated legal reasoning that would show that, because of some legal principle or another, the law isn't *really* inconsistent, would simply miss the point. We can construct a case where such principles would not be relevant, because the issue is not a legal one and traditions of legal reasoning do not apply.⁶³ Imagine that the secretaries of a number of clubs that do not allow their secretaries to be members form a club of their own, the Secretary's Club. The rules of the Secretary's Club are that one is eligible for membership if and only if one is the secretary of a club that does not allow its secretary to

⁶² Priest (2006a), pp. 184-185

⁶³ The following example is a common one, originally formulated by Chihara (1979).

join the club. The club grows to a sufficient size that they decide to hire a secretary of their own. Let's call her Fran, and say that Fran applies for membership in the club. Now, does she have a right to join the club or not?

Moreover, for the sake of simplicity, let's specify that this is a private club, receiving no government funding, in a country with terribly *laissez faire* laws on the matter of private discrimination, so a lawsuit is impossible and legal system will not get involved. The issue is not a matter of legal rights, but of Secretary's Club rights.

Does Jan have a right to vote? Does Fran have a right to join the Secretary's Club? Well, according to the letter of the law, Jan both does and doesn't, and according to the letter of the Secretary's Club rules, Fran both does and doesn't, but there's no outright contradiction in saying that *according to some document*, someone does and doesn't have something, rather than saying that they actually do and don't have it. The question is about each woman's *actual* rights, and this of course raises the question of what exactly it means to have an actual legal right, or an actual Secretary's Club right. This is a conceptual distinction we all immediately and unproblematically take for granted in other contexts. To take a familiar example, consider a totalitarian regime with official constitutional documents enshrining the right to free speech, but in which anyone who criticizes the Great Leader gets a midnight knock on their door from the secret police. The normal, natural way to summarize the situation is that "citizens of Country X don't have free speech rights, although the Constitution says that they do."⁶⁴

⁶⁴ JC Beall similarly argues that legal obligations have this sort of "according to" logical structure, and thus that we need not posit legal dialetheias. His central argument is, however, that different legislatures make different enactments. After all, if not, legal gluts come very cheap, they are very frequent, and Priest's complicated examples are unnecessary. "According to Australian law, it's illegal to drive on the right. According to American law, it's legal to drive on the right." (Beall (2009), pp. 126-127) While I take this to

Ultimately, either Jan will be allowed to vote, or not, but not both, and ultimately, Fran will either be allowed to join the club, or not, but not both. The Secretary's Club cannot perform the miracle of making Fran both a member and not a member at the same time and in the same sense, nor can the moderately misogynist legal system of Priest's hypothetical nation perform the miracle of both letting Jan vote and simultaneously, in precisely the same sense, not letting her vote. Not even Priest seems to be suggesting anything of the kind.

Remember that earlier, Priest plausibly claimed that a sufficient condition for you having a legal obligation was the relevant court would find you in default if you didn't carry it out. Let's take a step further and say that there is nothing more than this to having a legal obligation. (Of course, to fill this out, we'd need to specify the conditions a bit more. One wouldn't want to rule out the possibility that people *sometimes get away with violating their legal obligations*. Hence effective enforcement and punishment can't be aspects of the practice that are absolutely necessary in every case for us to talk accurately about the presence of a legal obligation. Still, the general idea is clear enough. Legal obligations are defined by legal practice.) It's hard to say what more one could plausibly pack into the notion of having a legal obligation. Similar considerations should apply to legal rights. Why postulate legal rights and obligations as abstract entities floating above

be a good and persuasive argument, I don't take it to be decisive. After all, Priest could simply argue that it's illegal to drive on the right *in Australia* and that the decisions of the Australian legislature don't (and aren't meant to) establish anything about where one drives when in the States. (Of course, it's possible that the governing legislature of Country A and Country B could pass laws explicitly regulating the rights and duties of each other's citizens, one making it legal to drive on the right, the other making it illegal, and neither law recognized by the other country. Perhaps Country A claims the territory of Country B as being within its borders and vice versa, though neither country is able to enforce its claim on the other's territory. In this situation, the "in Country A"/"in Country B" dodge wouldn't work. In this sort of situation, however, given its relative exoticism, Priest might be more inclined to postulate a legal glut than he would be with reference to Australian and American driving laws.)

actual legal practice? What would cause such abstract rights and obligations to come into existence? How would we causally interact with them such that we came to be aware that they existed, and to be aware of what exactly they were?

One could argue that in some sense these are abstract objects that don't entirely "float free" from the practice, and that they were brought into existence by the legislation and somehow continue to depend on the legal practice for their existence. Even if we could somehow make sense of the "brought into existence by" and "depend on for their existence" claims without claiming interaction between physical and abstract realms, however, this move still simply pushes back the question. If these relations between legislation and legal practice on the one hand, and causally inert abstract objects, on the other hand, *did* exist, how would *we* ever know that they did? To be sure, if we knew that legislation brought certain abstract objects into existence, we would know that they existed in any particular case, we could be informed of them by reading the law documents and so on, but the real question is *how would we ever find out that the general claim was true in the first place?*

Of course, one might press the point and argue that ordinary discourse about legal rights seems to assume that legal rights can be at odds with legal practice. Take a concrete example. I'm a member of the American Civil Liberties Union, and I strongly believe that there is some sense in which, in cases of e.g. indefinite detention without charges, the government violates detainees' constitutional rights. I don't stop thinking so when the ACLU loses the relevant legal battles, and obviously I don't regard my thoughts about this matter as being inconsistent. (Nor, presumably, would the most rabid Federalist Society-type defender of unlimited executive power to detain dangerous people

think I was being inconsistent *merely* because I refused to change my mind when the courts ruled against me. After all, if the same court cases go the other way, then presumably this partisan of the other side would continue to believe that the government had a right to continue to detain the dangerous people in question, even though the courts didn't agree.) What does it mean, though, to say that unconstitutionally detained people have a right to be freed, even if the courts do not recognize this right? It seems to me that, far from conflicting with the picture of the ontology of legal rights and obligations invoked earlier, this sort of case fits quite nicely with it. What I mean when I say that such detainees have a right to go free is that *according to the Constitution*, as I interpret it, they have a right to go free. If I'm angry at the judge, it's because, when engaged in his job of helping to determine what rights people have in practice, he has failed to fulfill his moral obligation to be fair and reasonable in his interpretation of the Constitution. None of this suggests, much less requires, that we postulate legal rights as abstract entities floating free of legal practice.

In the Secretary's Club case, it seems to me that neither those who wished to admit Fran nor those who wished to exclude her would have any case for thinking that, *if* the membership committee had fulfilled its moral obligation to fairly interpret the membership rules, they would have made the opposite decision. No possible, enforceable decision would have failed to violate the rules in this case. So much the worse for the usefulness of the club rules, but no threat to monaletheism lurks in the Secretary's Club's misfortune.

The Secretary's Club Paradox brings us very close to Russell's Paradox, which is the subject of the next section. The thing that stops us from positing an outright

contradiction in the former case is the practical impossibility of simultaneously allowing and preventing Fran from joining, and the unmotivated and problematic nature of postulating objective Secretary's Club rights that are constituted by something other than the actual practices of the Secretary's Club. Unfortunately, when it comes to set-membership rules that create similar problems, no such easy considerations are available to block the contradiction.

III. ($R = \{x: x \notin x\}$)?

According to naïve set theory, for every property, there is a set of objects that have that property. The Russell Set is the set of all sets that are not members of themselves. In the case of (almost) any given set, it is a simple and straightforward matter to determine whether or not it is a member of the Russell Set. For example, the set of all coffee cups is not a coffee cup, so it is a member of the Russell Set. The set of all sets is a set, which means that it is a member of itself, and hence it is not a member of the Russell Set.⁶⁵ The set of all books about set theory is not a book about set theory, so it is not a member of itself, and it is a member of the Russell Set. The set of all sets that have been used as examples in this paragraph has been used as an example in this paragraph, so it is a member of itself, and hence not a member of the Russell Set. So far, so good. When it comes to all the sets just mentioned, we seem to have no problem determining which of them are members of the Russell Set and which are not. We seem to be able to understand perfectly well what it is for a set to be, or fail to be, a member of itself, and perfectly able to use that knowledge, and the Russell Set's membership conditions, to sort out the members from the non-members.

⁶⁵ Remember, at this point we're explicating naïve set theory, and so assuming unrestricted comprehension.

The problem comes, of course, with the question of whether the Russell Set itself is a member of the Russell Set. If it is, it isn't. If it isn't, it is. Given the assumption that it must be one or the other, it must be both a member of itself and not a member of itself. Contradiction. Any temptation that we have at this point to postulate "set-membership gaps," so that the Russell Set can be seen as neither a member of itself nor a non-member of itself, should dissipate once we realize that, if the Russell Set is neither a member of itself nor not a member of itself, then by a single application of the Law of Double Negation, it is both a member of itself and not a member of itself. Contradiction again.

Given the validity of the reasoning from the existence of the Russell Set to its contradictory self-membership status, the difficulties involved in denying the existence of sets—even Quine, with his famous taste for desert landscapes, was forced to clutter up his ontology with set-theoretic realism—and the clear impossibility of sensibly denying that the presentation of the paradox is meaningful, this looks like a serious threat to monaletheism.

There is, of course, an orthodox response to this. Standard ZFC set theory rejects the Comprehension Axiom of naïve set theory (for any property, there's a set of objects that possess that property) and accepts the Axiom of Foundation. Instead of starting with some description, then gathering a set of everything that matches it, you can start with the empty set, or sets of concrete objects, then have the set of that set, and the set of the set of that set, and so on forever and ever, but chains of set-membership are never permitted to loop in on themselves. Paradoxical sets simply don't arise in this construction.

If you start by thinking about the problem Russell's Paradox poses for logical orthodoxy, this might seem like a terribly *ad hoc*, artificial solution that legislates the

problem out of existence instead of solving it any sort of serious sense. Mathematicians, however, may have good independent reasons to prefer to work with a cumulative hierarchy of sets, and philosophers should probably respect a certain division of expertise here. This, however, leaves us with the vexing problem of how to be sure that the mathematicians' hierarchy of choice exhausts the total universe of sets. The hierarchy could be an interesting set-theoretic construction, and mathematicians could have excellent reason to play exclusively in those fields, but there could be other sets "out there."⁶⁶ If so, how can we rule out the Russell Set?

Of course, many mathematicians and philosophers claim that the practice of doing ZFC set theory changes their set-theoretic intuitions, so that "set" seems to intuitively refer exclusively to sets in the hierarchy. We can grant them the right to assign the honorific "set" in this way, but that hardly solves the problem. Why should we think that the only *collections-of-things* that exist are "sets" in the ZFC sense? How, in particular, do we rule out troublesome entities like the Russell Collection-Of-Things?⁶⁷

One obvious point here is that if the Russell Collection-Of-Things exists and both includes and fails to include itself as a member, it's pretty clearly an abstract object. If we can make sense of set theory without having to appeal to abstract objects, that would have

⁶⁶ Other sets "out there" in the sense of actually existing, of course, not just "out there" in the sense of being postulated by some rival set theory. As Priest puts it, "It should be noted that whether there are any sets outside of the cumulative hierarchy is the central one. The agnostic position which allows that there may be sets outside of the hierarchy is quite compatible with the dialetheist position. It is not at issue that the hierarchy is an important and interesting set-theoretic structure. What is at issue is whether there are other sets...As solutions to the paradoxes go, there is really no extant alternative to the claim that the cumulative hierarchy exhausts the universe of sets." (Priest (2006a), p. 30)

⁶⁷ There is an extensive literature on the subject of proper classes, classes that are not sets. The problem, of course, is that if one postulates such classes, it's tricky to rule out a Russell Class. There are extant proposals for regimenting proper classes as well, setting out clear axioms for what proper classes are, but the problem now is that such regimented proper classes look like just another level of the set-theoretic hierarchy, and we still don't have a good, principled reason to rule out the existence of collections-of-things that are neither sets nor proper classes. The Russell Collection-of-Things still looms.

large and obvious advantages in terms of ontological simplicity, avoidance of Benaceraff-style epistemic concerns and so on.

David Lewis has given us just such a theory, in his ingenious paraphrase of (hierarchical) set theory in mereological terms, such that sets just *are* sums of concrete objects.⁶⁸ Given the attractiveness of this move on the grounds just mentioned, all of which are quite independent of paradox-avoidance, it looks like Lewis has given us an excellent, non-question-begging way to block the inference from Russell's Paradox to dialetheism.⁶⁹ Lewis' reservation about his own theory was that the physical world may not contain enough mereological atoms to reconstruct all the set theory needed for all of mathematics, and that philosophers should not try to revise mathematics on philosophical grounds.

Why, however, should we be mathematical realists at all? Why should we think that *any* sets exist? As Stephen Yablo puts it,

About fifty years ago, Quine convinced almost everyone who cared that the argument for abstract objects, if there was going to be one, would have to be *a posteriori* in nature. And it would have to be an *a posteriori* argument of a

⁶⁸ He develops this idea in Lewis (1991)

⁶⁹ A radical dialetheist could object that I am begging the question by assuming that the body of concrete objects with which I believe that all (real) sets are built up forms a consistent totality. What if there are contradictions in the physical world? I am not, however, trying to show that existing sets are free of contradictions in this sense, but only that the traditional set-theoretic paradoxes do not force us into belief in true contradictions. If it is plausible that all existing sets are built up out of concrete objects, then I have in fact shown this much. Now, apples and sets of apples, etc., may form an inconsistent totality if, for example, some apples are both red and not-red, but we can say two things about this. The first is that nothing in this scenario re-introduces problems about set-membership, from which the original claim of inconsistencies arising from set theory originated. Thus, the limited goal to which this solution sets itself (refuting the case for dialetheism *from the traditional set-theoretic paradoxes*) has been accomplished, even if it is conceptually possible that there are new and undiscovered cases of apparent contradictions left to be dealt with. Secondly, the fact that it is possible (in a weak epistemic sense of 'possible') that there are true contradictions involving things like apples does not give us the slightest reason to suppose that there actually are such contradictions. Until the hypothetical radical dialetheist has given us some sort of evidence that apples are inconsistent, it is not clear why we shouldn't be allowed to proceed as if they were not.

particular sort: an *indispensability* argument representing numbers, to use that example, as entities that ‘total science’ cannot do without.⁷⁰

In other words, if we should believe in any sets at all, the reason we should do so is because they are indispensable for our best science. We can’t paraphrase them out. However, not *all* of existing mathematics is necessary for science, and the universe might well contain enough mereological atoms to reconstruct the fragment we *do* need. If so, why be realists about the rest?⁷¹

Penelope Maddy, in the stage of her career in which she argued for mathematical Platonism, argued against the view that only those mathematical objects that are indispensable to science are real by claiming that it is totally contrary to the spirit of naturalism to expect mathematicians to have to wait “on the activities of the physics labs” to find out if their (mathematical) results are correct. They have after all, “justificatory practices of their own.”⁷² I argue, however, that this confuses two levels of analysis. The mathematician’s proper area of study is what is true in the interesting, imaginary mathematical world in which they work, and their own proof procedures suffice for that. (This is true even for inconsistent but interesting mathematical structures. It is certainly not true in the obvious sense that “the Russell Set is a member of itself and not a member of itself,” but if we read that statement as being prefaced by an implicit fiction operator,

⁷⁰ Yablo (2000) 275

⁷¹ I find the thought that we might need more mereological atoms than physical reality contains to reconstruct the sets necessary for the fragment of mathematics that’s indispensable to our best theories of that same physical reality deeply counter-intuitive, but given the sheer variety of mathematical constructions used in scientific theories and the magnitude of the problem of trying to sort out what can and cannot be paraphrased out, I don’t want to be over-confident about the point. As such, I develop a “Plan B” in Appendix A. (Of course, the Special Theory of Relativity is often taken to imply that time is continuous, such that between any two instants there is a third. This would give us an uncountably infinite number of bits of physical space-time, and it is hard to believe that we’d need more than that for a mereological reconstruction of the necessary mathematics. But, as mentioned above, at least on certain interpretations, the quantum results seem to point in the other direction on this question, so it’s probably best not to be over-confident about *that* either.

⁷² Maddy (1990), p. 21

then we can happily endorse it as true without thereby endorsing dialetheism.) How many of the objects in that world really exist, however—in other words, to what degree to which the mathematical world is a purely fictional construction, and the degree to which it overlaps with reality—is quite properly the domain of the empirical scientist, as is any question about the physical world.⁷³

Now, this suggestion—that our best reason to believe that the total universe of sets fails to contain the Russell Set is that we are only justified in being realists about the sorts of mathematical entities that our best-confirmed scientific theories are ontologically committed to—has a consequence that many of my fellow monaletheists might find extremely unattractive. So as not to look like I'm skirting the issue, it's worth pausing for a moment to spell it out, and to explain why I regard it not as a bug but as a feature.

While Priest usually likes to portray the intellectual adjustment involved in admitting the possibility of true contradictions as a smaller change than it might appear, one without sweeping consequences for our practices of reasoning—we'll come back to this point with a vengeance in Chapter Eight, when we consider his “classical recapture”—in other, more enthusiastically optimistic moods, he likes to speculate about the wide-ranging consequences that the jump to the “realm of the transconsistent” might have.

In modern science, the inferentially sophisticated part is nearly always mathematical. An appropriate mathematical theory is found, and its theorems are applied. Hence, a likely way for an inconsistent theory to arise now in science is via the application of an inconsistent mathematical theory. Though the

⁷³ Obviously, I'm assuming here that the only argument for the existence of mathematical objects we should find compelling is indispensability. If someone constructs an absolutely devastating argument for the existence of such objects that is independent of indispensability—i.e. that shows that (a) there is a *non-physical* portion of reality, and (b) that we are in contact with this domain in some other way than through empirical science, then my entire line of thought here is wrong. As indicated above, we should be fallibilists about everything, empiricism included. That said, the possibility that a view could be refuted is a very different thing from an actual refutation.

construction of inconsistent mathematical theories (based on adjunctive paraconsistent logics) is relatively new, there are already a number of inconsistent number theories, linear algebras, category theories; and it is clear that there is much more scope in this area. The theories have not been developed with an eye to their applicability in science—just as classical group theory was not. But once the paraconsistent revolution has been digested, it is by no means implausible to suggest that these theories, or ones like them, may find physical application—just as classical group theory did. For example, we might determine that certain physical magnitudes appear to be governed by the laws of some inconsistent arithmetic, where, for example, if n and m are magnitudes no smaller than some physical constant k , $n + m = k$ (as well as its being the case that $n + m \neq k$). There are, after all, plenty of episodes in the history of science where we came to accept that certain physical magnitudes had somewhat surprising mathematical properties (being imaginary, non-commuting, etc.). Why not inconsistency?⁷⁴

The suggestion may seem absurd, but consider the analogy of physical geometry.

For thousands of years, pretty much everyone was what we might think of as a “Euclidean monist,” meaning that they took it as obvious that there could only be one shortest path between any two points, and that they understood this not just as a claim about the particular hypothetical realm of one axiomatic system but as a (*a priori* knowable) claim about physical reality as well. Even when non-Euclidean geometries started to be developed, they were largely regarded as esoteric curiosities that couldn’t possibly model anything in the real world. If a student went to Immanuel Kant’s table at his tavern in Konninsberg and suggested to the great man that future science might plausibly one day falsify Euclidean monism and allow for multiple non-equivalent shortest paths between two physically real points, the young man would have presumably been laughed out of the place. Yet, history has rendered the opposite verdict; Einstein has shown that the true geometry of space-time is non-Euclidean, and so much the worse for our spatial intuitions.

⁷⁴ Priest (2006), p. 150

Of course, for the student to demand without special evidence or argument that Herr Professor Kant accept that this was possible (particularly in any richer sense than *epistemically* possible, where “epistemically possible” is taken to mean something like “having a status such that one should keep an open mind about it”) would be to beg the question against Euclidean monism, and the good professor would have been fully rationally entitled to decline to do so. Similarly in this case, if we take Priest’s rhetorical question as a demand that we (monaletheists) take his wild speculation about inconsistent magnitudes seriously as a possibility. The problem, of course, as the Einsteinian case shows, is that the lesson of the history of science seems to be that even intuitively well-grounded claims about impossibility can be falsified, and that we shouldn’t be dogmatically closed off to the (epistemic) possibility that we’re wrong about (logical and mathematical) possibility. Priest lays out a fanciful sort of scenario where monaletheist assumptions about logical and mathematical possibility are in fact falsified by empirical research.

Let us suppose that come to predict a collision between an enormous star and a huge planet. Using a standard technique, we compute their masses as x_1 and y_1 respectively. Since masses of this kind are, to within experimental error, the sum of the masses of the baryons (protons and neutrons) in them, it will be convenient to take a unit of measurement according to which a baryon has mass 1. In effect, therefore, these figures measure the number of baryons in the masses. After the collision, we measure the mass of the resulting (fused) body, and obtain the figure z , where z is much less than $x_1 + y_1$. Naturally, our results are subject to experimental error. But the difference is so large that it cannot possibly be explained by this. We check our instruments, suspecting a fault, but cannot find one; we check our computations for error, but cannot find one. We have a puzzle. Some days later, we have a chance to record another collision. We record the masses before the collision. This time they are x_2 and y_2 . Again, after the collision, the mass appears to be z (the same as before), less than $x_2 + y_2$. The first result was no aberration. We have an anomaly.

We investigate various ways of solving the anomaly. We might revise the theories on which our measuring devices depend, but there is no obvious way of doing this. We could say that some baryons disappeared in the collision;

alternatively, we could suppose that under certain circumstances the mass of a baryon decreases. But either of these options seems to amount to a rejection of the law of conservation of mass (-energy), which would seem to be a rather unattractive course of action.

Then someone, call them Einquine, fixes on the fact that the resultant masses of the two collisions were the same in both cases, z . This is odd. If mass has gone missing, why should this produce the same result in both cases? An idea occurs to Einquine. Maybe our arithmetic for counting baryons is wrong. Maybe the appropriate arithmetic is one where z is the least inconsistent number, and p (the period of the cycle) = 1. For in such an arithmetic $x1 + y1 = x2 + y2 = z$, and our observations are assumed without having to assume that the mass of baryons has changed, or that any are lost in the collisions! Einquine hypothesizes that z is a fundamental constant of the universe, just like the speed of light, or Planck's constant.⁷⁵

The story goes on, but the general idea should be clear enough. We have a localized change of the arithmetic assumed by our best science, one that doesn't require us to change the way we count match-sticks, any more than the non-Euclidean curve of space-time stops us from making Euclidean assumptions about how to make geometrically complicated bank shots when we play pool, but which is just as deadly to monaletheism as Einstein's discovery was to Euclidean monism.⁷⁶ Of course, the details of this story could doubtless be nit-picked in many ways, but I very much doubt that someone in the late eighteenth century trying to imagine a way that science could falsify Euclidean monism could do any better. It's in the nature of scientific revolutions that, before they come along, they're not only unpredictable but often almost completely

⁷⁵ Priest (2006), p. 162-163

⁷⁶ In his response to Denyer (1995)'s argument against the possibility that arithmetic could be inconsistent in this way, Priest says that "most people are not disposed to assert the existence of a number greater than or equal to all numbers. But this simply reflects a belief they have acquired during their education." (Priest (1996), p. 657) I think Priest is considerably over-playing his hand considerably here on the question of the counter-intuitiveness of the idea. On the other hand, I think Priest *would* have been well within his rights to dismiss the relevance of popular opinion to his debate with Denyer. As a small bit of anecdotal evidence, when I've taken classrooms full of college undergraduates through Cantor's diagonal proof that not all transfinite sets have the same cardinality, my consistent experience has been that most of them initially dismiss the thought that some infinite collections could be larger than others as obvious nonsense, and that even after seeing the proof, many of them try their best to find ways to get around the result. Folk intuitions don't count for much when it comes to complicated mathematics, and they shouldn't.

unimaginable. Similarly, at the level of detail that's Priest's given us, there's no way to be sure whether (even in the extreme, strange, counterpossible hypothetical situation being considered) revising the underlying arithmetic of our theory from a classical to a paraconsistent one would be the most rational response. As per our discussion at the end of the introduction, recalcitrant evidence can always be taken as evidence against many different parts of our overall package of beliefs, and any given proposal for belief-revision has to be reasoned out on the specifics of the case, specifics that are unknowable in consideration of these sorts of extreme hypotheticals. So we can't be too confident in advance that, even in this specific scenario, the best scientific move would be the one Priest hypothesizes.

That said, *if* Priest's scenario came true, and it turned out that the best, most reasonable response was to formulate a scientific theory to which inconsistent mathematical entities were indispensable, then, by my lights, we should be realists about such entities, just as we have a good reason to be realists about as much set theory as we need to make sense of our best current science. Thus we have the potentially unattractive consequence mentioned earlier: my solution to Russell's Paradox leaves the empirical/mathematical back door open for dialetheism. *If* the empirical results were to go in certain extremely unexpected directions, then inconsistent mathematical results *would* give us an excellent reason to abandon monaletheism.

As I said, this strikes me as not a bug but a feature of my solution. A belief that no evidence would ever dislodge is a dogma, and one doesn't need to subscribe to all the details of Karl Popper's epistemically impoverished theory of justification to agree with him that one of the great lessons of the last few hundred years of scientific progress is

that unfalsifiability is a mark of theories that don't deserve to be taken seriously. If monaletheism is intuitively compelling, there are no good arguments for dialetheism, and the consequences of accepting dialetheism are unappetizing, then we're within our rational rights to retain the monaletheist structure of our beliefs, but we should always keep the door open for the world to push back against those beliefs and show us that we're wrong.

Of course, we don't have a right to even this limited conclusion until we've satisfactorily blocked the single strongest argument for dialetheism, which is the argument from the Liar Paradox. It is to this, then, which we now turn.

Chapter Four: Liars and Gluts

Take the following claim:

(1) The sentence marked as (1) is false.

This is a version of the Simple Liar, the first of several Liars we will encounter.

The issue of how to treat such sentences is one of the oldest and hardest problems in the philosophy of logic, one which the apostle Paul refers to in the New Testament when he talks about the prophet of Crete who says that all Cretans are liars.⁷⁷ On the face of it, it doesn't seem like (1) could have any sort of consistent truth-value. If (1) is true, then it is false, and hence not true. Contradiction. If (1) is false, then it is true, since that's exactly what it asserts itself to be, and if it's true, then it follows that it's not false. Contradiction again.

Graham Priest's analysis of this problem is simplicity itself. It *seems* as if all roads lead to inconsistency because they *do*. The Liar really is both true and false. Thus, by the law of existential generalization (valid whether one's logic is classical or paraconsistent⁷⁸) some sentences are truth-value "gluts," both true and false.

This is, at least, a *prima facie* sound argument for dialetheism. The premises all *look* true, and the reasoning *looks* impeccable. The conclusion may be hard to swallow but, after all, as Intro to Philosophy and Beginning Logic students learn at the beginning of every semester at colleges and universities across the world, part of what it means to

⁷⁷ Titus 1:12

⁷⁸ Unless, perhaps, one's logic is both paraconsistent and free, since there are free logicians who don't regard existential generalization as a valid inference. However, the *reason* that they don't regard the inference as *universally* truth preserving—that we don't want, for example, to infer “there exists an X such that X is a detective” from “Sherlock Holmes is a detective”—doesn't apply here, unless one believes that statements like (1) are non-existent objects.

be a rational person is that, when confronted with an argument from premises you believe to a conclusion that you don't believe, you either revise your belief in those premises, show that the conclusion doesn't really follow from them or acknowledge that the conclusion is true after all. Those are the only options, and it looks extremely difficult, in the case of the argument for dialetheism from the paradoxes, to either show that any of the explicit or hidden premises (like "a sentence that says of itself that it's false is true iff it is false") are mistaken, or that there's anything wrong with the reasoning.

While there is not (yet) any hard data on what nonphilosophers' intuitions on Liar-like sentences really are—this looks like an interesting task for the burgeoning field of experimental philosophy, or "x-phi," given that the literature on the Liar paradox is full of empirically unsubstantiated claims on what "seems" to be the case about it to most nonphilosophers—anecdotally, my impression is that the average nonphilosopher, when confronted by a sentence like (1), reacts as if they are being presented by a clever sophisticated trick they haven't quite figured out and, depending on their level of interest in game-playing, either their curiosity is piqued or (more commonly) they dismiss it as a silly diversion. One way or the other, the common thread is a belief that "there's something wrong with it, but I'm not sure what."

Philosophers, professionally engaged in the search for truth, don't have the luxury of dismissing the argument in this way. We must either show where the argument goes wrong, or, if we are unable to do so, we must reluctantly accept the conclusion, and accept that just as Einstein showed us that some of our basic intuitions about the nature of time were mistaken, Graham Priest has shown us that some of our basic intuitions about the relationship between truth and falsehood are mistaken.

Or so one would think. Interestingly enough, at least one philosopher, Thomas Hofweber, has argued that the apparent options of rejecting a premise, demonstrating a flaw in the reasoning or embracing the conclusion *aren't* jointly exhaustive of the rational options available to us when it comes to the Liar Paradox. Hofweber believes that we can believe all of the premises of the argument, accept as valid all the rules of inference that are employed to get from those premises to the conclusion, even accept that to say that a rule is valid is to say that is truth preserving, but, for all that, still be rationally entitled to decline to accept the conclusion.

This would be quite a trick if Hofweber could pull it off. To get a sense for how this move supposed to work, start by thinking about the logic of generics. Quite a bit of natural language reasoning about the relationship between categories doesn't start with anything as precise as either universal generalizations ("all members of category S are members of category P") or existential ones ("some members of category S are members of category P"), but with "generic" statements like "members of category S are members of category P." Generic statements can't be reduced to either of the more logically precise categories. Take Hofweber's example, "Bears are dangerous."⁷⁹ Unlike "all bears are dangerous," this statement is not falsified by the existence of a single old, toothless bear, and unlike "some bears are dangerous," the dangerousness of one bear would not suffice to make it true...after all, if it did, then "bears are not dangerous" would be equally true.

Generic statements seem to rationally license non-monotonic "default reasoning." As Hofweber puts it,

⁷⁹ Hofweber (2007), p. 148

Such inferences are ones that one is entitled to make unless one has overriding information. So, I know nothing about Freddie except that he is a bear and I know (2) [which is “bears are dangerous”] then I am entitled to infer that Freddie is dangerous. But if I learn in addition that Freddie is old and lost all his teeth I am not entitled to make that inference any more....More information can make an otherwise appropriate inference inappropriate.

So far so good, but how does this help us out when it comes to the Liar Paradox? None of the premises of the argument sketched above from the Liar’s truth conditions to dialetheism seem to be generic statements. Also, remember, Hofweber is willing to concede that the premises are all true, that the inference is a valid one *and that valid inferences are truth preserving*. That last point is the key one.

What seems to be distinctive of inference rules in deductive logic is that they are truth preserving, and above we considered it to be a defining feature of valid inference rules that they are truth preserving.

(1) Inference rules are valid if they are truth preserving.

This is hard to deny, and we should not deny it, properly understood. In fact, (1) has two readings. One of them, which we could call the strict reading, requires that each instance is truth preserving. But (1) also has a generic reading. The right hand side has a generic reading which is nicely brought out by using the plural. Inferences are truth preserving according to the generic reading just like bears are dangerous. That doesn’t mean that each and every instance is truth preserving, just like the latter doesn’t mean that each and every bear is dangerous. But nonetheless, (1) is literally true on this reading.⁸⁰

This is all terribly clever, and it allows Hofweber to get off the dialetheist hook (and, in the case of Curry’s Paradox, which we will consider later, get off the trivialist hook) without sacrificing much of anything.⁸¹ He gets to keep his belief in all the premises, he gets to keep regarding all the relevant inference rules as valid, and he even gets to continue to assert that valid inference rules are truth preserving. There’s only one problem. Sadly, that problem is a fairly important one.

⁸⁰ Hofweber (2007), pp. 149-150

⁸¹ While dialetheism is the position that *some* contradictions are true, trivialism is the position that *all* contradictions are true. As we’ll see, Curry’s Paradox seems to threaten the latter, more extreme, conclusion.

Taken as a response to an opponent (whether a real dialetheist or a hypothetical trivialist) who believes that at least some contradictions are actually true, the argument that valid inferences are not always truth-preserving, and that we are entitled to conclude that the particular valid inferences in question are not truth preserving precisely because they take us from true premises to conclusions that cannot be true (because contradictions are never true) completely, utterly and transparently begs the question.⁸²

Similar considerations rule out what may be the best-known “solution” to the Liar Paradox, Tarski’s proposal that we can save consistency not for natural languages but for restricted and regimented artificial languages, where no language is allowed to contain its own truth predicate. At the bottom level, we have the “object language,” which does not contain words like “true” or “false” at all. Next we have the meta-language, which contains such words, but where they can only be applied to object language sentences.

For example, take sentence (2), below, which is in the object language, and (3) and (4), which are both in the metalanguage.

- (2) Cats believe in ghosts.
- (3) It’s true that “cats believe in ghosts.”
- (4) It’s false that “cats believe in ghosts.”

⁸² One could object that in making this accusation, I’m at least partially putting Hofweber’s solution to the paradoxes into the context of refuting dialetheism and that that may be exegetically unfair. After all, Hofweber’s main concern is not avoiding dialetheism but avoiding triviality, which he calls “the Great Collapse” of all possible reasoning (although he pretty clearly takes his solution to work for the ordinary Liar-like cases as well as Curry, and hence takes it to preserve consistency as well as non-triviality). He briefly alludes to Priest (see Hofweber (2007), p. 146) as one of many “Plan A” solvers of the paradoxes, Plan A being the plan where we show that there’s something wrong with the argument from the paradoxes to triviality, with Priest’s solution being that the ‘something wrong’ is *ex contradictione quodlibet*. Aside from that, he doesn’t mention Priest, or dialetheism, for the rest of his essay. I would say, however, that (a) as we will discuss in Chapter 7, begging the question against the trivialist is just as bad as begging the question against the dialetheist, and (b) his solution is being examined here for the purposes of seeing whether we can use it to avoid being rationally forced to surrender to Priest’s case for dialetheism. The fact that his argument *would* be question-begging is awfully relevant to our reasons for examining it, regardless of whether or not it is relevant to Hofweber’s purposes.

Now, speaking the metalanguage, we can say what we want to say about object language sentences like (2), but what if we want to express our disagreement with a metalanguage sentence like (3)? If we were allowed to do that in the metalanguage, then the metalanguage would contain its own truth predicate, and we could construct a metalanguage sentence like (1) and the problem wouldn't have been avoided at all. As such, we have to move up to the meta-meta-language, where we can say things like (5), below.

(5) It's false that "it's true that 'cats believe in ghosts.'"

If we want to talk about the truth-value of (5), we have to do so in the meta-meta-meta-language and so on. No sentence like (1) will ever be constructable at any point in the hierarchy.

Now, quite apart from its adequacy as a way to save the consistency of our beliefs from the problem posed by the Liar Paradox, there are many strange and suspicious things about this construction. For one thing, there are a lot of statements that seem intuitively true—indeed, statements whose apparent truth seem to be the selling points of Tarski's construction—that we can't say in any of his artificial languages. For example, take (6) and (7), below.

(6) No sentence anywhere in the hierarchy says of itself that it is false.

(7) No sentence anywhere in the hierarchy is both true and false.

These sentences are ones that one would naturally make in trying to convince someone to accept Tarski's solution to the Liar Paradox, but they are necessarily banned from any language in the hierarchy for precisely the same reason that paradoxical

sentences are. After all, (6) and (7) try to talk about the truth or falsity of sentences of every level of the hierarchy, including their own.⁸³

This should certainly be an embarrassment for the theory, but it gets worse. Even if we've saved consistency *for these new, artificial languages*, what about natural languages, where paradoxical sentences *can* be constructed? (1) doesn't appear in any language of the hierarchy, but it does appear in English. Is it true, or false, or what?

Uncharitably, one might think that Tarski's proposal amounts to a sophisticated philosophical version of closing one's eyes, putting one's hands over one's ears and yelling "I'm not list-en-ing!" Rather than talking about languages in which the problem arises, we simply shift to only thinking about languages in which it doesn't.

One step better than this, we could say that there's something wrong with the sentences, such that although they are grammatically "constructable" in English, they are defective or meaningless in some way such that they are properly excluded from more carefully constructed artificial languages. Perhaps the problem has something to do with self-reference—use and mention just can't be combined in the way they are in sentences like (1).

Tarski, of course, never so much as hints at this sort of view, but since he doesn't give us a clear view of his own on this issue, it's worth trying it out as a possible rational reconstruction of his position. In any case, however, there are at least four problems with this suggestion. We've already seen the first—that quite a few sentences seem to be

⁸³ Of course, this should not be confused with an accusation that Tarski cannot prove the consistency of all of the languages of his hierarchy. He was, after all, well aware of Gödel's results, and so this latter accusation would not have bothered him. The problem is a much more serious one, first that Tarski can't even *assert* consistency (which is quite a different thing from being unable to *prove* it, and worse yet, that Tarski cannot even say of his hierarchy that it isn't vulnerable to the *specific source* of inconsistency in natural languages that motivates the construction of the hierarchy as an alternative to those languages.

excluded that are, far from being meaningless, precisely the sorts of things one might say in support of Tarski's solution, like (6) and (7), above, because they are self-referential in the banned way if said in any language of the hierarchy.

The second problem is that we seem to be able to construct perfectly meaningful (and non-paradoxical) sentences that are self-referential even when they are said in English, like (8) and (9) below.

(8) This sentence has seven words in it.

(9) This sentence has twelve words in it.

(8) seems to be meaningful and clearly true, and (9) seems to be meaningful and clearly false. Thus, at the very least, being self-referential isn't *sufficient* for meaninglessness. Perhaps the gist of the suggestion that self-reference is the problem can be saved, if the real problem is not *just* self-reference but that (1)—unlike (8) or (9) self-referentially ascribes a *semantic* fact to itself. This won't work either. Consider sentence (10).

(10) This sentence is meaningless.

On pain of contradiction, that sentence can't be meaningless. After all, if it's meaningless, it's *true*, and hence *meaningful*.

The third problem is that we can construct a Liar-like paradox without self-reference. Consider the following pair of sentences:

(11) Sentence (12) is false.

(12) Sentence (11) is true.

(11) and (12) are both true iff they are false, just like (1) and neither is about themselves. It could be objected that they are both *indirectly* self-referential. Fair enough.

It is possible, however, to construct a Liar-like paradox without even indirect self-reference. Consider the following infinite series of sentences.

(Y1) All of the sentences ranked (Y2) or above are false.

(Y2) All of the sentences ranked (Y3) or above are false.

...and so on, such that every sentence in the series is of the format:

(YN) All of the sentences ranked (N+1) or above are false.

This is of course a version of Yablo's Paradox, and it is just as apparently impossible to attribute a consistent truth-value to *any* sentence in this series. Assume that any sentence in the series is true. Let's call a particular such sentence (YX). If (YX) is true, then all the sentences after (YX) are false, including (YX+1). Since (YX+1) says that all the sentences ranked (YX+2) or above are false, for (YX+1) to be false, at least one sentence that comes after (YX+1) must be true. The problem is that all of the sentences that come after (YX+1) are *also after* (YX). Therefore, if at least one sentence that comes after (YX+1) is true, at least one sentence that comes after (YX) is true, and as such (YX+1) is false, and hence not true, so it is both true and not true. Contradiction.

If, on the other hand, if we assume that (YX) is false, then it follows that at least one sentence that comes after (YX) is true. Let's call it (YY). If (YY) is true, then (YY+1) is false, which means that at least one sentence that comes after (YY+1) is true, which means that at least one sentence that comes after (YY) is true, which means that YY is false, and hence not true, so it is both true and not true. Contradiction again.

Of course, technically speaking no Yablo-type paradox can be constructed using the languages of Tarski's hierarchy, since no truth-value-talk is allowed in the object language, and in the languages above it, truth-value-talk is only allowed to concern

languages below it in the hierarchy. (Notice again, by the way, that the previous sentence is un-constructable in any language of Tarski's hierarchy.) You could construct an infinite series of sentences, each on a different artificial language level, where each one said that the sentences *below* them were all false, but this series would necessarily "ground out" in Kripke's sense (discussed below), since the series would bottom out in an object language sentence not containing words like "true" or "false." However, the existence of such paradoxes seems to show that the self-referentiality of sentences like (1) can't be the source of the problem.

The fourth and to my mind the most serious problem with the suggestion that the alleged meaningfulness of self-referential sentences gives us a principled reason to justify their exclusion from Tarski's hierarchy of artificial languages is that if this move succeeded in saving the Tarski solution from the appearance of *ad hoc*-ness, it would do so only at the price of making the whole complex and awkward maneuver of shifting our reasoning into the framework of new, artificial languages completely unnecessary. We would have solved the paradox within the framework of natural language reasoning.

There is another way that we could think about what Tarski is up to without either accusing him of hands-over-the-eyes avoidance of the paradox, or of the unworkable attempt to claim that self-referential sentences are meaningless. What, after all, does he mean when he claims that natural languages are "inconsistent" but his favored artificial languages are not? It can't simply be that inconsistent theories can be formulated in natural languages. After all, we can certainly make inconsistent claims in the languages of his hierarchy. (For example, there's nothing stopping us from claiming in the meta-language that two mutually inconsistent object-language are true, or from claiming in the

meta-meta language that a meta-language sentence that says of some object language that it is false is true *and* that a *different* meta-language sentence that says of the *same* object language sentence that it is false is also true. We can also presumably express more complicated and plausible-sounding theories about what sorts of sentences are true in lower-level languages that would imply such inconsistencies.) What if, on the contrary, calling a language inconsistent means that some (or, on the assumption that classical logic is the One True Logic even despite this, *all*) contradictions are true in it?

We could think of Tarski as being a kind of early dialetheist, albeit of a very different stripe than the sort we are familiar with today. Contemporary dialetheists try to isolate and contain contradictions by rejecting classical logic in favor of some paraconsistent logic, like Graham Priest's Logic of Paradox (LP). Of course logics like LP didn't exist at the time that Tarski was writing—so he wouldn't have been aware of paraconsistency as an option—but we can think of Tarski as pursuing a different strategy toward the same end of acknowledging, isolating and containing contradictions.

This dialetheist Tarski can be seen as, while accepting classical logic as the correct account of the logical consequence relation (for every language), relativizing truth to languages. Thus, (1) is, for *this* Tarski, both true and false *in English*, which means that everything is true-and-false-in-English by the explosion of inferences that are derivable from contradictions, but it is not true-in-the-object-language, or true in any other artificial language of his hierarchy, for the simple reason that no such sentence is constructable in any of those languages.⁸⁴ Thus, since no contradiction is true in any of

⁸⁴ Of course, on this Tarskian-dialetheist view, “there are some sentences in the object language that are both true and false” is a true sentence *of English*, as for that matter is “all sentences of the object language are both true and false,” but neither of these are true sentences of any language of Tarski's hierarchy.

these languages, not everything need be true in any of them, and we can non-trivially reason about science and other useful subjects in such languages.

If this interpretation is correct, then Tarski is not “solving” the Liar Paradox in the sense of trying to avoid the inference from any possible truth-value of sentences like (1) to the truth of at least some contradictions. Rather, Tarski accepts that inference.

Doug Patterson strongly rejects any reading of Tarski according to which Tarski thinks that some natural language sentences are both true and false. Instead, he attributes to Tarski his own view that the Liar Paradox shows us that natural language sentences don’t mean what we take them to mean, because if they did, some contradictions would be true. To understand English or any other natural language, according to Patterson’s theory, is to be in a sub-doxastic state such that one processes according to a false theory of what sentences mean, such that, for example sentence (1) seems to claim that it is not true. Of course, the delusion that the sentences of natural languages have the meanings they seem to is pervasive and, in a certain way, unbreakable.

Understanding natural language is a matter of modularized, sub-doxastic cognition. On the basis of this cognition speakers have a broadly perceptual understanding of natural language. Sentences of natural language, in context, seem to speakers to have certain meanings, and their so seeming disposes speakers to believe that they have those meanings. The perceptuality and sub-doxastic character of understanding is registered, for instance, in common concepts of *fluency* in a natural language, on which as long as one must explicitly think about what things mean, one doesn’t fully understand. This is why, even in the face of the paradoxes, and even in the face of acceptance of the present view of the paradoxes, it can continue to seem that there are sentences that “say of themselves” that they are untrue. This also why, although I am an inconsistency theorist, on my view I continue to understand English: I am still in the sub-doxastic state and so, like one in the presence of a visual illusion, though I no longer believe things to be a certain way, they still seem that way to me.⁸⁵

⁸⁵ Patterson (2007), p. 587

In other words, to understand English (or, like Tarski, Polish) is to have deeply internalized a false theory of what sentences mean. We know that it's false because, *if it were true*, sentences like (1) would be true iff they were false, and thus contradiction would arise either way.

On the other hand, there's no reason why we can't accept that sentences of the artificial languages in Tarski's hierarchy mean just what we think they do—the subdoxastic state we'd be in when we understood them would presumably be based on a theory that hadn't been falsified by paradoxes. Patterson's Tarski's solution to the Liar Paradox is, in the context we're considering—how to defend monaletheism against the argument from paradoxical sentences to the conclusion that some contradictions are true—a total non-starter. It totally, utterly and transparently begs the question against the dialetheist. To be rationally entitled to reject their conclusion, we can't just start by assuming that they're wrong and act as if this entitles us to infer any strange thing we want to about further substantive issues without giving our new views independent motivation. We are doing precisely that, and in a fairly extreme way, if we jump to the conclusion that all speakers of all natural languages are in the thrall of a pervasive illusion caused by their acceptance of the false theory that “sentences mean what they seem to mean,” and act as if this strikingly radical conclusion is sufficiently established by the fact that if our “theory” of what sentences mean were true, some contradictions would be true.

What we need to do instead is to give a principled reason why something is wrong with the dialetheist's argument. This reason needs to come in the form of a demonstration that either their conclusion doesn't really follow from their premises, or that one of their

premises is wrong. Gilbert Harman has made one such proposal in his book *Change in View*. He thinks the faulty premise is the assumption that sentences like (1) are true iff they are false.⁸⁶ This assumption is simple instance of the general Biconditional Truth Schema—“p” is true iff p—and our assumption that this universally holds is, according to Harman, precisely what gets us into trouble. In language reminiscent of Hofweber’s gloss on the claim that valid inferences are truth preserving, Harman suggests that we don’t completely throw it out—after all, in the great majority of cases its instances seem to be unproblematically true—but rather that we don’t see it as holding universally.

In practice the best solution may be to retain the Biconditional Truth Schema and yet avoid contradiction by interpreting the Schema not as something that holds without exception but rather as something that holds “normally” or “other things being equal.” It is then a “default assumption.” One accepts any given instance of the Biconditional Truth Schema in the absence of a sufficiently strong reason not to accept it. One does not apply the Schema to (L) because doing so leads to a contradictions.

This does not seem to be a satisfactory solution from the point of view of logic, since we take logic to require precise principles with precise boundaries, not principles that hold merely “normally” or “other things being equal.” But in ordinary life we accept many principles of this vaguer sort.⁸⁷

In other words, if the Truth Schema was a strict logical principle that held in all circumstances, the Liar would be true iff it was false, and that can’t be right, since nothing is ever simultaneously true and false. So much the worse, Harman thinks, for the place that logically precise principles are often held to have in the structure of good reasoning. Hence, although the Schema certainly seems to hold in all other cases, it must have exceptions.⁸⁸ He doesn’t give us any indication of what the Liar’s truth conditions

⁸⁶ Of course, many other philosophers have come to the same conclusion as Harman. I pick him just because his explanation of the motivation for his view is particularly striking for the purposes of elucidating why this sort of solution can’t work as a defense against the dialetheist’s argument.

⁸⁷ Harman (1986), pp. 16-17

⁸⁸ One might think that *any* classical logician who uses a truth predicate might have to say something along these lines. Certainly, a great deal of the current literature—from deflationist dialetheists and deflationist “paracomplete” theorists, and also from certain classical logicians—assumes that the paradoxes force us to

really are, if they aren't what they seem to be. It's hard not to suspect that this is because this question is unanswerable. There is no remotely plausible story to be told about what would make the Liar true other than its falsity.

At any rate, putting Harman's solution in the context of our attempt to find a principled reason to reject the dialetheist's argument from the Liar Paradox, it is a non-starter.⁸⁹ Like Hofweber's solution, or Patterson's Tarski's solution, Harman's solution begs the question.

A certain kind of orthodox logician might take all of this as evidence that arguing with dialetheists is a fool's game—after all, what better evidence could we have for *anything* than that it saves us from contradictions? If that begs the question against our argumentative adversaries, maybe those aren't people worth arguing with.

I think that this is fundamentally wrong-headed for at least two reasons. The first one has already been sketched out, above—the dialetheist *looks* like she has a sound argument for her conclusion. As such, we *have* to argue with dialetheists—engage their argument, show what's wrong with it—if we don't want to embrace the only other rationally permissible option, which is to simply accept that they are right.

choose between (a) keeping a truth predicate that behaves in the expected way, validating the T-Schema, unrestricted Capture and Release, etc., at the expense of classical logic, or (b) keeping classical logic, but either doing without a truth predicate entirely (perhaps replacing it with some sort of 'consistent successor concept') or (especially if the universal intersubstitutivity of equivalents is kept around) restricting the T-Schema. All I can say that I'm at least one classical logician who rejects the claim that we have to make this choice, and I hope to show in Chapters Five and Six how, given a plausible, independently motivated picture of truth and meaning, we can and should simultaneously keep classical logic, the T-Schema and unrestricted Capture and Release behavior even in the face of the paradoxes.

⁸⁹ Of course, like Tarski, Harman doesn't discuss dialetheism, but even without a dialetheist opponent to beg the question against, if we take *any* consistent solution to the Liar Paradox as an attempt to save the consistency of our beliefs from a possible problem, then simply assuming that inconsistencies can't be true does start to look like begging the question in the secondary sense that the paradox-solver *assumes* what he should be trying to *prove*. And, of course, the reason why we are interested in these solutions here is the hope that one of them will get us off the dialetheist's hook. Harman fails to do that for the same reason that Hofweber and Patterson's Tarski failed.

The second problem is that worrying about begging the question against the dialetheist forces us to do something that would be useful and valuable even if we weren't worried about the challenge posed to logical orthodoxy by dialetheism, which is to ground our solutions to the paradoxes in reasons that are independently plausible, having something going for them other than the fact that they allow us to preserve consistency. After all, if we are willing to throw enough of our beliefs into the fire for the cause of "saving truth from paradox," there many different possible ways to preserve the consistency of our beliefs. (Consider that the simplest and most easily constructed internally consistent set of beliefs is an empty one.) Revising our relevant beliefs to avoid inconsistency in such a way that there are *other reasons* motivating the revision *besides* preserving consistency should, at the very least, increase the chances that our new beliefs will be not only consistent but also true.⁹⁰

With this in mind, let's see if we can do better than the approaches considered so far.

⁹⁰ ...and *informative about subjects we're interested in*, of course. If, when I talk about the chances of our beliefs being true, this is just taken to mean "the chances that our belief set won't include very many falsehoods," this is, of course, best satisfied by not believing anything. If it's taken to mean "the chances that our belief set won't include very many falsehoods *and* that it will include lots of truths," then the best route would be to try to fill our belief set with lots of tautologies. What I mean is something very roughly along the lines of "the chances that our belief set will include beliefs that capture the truth about all sorts of subjects, such our knowledge of how things actually are is broad and deep and interesting."

Chapter Five: Liars and Gaps

Given that regarding sentences like (1) as true leads to contradiction, as does regarding it as false, an obvious “out” would be to reject the premise that truth and falsity are jointly exhaustive of the options. Perhaps, in some sense, there are “truth-value gaps,” and we should accordingly reject classical logic not in favor of some paraconsistent logic but in favor of some many-valued logic where the additional truth-value (or additional truth-values, if we end up with more than three values) is something other than the conjunction of the original two.

The phrase “truth-value gap” is a bit unfortunate here, since it gives us the impression that some sentences, or propositions, or whatever one takes truth-bearers to be, *don't have truth-values*, rather than there being some third value that's incompatible with either of the two traditionally recognized truth-values. Attempts to formalize this notion often leads to the unfortunate image of truth tables with G's, for “Gappy,” along with the T's and F's, which suggests that “not having a truth-value” (which is what's suggested by the word “gap”) *is itself a truth-value*. Whether or not we ultimately judge the “gap” proposal to succeed in saving consistency, it seems unfair to prejudice that discussion from the start by describing the proposal that some truth-value-bearers are neither true nor false in a way that sounds flagrantly inconsistent.

Sadly, the word “gap” is too entrenched in the literature for there to be much hope of trying to convince anyone to use a better word, I have no better word to offer in any case, and the existing language has the benefit of giving us a heuristically useful shorthand whereby we can on the one hand refer to the proposal that some sentences have

a third truth-value that's neither true nor false as the claim that there are "gaps" and on the other hand refer to the dialetheist's proposal that some sentences are both true and false as the claim that there are "gluts." As such, in everything that follows I will swallow my reservations about the terminology and go on referring to the "neither true nor false" idea as the theory of truth-value gaps, and the corresponding solution to the Liar Paradox as the idea that Liar sentences are "gappy."

So, *does* this solution work as an analysis of (1), above?

Well, we can say quite a lot in favor of the gap theory that can't be said for any of the other approaches we have considered so far. When we say that (1) is gappy, we aren't doing anything transparently ad hoc, like claiming (as Harman does) that (1) has some mysterious unknown truth conditions unlike the obvious ones. The gap theorist can cheerfully admit that (1) is indeed true iff it's false, while pointing out that we don't have the slightest reason to claim that it's either true or false independently of our (according to the gap theorist, improperly motivated) general assumption that every sentence is one or the other.⁹¹ They aren't doing anything strange and otherwise unmotivated with the notions of "validity" and truth preservation, like Hofweber is, but can indeed graciously concede to the dialetheist that the argument from (1) being either true or false (because every sentence is) and (1) being true iff it is false, to the conclusion that it is both true and false is quite valid, and as such truth-preserving. They would simply claim that since the assumption that every sentence is either true or false (Bivalence) is a false philosophical dogma, there's no truth there to be preserved in the inference. Moreover, and best of all,

⁹¹ An important technical caveat is in order here. Whether they can admit this actually depends of which gappy logic they adopt. If they're using Kleene biconditionals they can't, but if they're using Lukasiewicz biconditionals, they can't. The general point is that, at least on a "blurry water color" sort of picture of the conceptual basics of gap theory, there's no fundamental reason that they couldn't say this. As always, of course, the devil is in the details.

it looks like they can motivate the claim that Bivalence is a false philosophical dogma, and even specifically that we should reject the Bivalence instance “(1) is either true or false,” independently of contradiction-avoidance. As such the gappy solution *doesn't* beg the question when taken as a defense against a Liar-based argument for gluts, and that alone would make it, by leaps and bounds, the most plausible defense against that argument we have considered so far.

To see how to motivate the gappy solution independently of contradiction-avoidance, think about sentence (13), below.

(13) The sentence marked as (13) is true.

Is *this* sentence (which is a version of the Truth-Teller) true or false? And, if we doggedly stick to our guns and say that it must be one or the other because every sentence is one or the other, how would or could we ever possibly decide which one it is? In fact, if no one had ever formulated a sentence like (1), but sentences like (13) were the source of philosophical debate, it seems extremely likely that some logicians would have still formulated the truth-value gap theory as a response to (13) alone. It's obvious truth conditions—(13) is true iff it is true—shed no light whatsoever on whether it is true or false. Remember that, when it came to (1), the truth conditions and the assumption of Bivalence sufficed to get us the conclusion that it had both truth-values. With (13), they don't even seem to get us to either one.

Of course, there are many sentences that we all ordinarily assume to be either true or false—that, indeed, even the gap theorists among us would accept are either true or false—but for which we have no possible method of *deciding* whether they are true or false. Careful consideration of this point, however, actually strengthens the gap theorist's

case. Imagine a situation where we would all agree that a sentence is either true or false, but that it's impossible to determine which one it is. For example, imagine two people, Jack and Jill, are walking down a path along a beach somewhere in Miami. Jack points to a discarded Cuban cigar butt, and then to a "no smoking" sign, and asserts that there are exactly 6,542 grains of sand between the two. For the sake of simplicity, put aside questions of vagueness and assume that there are no grains of sand straddling the edge of either landmark, but that they are all clearly located well beyond the boundary of the first landmark and before the second. Further assume that the number was reasonably plausible given the apparent distance between the two.

Jill clearly has no way of deciding whether Jack's statement is true or false. Even if she set up advanced machinery to sift and count the grains of sand, the process of setting it up would dislodge a few grains here and there, and thus make it impossible for her to get an exact count. None of this should shake her certainty that Jack's statement is either true or false, because it seems clear that there is some number of grains of sand between those two landmarks. If there are either more or less than 6,542 grains of sand, the number of grains of sand makes the statement false, and if there are exactly that many, it makes the statement true.

(13) isn't like this. The reason we don't seem to have any grounds for attributing either truth or falsity to (13) seems fundamentally *unlike* Jill's difficulty in deciding whether Jack's statement is true or false. In Jill's case, there's a fact, of a quite familiar kind, such that she has good reason to be sure that this fact either makes Jack's statement true or false, and that she knows both *that* and *why* this fact lies outside of her awareness. We don't even know what sort of fact would do the job of making (13) true or false.

In other words, what careful consideration of the claim that some sentences clearly conform to the Principle of Bivalence even though we can't know their truth-value amounts to is that we need to separate out the *metaphysical* claim that for every sentence, there is either a fact that makes it true or a fact that makes it false from the (obviously false) *epistemic* claim that for every sentence, we know how to determine whether it is true or false. Indeed, it looks like the very same intuition that makes attempts to say that the Liar doesn't have its apparent truth conditions so implausible is the one that leads us to think that every truth-bearer has a truth-maker. Both claims seem to be neatly packaged in Quine's often quoted observation that when we say that snow is white, we aren't conveying any information "above and beyond" attributing whiteness to snow. The actual whiteness of actual snow makes true "Snow is white" and makes false "Snow is green."

It is the metaphysical claim that really gets us into trouble with (13), and by extension, with (1). It's not just that we don't know whether (13) is in fact true or false, but that we don't have the slightest notion what sort of fact would ever make (13) true *or* what sort of fact would make it false. And, once we reject the assumption that every sentence is either true or false, the fact that (1) it is true iff it is false gives us no reason whatever to believe that it is true, and no reason whatever to believe that it is false, and, just like (13), it starts to look like *nothing* would or could ever make it true and nothing would or could ever make it false.

Moreover, various other motivations have been offered in favor of the claim that some sentences are neither true nor false, varying from problems arising from reference failures to esoteric metaphysical concerns like Aristotle's worry that if we allow the the

Principle of Bivalence to apply to future contingents, we will be committed to embracing a sort of fatalism incompatible with belief in free will and moral responsibility. The plausibility of any individual claim of this sort is of course tied to all sorts of other philosophical and even scientific premises—e.g. if the Special Theory of Relativity entails a four-dimensionalist ontology, as many philosophers and physicists think it does, then “the future” exists in precisely the same way as “the present,” and there are facts about what happens in “the future” that fix the truth-value of “future contingent” as either true or false—but the existence of a wide variety of motivations having nothing to do with sentences like (1) or even (13) should be more than sufficient to demonstrate that we are not dealing with an ad hoc theory that only exists as a desperate attempt to save consistency.

These other arguments for gaps may succeed and they may fail, but they each surely have at least some weight until they are each shown for their own reasons to be spurious. Moreover, given the truth-maker intuition, the fact that we seem to have no earthly idea what sort of facts would make (13) true or false, and the fact that the *only* thing that allows the dialetheist to assign one truth value to (1), much less both, is their assumption of what the gap theorist would regard as our unmotivated philosophical dogma of Bivalence (which, in combination with the fact that (1) is true iff it is false, gets us the result that it is true and false), we have excellent, non-question-begging reason to suppose that not only are there gappy sentences, but that Liar sentences should be counted among their number.

Even Graham Priest admits that the Truth-Teller (unlike the Liar) *looks* like an excellent candidate for gap status.⁹² He goes on to claim that it is in fact glutty, like the Liar, but this takes some sophisticated argumentation. At any rate, the underlying intuition, which seems extremely reasonable, is that since neither (1) nor (13) have the slightest content above and beyond ascribing truth-values to themselves, they should either *both* come out as being gappy or neither should. Instead of re-interpreting the apparent alethic status of (13) to bring into line with (1), however, why not re-interpret the apparent alethic status of (1) to bring it into line with (13)? At the very least, it starts to look like we're in a dialectical impasse between the two approaches.

Unfortunately, I think that (on the assumption that either both are gappy or both are glutty) this impasse breaks in the direction of the dialetheist, for at least two reasons. To get to the first one, start by thinking carefully about the gap theorist's claim that there can be certain statements such that no fact makes it true, *and* no fact makes it false. One may doubt the background assumption that truth requires facts for a variety of reasons (e.g. one could be non-descriptivist about moral or about modal truth), but for a moment, let's assume that it's completely accurate. On this picture, if we don't want to admit gaps, what *do* we say about all sorts of cases where it may seem intuitively doubtful that there is a fact that makes some claim true or a fact that makes it false?

As it turns out, we can fully embrace the truth-maker intuition while still having a good, principled reason to deny the existence of truth-value gaps. As Graham Priest puts it,

The correspondence theory of truth may not be correct, but it captures an important insight concerning truth: for something to be true, there must be something in the world that makes it so. This need not be a state of affairs as

⁹² Priest (2006a), pp. 13

traditionally conceived of by correspondence theorists.... [For example, in] the case of a legal right, it might be certain activities of a legislature. But there must be something, some Fact, that if (counterfactually) it did not hold, the sentence would not be true....

In each case [appealed to by the gap theorist], one might take issue with the *particular* grounds offered for this conclusion. This would involve us in questions such as Meinongianism, realism in quantum mechanics and mathematics. These are deep questions...[but]...whatever the particular case, there is a *general* reason why this argument fails. In a nutshell, if there is no fact that makes α true, there is a fact that makes $\neg\alpha$ true, viz. the fact that there is no fact that makes α true. Less cryptically, the point is this. Suppose that α is a sentence, and suppose that there is nothing in the world in virtue of which α is true—no fact, no proof, no experimental test. Then this is the Fact in virtue of which $\neg\alpha$ is true.⁹³

Re-phrasing Priest's point (or extending it if we aren't assuming that falsehood is the truth of negation⁹⁴), we can say that if there is no fact that makes a sentence true, then the fact that there is no fact that makes it true *is* the fact that makes it false. This seems absolutely right to me, although to avoid being interpreted in bizarre ways, we should introduce a few qualifications (none of which will save gap theory).

The first is that, *whatever* one's theory of truth-bearers—whether one sees them as sentence-types or sentence-tokens or the propositions expressed by sentences or whatever—it's not quite accurate (perhaps not even meaningful) to say that every sentence is either true or false. Taken literally, that's not what anyone, ever, has *meant* by the Principle of Bivalence, and it isn't the claim that gap theorists mean to deny when they deny that principle. To see why, consider sentences (14), (15), (16) and (17).

(14) Is Graham Priest right about truth-value gaps?

(15) Explain why Graham Priest is wrong about truth-value gaps!

⁹³ Priest (2006a), pp.65-66

⁹⁴ On some versions of gap theory, this equivalence breaks down. For example, the gaps countenanced in Beall (2005) are such that the statements themselves are neither true nor false but their negations are (just) false. (This trick is pulled off by the introduction of two distinct types of negation operators.) On the other hand, the "classical gap theories" considered in Chapter 7 of Field (2008) are such that the equivalence between falsehood and negation is intact, but the equivalence between $\text{Tr}(\neg\alpha)$ and $\neg\text{Tr}(\alpha)$ breaks down.

(16) Graham Priest green truth-value explaining gap.

(17) Green ideas sleep furiously.

All four are sentences, and surely we would be making a very basic kind of mistake if we said that of any of them that they were “true,” but equally obviously, we would be making a mistake of precisely the same kind if we said of any of them that they were “false.” (14) is a question and (15) is a command. Surely, no one who wasn’t deeply, deeply confused would ever be tempted to put together a truth table with questions or commands symbolized with Greek letters and T’s or F’s (or even G’s) written under them.⁹⁵

Maybe we could say that all *declarative* sentences are either true or false, but that won’t quite get it done either. What about (16)? A certain sort of logician might try to make heavy water of the fact that (16) isn’t syntactically “well-formed,” but I think this misses the point. Any talk of being “well-formed” or not rests on an exaggerated and idealized analogy between natural language and logical “languages.” Letters and symbols can only combine in certain ways to form logical formulas, and a combination of a disallowed type (say, $\alpha\neg\beta$) is simply not a formula and can’t be used to express anything in the logical “language.” A combination of symbols that isn’t a well-formed formula (wff) of a system isn’t an anything of the system.

Natural language isn’t even a little bit like that. If you need to convince yourself of that, grab a tape recorder, go out and record a few thousand casual conversations—

⁹⁵ It could be objected that in imperative logics, one *does* symbolize commands with Greek letters, and no obvious category mistake is being committed by the imperative logician. Note, however, that, at least on an orthodox story about what logic is, imperative ‘logics’ are doing something merely analogous to regular alethic logic. In fact, standard discussions of and justifications for imperative logics start from the *assumption* that only declaratives can be said to be true or false. Thus, a standard line on imperative logics is that, since ‘validity’ in them can’t be about *truth-preservation*, it must be about something like ‘wish-preservation.’

among janitors, college students, soldiers, doctors, children and even philosophers and logicians—and count the number of sentences they utter that embody one sort of grammatical mistake or another, even a mistake according whatever informal rules seem to generally describe whatever eccentric little dialect is common in their neighborhood or social circle. Then ask yourself how many of the incorrectly formed, utterly ungrammatical sentences can be plausibly said to convey no information at all.

By contrast, consider (17). It is a perfectly “well-formed” subject-verb-object sentence composed solely of recognized English words and conforming to all the rules of standard English grammar. It’s just that that particular combination of words doesn’t come together to convey any information.⁹⁶ It’s a meaningless, nonsensical jumble of words, well-formedness notwithstanding. Anyone who symbolized (17)—or (16), of course—with a Greek letter, and wrote T, F or even G under it in a truth table would be showing themselves to be very confused about the nature of truth talk.

Thus, when it comes to natural language sentences, being “well-formed” is neither necessary nor sufficient for being meaningful. My suggestion is that, if we take sentences to be truth-bearers, we say that the Principle of Bivalence tells us that all *meaningful declarative sentences* are true or false, if we take propositions to be the truth-bearers, we say that the only kind of sentences that can express propositions are

⁹⁶ One could ask about poetic discourse. It conveys information, doesn’t it, despite often sounding a bit nonsensical? My answer would be that, yes, sometimes it does (even when—as they so often are—bits of poetry aren’t anything even close to ‘syntactically well-formed’!), but sometimes it doesn’t and it isn’t intended to, since all it’s supposed to do is be a beautiful-sounding and evocative combination of individual word-sounds and images. When it does, it does so *metaphorically*. The information can be expressed in non-poetic terms. Of course, sometimes, poetic discourse is meant to convey a haunting sense that *something* has been expressed even if one isn’t sure what. In those cases, I’d say that whether it does convey information depends on whether this sense is veridical—whether the unsureness reflects an epistemic problem or the fact that there’s nothing there to be sure about. In any case, needless to say, the *value* of poetry or poetry-like discourse needn’t be dependent on its literally expressing any claims about anything.

meaningful declarative sentences (and Bivalence tells us then that all of those propositions are either true or false) and so on. Notice that this doesn't give the slightest ground to the gap theorist, since no revision of classical logic is necessary to accommodate questions, commands or meaningless jumbles of words, and we are not claiming that sentences of any these types have some truth-value other than true or false, but merely that they aren't the sorts of things to which a truth-value of any kind can be meaningfully attributed. "It's true that 'Is Graham Priest right about truth-value gaps?'" is a meaningless pseudo-statement if anything is.

Most importantly, we're certainly not claiming that they are "neither true nor false" in the sense that we are asserting the negation of the claim that they are true and the negation of the claim that they are false. Notice that, the intersubstitutivity of "P" and "'P' is true" and the Principle of Bivalence dictating that "true" and "false" are jointly exhaustive of the options for every P, it would follow from questions, for example, being neither true nor false that questions were both true and false. Dialetheism would be trivially, obviously true, and we wouldn't need to appeal to anything as esoteric as the Liar Paradox to demonstrate that. It seems significant that no dialetheist has ever argued for truth-value gluts in this way, and likely that the *reason* that no one has made this argument is that the intuition that truth-value talk can only meaningfully apply to sentences (or the propositions expressed by sentences, or whatever) whose words combine not into queries or orders or meaningless jumbles, but into claims about what is the case, runs even deeper than the intuition that no sentence is both true and false. The sense in which it seems meaningful and true to say that question are neither true nor false isn't a sense where we're asserting for every question the negation of the claim that it's

true and the negation of the claim that it's false. It is, rather, a sense in which we think someone's making a more fundamental mistake when they try to apply these categories to questions.

Now, even with this reservation in place, a gap theorist might be within their rights to be unmoved by the argument that if there is no fact that makes a (meaningful declarative) sentence true, the fact that there is no such fact is the fact that makes it false. After all, they could claim, this way of thinking about how sentences are made false might have many attractive features to recommend it, in terms of explanatory simplicity, elegance and so on *if we've already decided that there aren't any cases in which anything simultaneously fails to be true and false*, but if there are such cases, Priest's account of falsehood is simple, elegant and wrong. Moreover, whether there are any such cases is precisely the bone of contention when arguing with gap theorists, so one shouldn't beg the question by assuming it.

In response to this, we can point to the fact that, while for some sentences there seems to be some positive fact above and beyond the failure of anything that makes them true, there are plenty of sentences that all of us in the normal course of things take to be false, regardless of our views about Bivalence, and which seem to be made false by nothing more than the failure of the world to include something that makes them true. Think about the difference between sentences (18) and (19), below.

(18) John McCain won the 2008 Presidential election.

(19) Vampire are real

Now, in the case of (18), we might be tempted to say that the fact that Barack Obama won the election (combined with the fact that only one person can win) is what

makes it false. The gap theorist doesn't have to admit that it is merely the fact that the world fails to include a McCain victory in November 2008 that makes (18) false. Of course, plenty of other facts could have also made it false—e.g. Hillary Clinton winning the primary and then the election, or independent candidate Ralph Nader winning the election, or a global thermonuclear apocalypse—but this is quite compatible with the need for something above and beyond the fact that there is no fact that makes it true for it to be false.

In the case of (19), however, no such extra fact seems to be needed. For (19) to be made true, the world would have to include at least one creature that looked human, but did not need to eat, was incapable of walking around in the sunlight, shunned crosses and holy water, and lived forever by drinking the blood of the innocent.⁹⁷ Now, we all think it's false, but there doesn't seem to be any special way that the way the world actually is makes it false, however, except the mere fact that it doesn't include the fact it would have needed to make it true. Thus, at least in this case, the fact that there is no fact that makes it true seems to be all that makes it false.

The gap theorist could make at least two moves here. The first might be to concede the point that *some* sentences can be made false by the mere fact that there is no fact that makes them true, but to deny that this is *in every case* sufficient for falsehood.

⁹⁷ Of course, depending on how much one believes is packed into the notion of a "vampire," these requirements may be taken to be either too narrow or too broad. Moreover, if one subscribes to Kripke's theory of reference, and takes "vampires" to be the proper name of a natural kind of fictional characters or some such (which one doesn't *have* to—one could accept Kripke's theory but still take "vampires" to reduce to a description—but one certainly might), one might take even an exhaustive list of this kind to be insufficient. Even this objection, however, can be circumvented if we expanded our list to add some condition like that the creature(s) the world would have to include would have had to be the cause of original slavic peasant superstitions about vampires, such that our word "vampire" originally baptized them. Better yet, we can simply change our example from "vampires are real" to something of the format "[monsters of a type invented for the purposes of this example] are real." Nothing about the example hinges on the selection of vampires as opposed to any other type-of-things of which there aren't any.

For some sentences, or some categories of sentences, there could still be neither a fact that makes them true, or a fact that makes them false (and, in these cases, the fact that there is no fact that makes them true would not count as such a false-making fact). Now, the onus would be on the gap theorist to tell a principled story about when failure-of-truth-making adds up to false-making and when they don't, much less how we can tell the difference, and it's hard to say in advance what this story could amount to.

Alternately, the gap theorist could claim that there *is* a further fact that makes (19) false, and that this is a fact about what sorts of things *do* exist. For each existing object, there is a fact about what sort of thing it is and a fact that being that sort of thing is incompatible with being a vampire. All of these facts add up to the fact that everything that exists is a non-vampire, and this is the fact, above and beyond the absence of fact that makes (19) true, that makes (19) false.

Now, it is not entirely clear whether there is a substantive distinction between every fact of the relevant kind failing to be the sort of fact that would make (19) true and the mere absence of fact that would make it true. At the very least, it seems plausible that the reason that the sense in which we take the former fact to make (19) false, and even that the sense in which we take the fact Obama's victory to make (18) false, is simply that we take the relationship between these facts and the sorts of hypothetical facts that would have made either true to be such that we know that the presence of the former guarantee the absence of the latter.

Moreover, even if that's wrong, and we should take these facts to be false-making for some other reason (or there is indeed some principled reason why we should take the fact that there is no fact that makes a sentence true to in some but not all cases constitute

a fact that make it false), and accordingly there are at least some truth-value gaps, meaningful declarative sentences that are neither true nor false, our second argument would still decisively show that there is no workable gap-based defense against the argument from the Liar Paradox to dialetheism.

Take sentence (20) below, which is a version of the Strengthened Liar.

(20) The sentence marked as (20) is not true.

If (20) is true, it's not true, so it's both true and not true. Contradiction. If (20) is false, then it's not true, and so it is true, and so it's both true and not true. Contradiction again. If (20) is gappy, "neither true nor false," then it's not true, so it is true, so it's both true and not true. Contradiction once more. No matter whether one takes there to be three truth-values or nine or a thousand, (20) having some truth-value other than "true" or "false," and incompatible with either of those, gets us the result that it isn't true, which is precisely what it claims of itself.

If it's some truth-value that's *compatible* with true, then the question arises of whether it's *also* true. If the answer is "yes," then the ascription of that truth-value to (20) doesn't help, since, as we've already seen, the assumption that (20) is true generates the conclusion that it both is and isn't true. If the answer is "no," the ascription of that truth-value to (20) *still* doesn't help, since we've already seen that the assumption that it isn't true generates the conclusion that it both is and isn't true. One way or the other, there's no way to use the idea that "true" and "false" aren't jointly exhaustive of the possible truth-values that a meaningful declarative sentence can have to attribute a *consistent* truth-value to (20).

If there really are truth-value gaps, we can't allow that (20) is one of them without accepting in the bargain that there are also truth-value gluts, and that (20) is both, since if it has some truth-value other than true, it is also true. The gappy solution doesn't get us off the dialetheist's hook.

That said, it doesn't seem to be a total non-starter like the question-begging approaches discussed in the previous chapter. There seems to be something right about the gappy analysis of sentences like (1), (13) and (20), and if we can sift out the core insight *underlying* gappy approaches from the proposal that there really are meaningful declarative sentences that simultaneously fail to be true or false, and that paradoxical sentences are of this type, we might be well on our way to a workable solution.

Chapter Six: Liars and Meaninglessness

Graham Priest has claimed that all of the “paradoxes of self-reference” are instances of the same pattern, which he calls the Inclosure Schema.⁹⁸ The components of the schema are that the locus of the paradox (whether that is a sentence or a set) *exists*, that it *has* some property (such as “being true” or “being self-membered”) and that it *fails* to have that property. He has derisively referred to any sort of solution to the Liar Paradox that relies on denying the *existence* component by denying that Liar-like sentences count as meaningful (or that, for example, that they express propositions) as “the heroic solution.” The idea is presumably that it’s simply so obvious that they are meaningful, truth-value-bearing sentences that it takes (misguided) heroism to will yourself into denying it. Like a hero throwing himself on a grenade to save his friend at the cost of his life, the proponent of the “heroic solution” is sacrificing the plausibility of their beliefs to save themselves from having to admit the justice of the otherwise inexorable inference from the difficulties involved in attributing any consistent truth-value to Liar-like sentences to the conclusion that such sentences are dialetheias. The whole project of denying that Liar-like sentences are the sort of thing to which we can meaningfully apply truth-talk ends up being a philosophical version of the “stopcrime” techniques in George Orwell’s *1984*, whereby citizens learn to stop themselves from realizing treasonous things that would constitute “thoughtcrime.”

I will be arguing, below, that this is exactly wrong. Far from being really heroic, the “heroic solution” can and should be adopted on the basis of extremely cautious, reasonable and well-motivated considerations that are quite independent of

⁹⁸ See, for example, Priest 1995, Priest 1998b, or Priest 2000.

inconsistency-avoidance. Before going into all of that, however, it's worth taking a moment to offer an analysis of the continuing appeal of gap theory.

After all, despite its many persistent problems, gap theory persists and continually branches out into new and more sophisticated forms and varieties, from Hartry Field's "paracomplete" theory to Laurence Goldstein's neo-Cassassionist approach.⁹⁹ Both display interesting ambiguities, which at least in Field's case clearly indicate an understanding of the severe difficulties facing any attempt to assert gaps without admitting gluts. His view is that while Liar-like sentences have truth-values between the classical values 0 and 1, this shouldn't be read as the claim that they are gappy or neither true nor false, and that indeed, while sentences with such truth-values should be rejected, so should, e.g. the statement that these sentences are neither true nor false, so as to avoid some of the objections discussed in the last chapter. Nonetheless, given that his view is that Liar-like sentences are *meaningful* and *truth-evaluable*, the sort of things that are properly symbolized with Greek letters and evaluated in truth tables, and that doesn't regard them as either true or false, it's hard not to regard his "paracomplete" proposal as yet another reformulation of gap theory.¹⁰⁰ In Golstein's case, within the span of the same

⁹⁹ See Goldstein (2006) and Field (2008).

¹⁰⁰ Field identifies 0 and 1 as "determinately true" and "determinately false," respectively, but rejects that .5 or any other value in between should be read as "neither true nor false." His stance is that there is simply no fact of the matter about whether there are gaps (or gluts). Thus, for any paradoxical sentence, he *both* rejects the claim that the sentence is either true or false, *and* rejects the claim that they are neither true nor false, and continues to do so "all the way up." It seems to me that this view faces a Strengthened Liar-type problem of its own, when it comes to the sentence "Anyone who ever asserted of this sentence that it had truth-value 1 would be making a mistake of some sort." (To be clear, the question is not whether Field's account has the formal resources for avoiding the derivation of a contradiction here—his resources for that task are formidable, and he has an infinite hierarchy of "determinately" operators, with no overall one such that we can talk about "determinately true for any sense whatsoever of 'determinately,'" precisely for the sake of navigating around these sorts of revenge paradoxes. The question is, rather, whether Field's account leaves him any room to say anything intuitively plausible about the sentence just mentioned. After all, presumably the *whole point* of the account is that it would be mistaken for anyone to ever assign truth-value 1 to any paradoxical sentence, so intuitively, if his account is right, anyone who ever claimed that the sentence just mentioned had truth-value 1 *would* be making a mistake.)

paper, he sometimes says that Liar-like sentences have the value “GAP” and other times says they “refuse to yield statements” and don’t “say anything,” just as, in his interesting and suggestive analogy, if the Fibonacci sequence didn’t start with “1 1,” but rather with “the sum of the previous two numbers,” this phrase would fail to designate any actual number, so his overall view is somewhat ambiguous between the view that sentences that don’t ground out in being about anything other than truth or falsity are meaningless pseudo-statements and the view that they are meaningful but have a truth-value other than the two classical values. Even Goldstein, who has one foot in “heroism,” keeps another firmly planted in gap theory.

A plausible reason for this strange persistence of gap theory in light of all of its troubles is that it isn’t *just* an ad hoc attempt to save consistency. There is something intuitively appealing about the claim that *some* sort of mistake is being made when we call sentences like the Liar or the Truth-Teller true or false, and the best, most sophisticated formulations of gap theory have given this intuition an extremely plausible motivation. It’s worth taking a look at this, in the hopes of separating out the genuine insight lurking in the best versions of gap theory from the claim that there are actually truth-value gaps. Once this has been done, we’ll be most of the way to a good, well-motivated solution.

The most influential—and I think, the most plausible and insightful—version of the gappy proposal on record is the one Saul Kripke advances in his classic 1975 paper *Outline of a theory of truth*. In fact, despite our dismissal of Tarski, above, as either avoiding the problem, advancing a proposal that was justified iff it was unnecessary, being a dialetheist or begging the question, depending on which of the four

interpretations of what he's up to we focus on, I think that a careful analysis of what's right about Kripke will also reveal what's right about Tarski.

Side-stepping the details of the formal construction Kripke proposes in terms of three-valued logic—which would only be relevant if we hadn't already rejected the notion that any sort of solution based on assigning paradoxical sentences a truth-value other than “true” or “false” can get us off the dialetheist's hook—the important part is that he thinks that the way we can sort out the gappy sentences from the non-gappy ones is to look at whether they are “grounded.” To get a sense of how this works, consider the following series of sentences:

(21) Sentence (22) is true.

(22) Sentence (23) is false.

(23) Sentence (24) is false.

(24) Sentence (25) is true.

(25) Sentence (26) is true.

(26) Hitler lost World War II.

Now, contrast sentences (21)-(25) to the the Liar sentences we've considered—(1) and (20)—or to the Truth-Teller sentence (13) or to the infinite series of Yablo sentences discussed above, or a Yablo-like non-paradoxical infinite series of sentences where every sentence says of the sentence above it that it is true, or the multi-sentence paradox consisting of sentences (11) and (12), above, or the pair formed by (27) and (28), below, where (27) is true iff (28) is false and false iff it is true, but where no paradox can be derived from this.

(27) Sentence (28) is false.

(28) Sentence (27) is false.

What every single one of the sentences just mentioned—the first five sentences in the series (21)-(26), the pair formed by (11) and (12), the pair formed by (27) and (28), the Yablo series, the infinite series of sentences attributing truth to the sentences above them in the series, and the Liar sentences (1)—have in common is none of them does anything but attribute a truth-value to a sentence. However, the sentences (21)-(25) all seem fundamentally different from any of the other sentences I just mentioned. What makes them different is that they all refer (albeit *indirectly* in the case of (21)-(24)) to a sentence that's *not about truth*. In Kripke's terminology, they are all "grounded," whereas all of the other sentences just mentioned are "ungrounded."

A grounded sentence is either directly or indirectly about something other than truth.¹⁰¹ No sentence ever "becomes" grounded by attributing a truth-value to an ungrounded sentence, but a sentence that attributes truth or falsity to a grounded sentence is itself grounded, and either true or false depending on the truth or falsity of the sentence the series grounds out in. For example, (26) is true, which makes (25) true, and that

¹⁰¹ I should add an extremely important technical note here. Since I'm talking about "grounding," and will continue to do so when I come to my own views, some readers may wonder if I mean "grounded on Strong Kleene semantics," "grounded on Weak Kleene semantics" or what. The answer is that I mean none of these things, but something far more basic. When discussing Kripke, since the details and pitfalls of his formal construction aren't of interest for my present purposes, I'm only interested in the intuitive conceptual core of "grounding," which is simply that a grounded truth-attribution directly or indirectly attributes truth to a sentence which *does not use the truth predicate* (or the truth operator, which on my view, behaves in precisely the same way as the truth predicate when it comes to questions of 'grounding'). Moreover, this intuitive core is all I plan to take from Kripke. If this is not precise enough, I offer the following definition: a sentence that uses the truth predicate (or operator) is 'grounded' iff it can, in principle, be accurately paraphrased with one or more sentences that do *not* use the truth predicate (or operator)." Obviously, the "in principle" part is important, given that the most important linguistic functions of truth talk—blind endorsements, complicated generalizations and so on—are ones where the paraphrasing-away is impossible *in practice*. The notion that truth attributions can be accurately paraphrased (i.e. they lose none of their content) into 'true'-free sentences, of course, assumes the "nothing above and beyond" principle, which those with more inflationary views about truth couldn't accept. I take the simple view of grounding "nothing above and beyond" allows to be, in itself, a pretty good advertisement for a more deflationary stance on truth.

makes (24) true, which makes (23) false, and that makes (22) true, and that in turn makes (21) true. If (26) had been “Hitler won World War II,” (21) would be false. If (26) were a claim about who won some minor tribal war that took place several thousand years ago, and we had no records to indicate which side won, we could still be sure that (21) was either true or false, although we would not know which.

Kripke’s proposal is that we take ungrounded sentences to be neither true nor false, getting ?’s instead of T’s or F’s when we put together our truth-tables. This gets us into trouble for reasons that we have already seen, most importantly the Strengthened Liar. However, there seems to be a basic insight here about how truth-attributions work that can be separated out from the unworkable proposal that some meaningful declarative sentences are neither true nor false. The positive point here is that the problem with sentences like (1) and (20) is not that calling them true or false commits us to saying that they are both. The problem is that they are truth-value-attributions that don’t ground out in sentences about something other than truth. The thought that *this* should make such sentences defective in some sense is extremely intuitively compelling, it can be motivated by philosophically plausible considerations (see below) *and* it begs no question against the dialetheist.

Moreover, a similar insight seems to lurk in Tarski’s requirement that sentences in languages with truth and falsity predicates only apply them (directly or indirectly) to sentences of the object language, which lacks a truth predicate. Putting all of this together, we can call this the “Kripke/Tarski Thesis.”

Kripke/Tarski Thesis: We are making some sort of mistake when we attribute truth or falsity to a sentence that isn't (directly or indirectly) about something other than truth.

To get a better sense of *what sort of mistake* we are making, recall Quine's claim, mentioned above, that when we say that the sentence "snow is white" is true, we aren't saying anything that goes "above and beyond" attributing whiteness to snow. Let's generalize this into the "Quine Thesis."

Quine Thesis: The only information conveyed by an attribution of truth to a sentence is the information conveyed by the sentence that truth is being attributed to.

While we're at it, recall Priest's conceptual point about truth, negation and truth-makers. (The only "fact" we need to make a statement false is the fact that there is no fact that makes it true. Given that, like Priest and even *some* gap theorists, I understand falsity to be nothing more) We can codify this, in the same language we used for the "Quine Thesis," as the "Priest Thesis."

Priest Thesis: The only information conveyed by an attribution of falsity to a sentence is the information that would have been contained by an attribution of truth to the negation of that sentence.

For the sake of convenience in referring back to it later, we can combine these claims into a single "Quine/Priest Thesis," though this shouldn't be taken as suggesting that Priest would agree with it.

Quine/Priest Thesis: The only information conveyed by an attribution of truth to a sentence is the information conveyed by the sentence that truth is being attributed to, and

the only information conveyed by an attribution of falsity to a sentence is the information that would be conveyed by an attribution of truth to the negation of that sentence.

The Quine/Priest Thesis, and the Kripke/Tarski Thesis, jointly take us almost all the way to a workable, plausible and non-question-begging defense against the argument from the semantic paradoxes to dialetheism. The final touch is to precisify the core insight somewhat vaguely expressed as the Kripke/Tarski Thesis above by saying that *truth-attributions are only meaningful parasitically*. Just as attributing truth to a sentence adds no content to the sentence, a sentence that says of itself nothing beyond attributing truth to itself *has no content*. (As Carnap was so fond of pointing out, the negation of nonsense is nonsense.¹⁰² A sentence must have meaning for its negation to have meaning.) The former view only makes sense if we take the meaning of a sentence attributing truth to be *entirely* inherited from the sentence it attributes truth to, and when a sentence does nothing but attribute truth to itself (or to the negation of itself), there's nothing for it to inherit its meaning from, and it is simply devoid of any possible meaning. As such, it's simply not the kind of thing we can meaningfully apply truth talk to, and claims that it is true, that it is false, that its negation is true, that it's negation is either true or false, that its negation is neither true nor false and so on are all equally meaningless.

Note, importantly, that this parasitical-meaningfulness principle applies equally well when truth is being attributed to a statement through the use of a truth *predicate* and when it's been attributed to a statement through the use of a truth *operator*. On some views of truth and meaning, "it's true that green ideas sleep furiously" is meaningless, but "'green ideas sleep furiously' is true" is simply false. I strongly reject this claim of

¹⁰² See, for example, Carnap (2001), pp. 425-435

inequivalent predicate/operator behavior of truth-attributions, and think that this rejection is a natural consequence of my picture. To begin with, remember that Quine's original formulation of the "nothing above and beyond" principle is not *about* truth as a predicate (used within a sentence) but truth as an operator, applied to a whole sentence.

To ascribe truth to the sentence ["Snow is white"] is to ascribe whiteness to snow...Ascription of truth just cancels out the quotation marks. Truth is disquotation.

Moreover, *if* one assumes that truth plays Capture and Release—that is to say, that $\text{Tr}\langle\alpha\rangle$ and α can each be inferred from the other, something anyone who subscribes to "nothing above and beyond" certainly *should* assume—the claim that "green ideas sleep furiously" is meaningless but that "'green ideas sleep furiously' is true" is false generates an absurd consequence, namely that a meaningful (albeit false) sentence *entails* a meaningless one. If this is not absurd enough, observe that if "'green ideas sleep furiously' is true" is *false*, then "'green ideas sleep furiously' is not true" is presumably *true*. This means that, on the view that attributions of truth to meaningless statements are meaningful if truth is being attributed through an operator rather than a predicate, and that truth plays Capture and Release, a statement that's not just meaningful but also *true* validly entails a conclusion that's *meaningless*. Something, somewhere, has gone wrong in any chain of reasoning that generates that result. My candidate for *what* has gone wrong is the assumption that it matters whether truth is being attributed to a sentence in operator or predicate form.

Finally, and most importantly, the claim that it should matter for this issue whether truth is being attributed to ungrounded truth talk (i.e. truth talk that neither directly or indirectly refers to a "true"-free meaningful declarative sentence) through an

operator or a predicate, if taken as an objection to the solution, subtly begs the question against the proposed solution by implicitly *assuming* that sentences that attribute truth to themselves or other ungrounded sentences are *meaningful*. After all, the function of *any* logical operator is to turn a complete sentence into a sentence, where “sentence” doesn’t mean “question” or “command” or “meaningless jumble of words,” but the sort of sentence that can legitimately be the interpretation of a Greek letter in a logical context. In other words, to say that, even if α is meaningless $\text{Tr}\langle\alpha\rangle$ should just be false, is to *assume* that the sentence in question is the kind of thing which we can symbolize with a Greek letter and perform logical operations on without committing a nonsensical category mistake. It is, in other words, to assume that it is *meaningful*!

It’s important to emphasize that the words “meaningful” and “meaningless” are being used in the completely normal sense of those words. Some people simply refuse to believe this. For example, Hartry Field talks about the relevant sentences being so obviously meaningful “in any ordinary sense of ‘meaningful’” that, when people (like me) dispute this “it must be some special technical sense of ‘meaningful’ that is intended.”¹⁰³ As such, a meaninglessness view must be “a kind of paracomplete view” being unhelpfully presented in a misleading disguise.¹⁰⁴

I would respond that, while I can’t speak for any other theorist who says similar things, in my case I don’t mean “meaningful” and “meaningless” in any strange technical sense. I mean them in precisely the ordinary mundane senses of the word, despite the fact that some people have a strong intuition that ungrounded truth talk is meaningful.¹⁰⁵ I

¹⁰³ Field (2008), p. 12

¹⁰⁴ Field (2008), p. 13

¹⁰⁵ Of course, given the strong intuitive pull of the idea that Excluded Middle and Non-Contradiction both hold universally, anyone who shares my intuitions on these questions but shares Field’s intuition about the

simply don't think that we should be Cartesians about questions of meaningfulness, taking the privileged access of competent speakers of a language to somehow render their judgments infallible when it comes to questions of which sentences of the language are meaningful. Rather, I take such judgments to be excellent (but defeasible) evidence of the same, and take it that competent speakers can be mistaken on such issues.

To take a particularly obvious example, the philosophers of the Vienna Circle were presumably all competent speakers of the German language. Despite this, they came to severely mistaken conclusion that many perfectly meaningful German sentences expressing unverifiable metaphysical views were literally meaningless, the cognitive equivalent of music. Of course, one could grant that *philosophers* are capable of making mistakes about meaning, but hold that the folk in their wisdom are incapable of such extreme errors. I address this below, sketching out what I take to be some mundane ways that ordinary people can become mistaken about ordinary questions of meaningfulness for ordinary reasons.

With this in mind, to be clear, my view is that sentences that attribute truth (or falsity) to a sentence—whether themselves, or others, since self-reference isn't the issue here—are only meaningful (in the ordinary, garden-variety sense of meaningful) if the sentence to which truth is being directly or indirectly attributed is about something other than truth. (Sentences about truth that fail this test are meaningless in precisely the same ordinary way that “green ideas sleep furiously” is meaningless.) To put the point in stark terms, if saying that something “is true” is essentially a device for expressing agreement (and saying that something “is not true,” or, equivalently, “is false” is a device for

meaningfulness of ungrounded truth talk—or Priest's, given Priest's dismissal of the meaningfulness move as 'heroic'—faces a choice about which intuition is more basic when deciding which of our three views is the least implausible.

expressing agreement with the negation of the sentence under consideration), then when sentences fail to be directly or indirectly *about* anything but the truth-value of sentences, there's nothing there to express agreement or disagreement with and no information conveyed, nothing there to meaningfully affirm, deny or infer anything from.

To tie up one obvious loose end, we should note that, despite my use of “sentences,” this solution is completely neutral on the subject of what sorts of things are truth-bearers. If you think sentence-types or sentence-tokens or sentences-in-contexts or whatever are truth-bearers, then for reasons outlined above in qualifying Priest’s argument against gap theory, the word “sentence” in any of those phrases should really be understood as shorthand for *meaningful declarative sentence* unless you want to run into problems for other types of sentences, like questions, commands and meaningless jumbles of words. Similarly, if only the propositions or claims or whatever expressed by sentences are truth-bearers, then presumably questions, commands and meaningless jumbles of words can’t express propositions or claims or whatever. Only meaningful declarative sentences can. For the sake of simplicity and convenience, I will continue to talk about “sentences,” but absolutely nothing hinges on this.

Whatever one takes truth-bearers to be, my view is that truth-value ascriptions no more convey information than do transparently nonsensical sentences like (16) and (17) above, or, for example, cries of pain. When a carpenter accidentally smashes his thumb with a hammer, his scream of “motherfucker!” does not constitute a claim to which truth-talk can be meaningfully applied—say, an accusation that someone else in the room has had carnal relations with their own mother—but merely expresses his pain. (Even if he was only *pretending* to be in pain, we wouldn’t say that he was *lying* or saying something

false. That's just not the sort of language that makes sense here.) Of course, with other intentions, the same outburst might convey precisely that information, just as if, someone points at a mathematical equation on a chalk board and says "this sentence is false," that conveys information, but when a logician writes it down with the intention of having the "this" refer to the sentence that word appears in, no information is conveyed.¹⁰⁶ As we established earlier, being "well-formed" in terms of English grammar is neither necessary nor sufficient for being a meaningful declarative sentence, the kind of thing that's properly symbolized with a Greek letter or the sort of thing that logical laws are about. Empty truth-value ascriptions no more belong in truth tables with T's, F's or even G's under them than do questions, commands, or more obviously nonsensical sentences.

Notice that the grounding-out requirement this solution imposes on meaningfulness is that sentences be directly or indirectly about something other than *truth*. We're not making the broader claim that they must ground out in "extra-semantic reality." This is not an oversight. The reason that *truth-value-ascriptions* must inherit their meaning from the meaning of the sentences to which truth or falsity is being attributed, and hence why *empty truth-value ascriptions* are necessarily meaningless, do not necessarily apply to semantic concepts other than truth. (Perhaps, in the case of *some* of other semantic concepts, a similar story might be told, but such claims must be evaluated on a case-by-case basis. Although I don't work out the thought here, one extremely plausible candidate for truth-like-treatment is "applies to itself" predicate.

¹⁰⁶ Of course, there's a loose, irrelevant sense of "conveys information" in which all sorts of things that aren't meaningful declarative sentences convey information. A scream of pain conveys the information that the screamer is in pain, a cough may convey the information that the cougher smokes too many cigarettes and an utterance of "the sentence I am speaking right now isn't true" may convey the information that the speaker is confused about how the truth predicate works. All uses of the phrase 'conveys information' from here on out, however, can be read in the stricter sense of 'actually making some sort of claim about something.'

After all, normal meaningful uses of the predicate seem to “ground out” in much the same way as normal uses of truth. For example, “the predicate ‘being a predicate’ applies to itself” seems not to mean anything beyond “the predicate ‘is a predicate’ is a predicate.” Now, given that it’s not clear that “applies to itself” means something different than “is true of,” one might object that the story just gestured at for “applies to itself” is not really a *second* example of a semantic concept that behaves like truth, but an uninteresting extension of our story about truth. I won’t argue. I do, however, keep an open mind about the possibility of other cases that *are* more independent of truth.) Given that, on the one hand, extending this analysis to *all* semantic concepts would be unmotivated, and on the other hand, as we will see below, this would needlessly get us into trouble with certain revenge paradoxes, I’m inclined to say that we’re better off narrowly and cautiously applying this analysis to *truth* only rather than brashly applying it to all semantic concepts.

Moreover, this solution works equally well for *all* Liar-like paradoxes without exception, not only the Simple Liar but also ever single Strengthened and Revenge Liar ever formulated, since none of them are directly or indirectly about any subject other than truth. It works equally well for Yablo’s Paradox, discussed above, since although none of the Yablo sentences are about their *own* truth-values, none of them are directly or indirectly about anything but truth. The Yablo Series goes on forever without ever finishing in a sentence about something other than truth from which the rest of the sentences in the series could inherit their meaning, so every sentence in the series is meaningless for precisely the same reason that self-referential truth-ascriptions like the Liar and the Truth-Teller are meaningless.

Moreover, *unlike* the dialetheist solution of simply biting the bullet of inconsistency, which (for reasons we'll come to see in Chapter Eight, when we unpack this paradox in detail) *doesn't* work for the Curry Paradox, my solution applies equally well to "Curry sentences" like (29).

(29) If (29) is true, then the earth is flat.

For reasons we will see in Chapter Eight, the mere statement of the truth conditions of (29), after it's been fed into the T-Schema, entails that the earth is flat by a quick series of straightforward and unobjectionable inferences. Similar Curry sentences can be used to establish that *everything without exception* is true. This is just as much a problem for the (non-trivialist) dialetheist as it is for the monaletheist. Priest's solution of biting the bullet, accepting that the sentence has its apparent truth conditions and reasoning paraconsistently about it to avoid having to infer everything transparently fails to apply here, and Priest accordingly has to solve it in a completely different way, egregiously violating the Principle Of Unified Solution, given its obvious similarity to the Liar Paradox. After all, just as the inference to contradiction from the Liar involves accepting that the sentence is meaningful and reasoning in apparently unproblematic ways from the only truth conditions it can have given the T-Schema, the inference from (29) to triviality works in *precisely the same way*.

The solution advocated here, however, transparently applies equally well to the Liar and Curry. If the referent of "this sentence" is supposed to be the antecedent alone, then the antecedent is simply the Truth-Teller, which is meaningless for reasons we've already seen, and any sentence with the syntactic form of a conditional that has a meaningless statement in the position of the antecedent is itself pretty clearly

meaningless. If the referent is the whole conditional--as it's pretty clear that it is intended to be, and indeed as it has to be for the inference to triviality to go through--the situation is no better. It's true enough that "the earth is flat" is meaningful, so if the phrase "this sentence is true" could inherit its meaning from *that*, it would itself be meaningful, but the sentence is not "if the consequent of this conditional is true, the earth is flat" (which would be a fairly uninteresting tautology, and not the basis for any sort of paradox) but "if *this sentence* is true, the earth is flat." *That* truth-ascription no more grounds out in a subject other than truth than does the Liar or the Truth-Teller.

So, to review, we have a solution that applies to all of the Liar-like paradoxes, which captures the intuitive force of the most plausible versions of gap theory but to which Priest's arguments against gap theory do not apply, which doesn't force us to abandon classical logic in favor of any gappy or "paracomplete" or otherwise non-classical and inferentially weaker logic, which is well-motivated by serious, widely and independently held philosophical considerations (notably the "nothing above and beyond" principle) that have nothing to do with inconsistency-avoidance, but which *does* save consistency, and which even fares better on any reasonable version of the Principle Of Unified Solution (to the semantic paradoxes, at least) than even the dialetheist's position. This might sound too good to be true, and, while I don't think it is, there are further hurdles to be overcome before we can be sure we've arrived at a fully workable solution.

One could make at least five obvious objections at this point.

1. Wait wait wait. This whole time you've been reasoning about what follows from saying that the paradoxical sentences are true or false or some third value, what truth

conditions they must have if they have any and so on. Now you turn around and say that they don't convey any information? Doesn't just about everything you've said before that claim was made falsify the claim?

2. The Liar seems meaningful. How do you explain away this intuition?

3. What about accidental and contingent paradoxes, like Kripke's Nixon example? Isn't it a bit much to say that a sentence can be meaningless only because of which other sentences happen to be true or false?

4. Earlier, you criticized Tarski for banning himself from talking about the virtues of his solution in his languages. Don't you have a similar problem here, given that a statement like "(30) is meaningless" doesn't ground out in a claim about extra-semantic reality?

5. If something's meaningless, it isn't true. Doesn't this solution still fall prey to the Strengthened Liar? And even if you can somehow get around it, what about further revenge paradoxes, like "this sentence is either false or meaningless" or "this sentence does not express a true proposition"?

Let's consider these one at a time.

1. Whether or not one believes that empty truth-value-ascriptions convey information, surely hiccups do not.¹⁰⁷ Let's say that, despite this, some deeply confused person (let's call him Tim) symbolizes their hiccup as H, and proceeds to try to reason

¹⁰⁷ Again, one could think that, when truth is being deployed as an operator, it's attribution to anything is meaningful, and when its negation is attributed to anything other than a true sentence, the resulting statement is *true*. I replied to this extensively above on the question of meaningless-but-syntactically-declarative sentences, and obviously all of those arguments apply equally well to hiccups, but here I'd just add that (a) operators function to turn sentences into sentences, and a hiccup isn't a sentence, and (b) whether we frame the question in terms of operators or predicates, the idea that there's a meaningful question about whether or not hiccups are true presupposes the inflationary assumption that 'is true' (regardless of what logical form the claim is expressed in) picks out some sort of robust property, rather than simply acting as a device to endorse claims. The notion of expressing agreement with a hiccup is nonsensical on its face.

about whether H is true, false or neither. Then a critic (let's call him Tom) wades into the fray, drawing out the various contradictions and ambiguities in the claims that Tim has made about H--"wait, earlier you said H wasn't true, but if it's not true, then it must be false, and..."--and nit-picking various moves that Tim made. A lot of this sort of thing could go on equally well whether or not Tom knew that (nonsensically) H was meant to symbolize not a statement but a hiccup. If Tom did realize the subject under discussion was a hiccup, and concluded the discussion by pointing out that H is just not the sort of thing to which truth-talk can be meaningfully applied, the earlier discussion would hardly render his attitude inconsistent.

When indulging Tom in the game of treating H as if it was a meaningful declarative sentence, it made perfect sense to, for example, respond to Tim's attempt to gerrymander truth conditions for H that would firm up his claim that H was neither true nor false by pointing out that, by application of the T-Schema, "'hiccup' is true iff hiccup" and no other truth conditions could possibly apply. Seeing the justice of this point hardly requires recognizing "'hiccup' is true iff hiccup" as a meaningful claim...which, pretty clearly, it is not.

2. This is an objection one hears incessantly in these sorts of discussions, but its force is never quite clear. In any given context, does it mean that it seems meaningful to the person making the objection, that it really, secretly seems meaningful to the person claiming that it is in fact meaningless, that it seems meaningful to absolutely everyone, or merely that it seems meaningful to most people? The first interpretation seems irrelevant, if the objector can't back up his private intuition with any sort of further argument, the

second interpretation seems paranoid and unmotivated, and the third interpretation seems fantastically unlikely on its face, so, charitably, let's go with the fourth interpretation.

The first point worth making is that it is far from clear that most people really do pre-philosophically regard the Liar as meaningful. This is a claim about social psychology that one frequently hears being made from the armchair, but to the best of my knowledge, no one has ever done any empirical research on the subject. This sounds like a job for experimental philosophy to me, and until some has been done, I will reserve judgment. Anecdotally, my impression is that most people's pre-philosophical reaction to the Liar Paradox is along the lines of "OK, something's wrong with it, but I don't know what...where's the trick?" On the face of it, this reaction seems quite compatible with at least a vague suspicion that the sentence might be meaningless. At the very least, given the track record so far of "x-phi" falsifying claims commonly made in the literature about allegedly widely shared pre-philosophical intuitions, we shouldn't place too much stock on anyone's empirically uninformed *impressions* of what most people think.¹⁰⁸

Let's assume for the sake of argument, however, that it is indeed true that most people have the intuition that the Liar is meaningful. First of all, I think it can be easily demonstrated that we are not infallible about the meaningfulness of our sentences. Mistakes about meaningfulness are possible, unremarkable and common. In light of this, a vague pre-reflective sense that a sentence is meaningful shouldn't count for much against a rigorous argument that it is meaningless. Secondly, in the specific case of the

¹⁰⁸ The clearest example of this I know of is in the free will debate, where the alleged 'common sense,' intuitive force of incompatibilism has traditionally been central to the libertarian's argument. Studies, however, consistently show that respondents asked to assume strong causal determinist claims still judge wrong-doers to be morally responsible for their actions.

Liar, I think we can tell a pretty straightforward and plausible story about how this mistaken intuition might arise.

On the first point, imagine that Noel A. and Mary B. are going out to dinner, and Noel A. is pretending to be interested in what Mary B. is talking about while he really thinks about something else entirely. He may slip in a “yes, I agree” or a “that’s true” at moments in the conversation dictated not by the content of what she’s saying (which he doesn’t register) but rather by the pacing of and pauses in the conversation. At some point, she realizes that he isn’t paying any attention to her and emits (in a normal tone of voice, and with normal pacing) a series of nonsense syllables, like “bork geblord semlord” and he says “that’s true.” Now, not knowing what Mary B. said, Noel A. presumably thinks that what he said in response to her was either meaningful and true (if what she said was true) or meaningful and false (if what she said was false), but since what she said was meaningless, so was his attribution of truth to it. (If “it’s true that ‘bork geblord semlord’” isn’t meaningless, nothing is.) If, as in this case, speakers can be fallible about the meaningfulness of their own utterances, then surely listeners can be at least as fallible about the meaningfulness of the sentences they are hearing.

Let’s tweak the carpenter example from the last chapter. Let’s call the carpenter Josh, and say that his co-worker Joe is in the other room when Josh accidentally slams his thumb with his hammer. Joe is a sexual deviant who has been engaged in regular carnal relations with his own mother, and he’s terrified that his friend Josh will find out about it. When he hears Josh’s scream of “motherfucker!” from the other room, Josh immediately assumes that he’s been found out. Once again, something that doesn’t convey information is mistaken for something that does.

On a less esoteric level, the common adolescent boy's joke of responding to a claim one finds ridiculous by saying the word "bullshit" into one's hand while sort-of-pretending to cough plays on the routine ease with which we can create confusion between things that convey information and things that do not. Situations where a listener, or even (as in the Noel A. case) a speaker can confuse the meaningless for the meaningful are easily constructed and quite unremarkable.

In the specific case of empty truth-value ascriptions like the Liar and the Truth-Teller, if (for the sake of argument) they *seem* meaningful, that is because they are grammatically well-formed (although so are some more obviously meaningless statements like "Pumpkin pie penetrates Modus Ponens"), because self-reference per se isn't the problem (e.g. "This sentence has seven words in it" is meaningful and true whereas "This sentence has nine words in it" is meaningful and false), because it's very common for sentences to be meaningful even though we have no rational, justifiable way of deciding whether they are true or false (like the grains of sand case in the previous chapter) and because in other cases (i.e. those of grounded-out series' of sentences) a sentence that says of some sentence that it is not true but otherwise contains no information is perfectly meaningful. In other words, any appearance that the Liar is meaningful can be explained by the fact that it sits at the intersection of four categories (grammatically well-formed sentences, self-referential sentences, sentences about truth, and sentences for which we have no way justifiable way to decide if they are true or false), members of which are, both according to my account and according to our common experiences as speakers of natural languages, *usually* meaningful. In ordinary, casual, pre-philosophical reasoning, most people are systematically prone to composition

and division fallacies, sloppily sliding back and forth between the properties held by categories and those held by their members, and even professional philosophers who have the reasoning skills to systematically stop themselves from slipping into these errors might still come to the wrong conclusion here by applying an implicit probabilistic inference that if type-A sentences are almost always meaningful, and so are type-B sentences, type-C sentences and type-D sentences, then all else being equal, we can be have excellent reason to think that a sentence that falls into the intersection of all four of these types will be meaningful.

The problem, as I have argued above, is that in the light of theoretical considerations about truth and meaning, considerations whose result in this case is quite specific to the peculiar sub-variety of sentences under consideration, reveal that all is far from equal. This conclusion, however, has been won by argumentation rather than given by an immediate intuitive reaction.

3. As I've just argued, sentences that commit accidental meaninglessness are neither fantastically rare nor especially remarkable in non-paradoxical contexts. That said, I think the real force of objections stemming from examples like Kripke's Nixon case is not a feeling that it should be impossible for a sentence to be accidentally meaningless but from a feeling that it should be impossible for a statement to be *contingently* meaningless.

Let's review the case. Jones says that most of Nixon's assertions about Watergate are untrue. Nixon's statements about Watergate, however, include the claim that what Jones says about Watergate is true. Further assume that, by some bizarre fluke, everything else that Nixon says about Watergate is exactly evenly divided between the

unproblematically true and the unproblematically false. Thus, Jones' statement is true iff it is untrue. This is paradoxical in just the same way as standard Liars, and it clearly comes out meaningless on my account, although not *because* it's paradoxical. (After all, on the one hand, I don't claim, for example that sentences that pose paradoxes of vagueness are meaningless. On the other hand, non-paradoxical empty truth-value ascriptions like the Truth-Teller do come out meaningless.) It is meaningless for the standard reason, which is that it relies on ascribing truth-value to a sentence that isn't directly or indirectly about anything but truth. Nixon's statement and Jones' statement are in a circle where neither one grounds out in any subject other than truth, and this fact is obscured by the misleading use of the phrase "about Watergate" and the irrelevant fact that Jones' sentence attempts to *also* talk about the truth-values of a bunch of meaningful declarative sentences at the same time as applying truth talk to Nixon's ungrounded, meaningless sentence.

This case is no different from in principle from the "Card Paradox," where "The sentence on the other side of this card is true" is written on one face of the card and "The sentence written on the other side of this card is false" is written on the other one. The pair of statements involved in the Card Paradox is obviously ungrounded in any subject other than truth, which makes it meaningless on the analysis being offered, and this would be *no less true* if one of the faces said "The sentence on the other side of this card is true" and the other one said "The sentence on the other side of this card is false, and, by the way, at least half of what Nixon said about Watergate is also false." Here, you have a sentence with the grammatical form of a conjunction, with a meaningless sentence in the place of the first disjunct, which makes the whole "conjunction" as meaningless as

“both bork bork beglork and half of what Nixon said about Watergate is false.”

(Importantly, “The sentence on the other side of this card is false, and, by the way, at least half of what Nixon says about Watergate is also false’ is meaningless *regardless* of how many, or how few lies Nixon told about Watergate. The meaninglessness of it relies in no way on Nixon’s statements happening to be divided evenly between the true and the false.) The unroundedness (and hence, the meaninglessness) may be more *obvious* in the Watergatized Card Paradox than in Kripke’s original formulation, but the logical/conceptual structure of the two cases structure is identical. Either both are meaningless, or neither is.

The real source of the intuitive uneasiness that many people may feel at the notion of “contingent” meaninglessness, I suspect, comes from a suspicion that it reveals that the meaninglessness analysis of Liar-like sentences is an *ad hoc*, “heroic” maneuver to save consistency. The idea is that if someone says something that contingently, due to a bunch of empirically determined-truth values of other statements, happens to be paradoxical, we shouldn’t *on that basis* say that it is meaningless, when, if the world had given some of the other statements other truth-values, we would have considered it meaningful. The reality of a situation can contingently make our statements true or false, but the world turning out the wrong way shouldn’t make an otherwise meaningful statement meaningless.

This intuition has some force, but a few points might help to dispel it. First of all, at the risk of irritating repetition, it can’t be emphasized enough that the meaninglessness analysis of “what Jones says about Watergate is true” and “most of what Nixon says about Watergate is true” isn’t given *on the basis* of the fact that a paradox results from

applying truth-talk to these statements. Many non-paradox-generating sentences are meaningless on this analysis, and the motivations are quite independent of paradox-avoidance. Secondly, on the analysis given above (given the analogy to the Card Paradox), Nixons' and Jones' statements are both meaningless for reasons totally independent of what portion of Nixon's meaningful claims about Watergate happen to be true. As such, even if we completely accept the strongest version of the claim that contingent factors should *never* have an impact on the meaningfulness of our utterances, this evaluation of "what Jones says about Watergate is true" and "most of what Nixon says about Watergate is false" shouldn't be threatened by this intuition.

Of course, we could construct a similar case that's more strongly contingent than Kripke's Nixon case, such that this last point wouldn't apply, and the meaningfulness really would, in a sense, rely on certain background empirical facts. For example, imagine a "contingent Truth-Teller" case. If I am the stupidest person in a room, but I'm too arrogant to realize this fact, I might think "what the stupidest person in the room is thinking right now is true." Since my subvocalized sentence nonsensically applies truth-talk to a sentence that fails to ground out in any subject other than truth, it is meaningless. If, on the other hand, the world were different, such that, say, the man sitting to my left had been given a full frontal lobotomy, my sentence might well be meaningful.

Consideration of this last case brings us to the last point which needs to be made in response to objection 3, which is that it is flatly false that contingent factors never impact the meaningfulness of our utterances. If truth-attributions inherit the entirety of their meaning from the meaning of the statements they attribute truth to (and falsehood-attributions inherit the entirety of their meaning from the meaning of the negations of the

statements they attribute falsehood to), then there is an easy way that contingent factors can impact the meaningfulness of our utterances. If we try to attribute truth to a meaningless utterance, our truth-attribution will itself be meaningless. Since we aren't infallible about the meaningfulness of others utterances, as is established by cases like Josh and Joe the carpenters, or Mary A. and Noel A.'s dinner date, there can easily be situations where the contingent background assumption that what someone said is meaningful is false, so our truth-talk about their utterance is meaningless, but where if that background assumption had happened to be true, our truth-talk would be meaningful.

Of course, one might find the "nothing above and beyond" principle, or our application of it to the question of empty truth-value ascriptions, to be implausible, and if so, this case for contingent meaningfulness will seem correspondingly implausible, but these cases pretty clearly don't count as any sort of *extra* objection to the view under discussion.

4. 5. These objections *do* look like serious problems for the view. No matter how independently plausible the "nothing above and beyond" principle and the meaningfulness analysis of Liar-like sentences that flows from that principle may be, if this analysis falls prey to revenge paradoxes, then we must either abandon that analysis or give up and accept dialetheism. Either way, our proposal will have been seen to fail.

As such, it is to these revenge problems that we must now turn.

Chapter Seven: Meaninglessness and Revenge

OK, so what about (10), which we met above, and which, by analogy to the Liar and the Truth-Teller, we can call the Babblor?

(10) This sentence is meaningless.

On the assumption that the meaningless solution is right, (10) is, after all, the relevant instance of what we can call the Universal Revenge Paradox (URP):

URP: Whatever the correct solution to the Liar Paradox turns out to be is also the correct analysis of this sentence.

Surely, one of the criteria for any plausible solution to the Liar Paradox is that accepting it doesn't deprive a theorist of the ability to say something plausible about the relevant URP sentence. Gap theory, for example, does very badly, since the usual motives for calling the Liar and the Truth-Teller gappy seem to apply just as well to a sentence that says of itself that it is gappy, and this leaves the gap theorist with no plausible way around the paradoxical result that such a statement is both gappy and true. By contrast, dialetheism passes the URP test with flying colors, since saying that a sentence that says of itself that it is both true and false is both true and false is no more implausible than saying the Liar is both true and false, and creates no further difficulties.

Obviously, a certain sort of cranky orthodox logician might object that saying that the dialetheist analysis of the relevant URP instance is no more implausible than the theory itself is like saying that shooting a corpse in the head makes it no more dead, but regardless of the merits of that analogy, that objection is an entirely separate issue. *On the*

URP question, dialetheism does well. If we find dialetheism implausible, we need to do just as well in our treatment of the URP instance of our favorite theory.

Now, if the reason we regard sentences the Liar and the Truth-Teller as meaningless is that, a la some formulations of Kripke's truth-value gap theory, they fail to ground out in "extra-semantic reality," the Babblers poses a serious problem. It also clearly fails to ground out in extra-semantic reality, so by parity of reasoning, we should say that it is meaningless.¹⁰⁹ But this is precisely what it claims of itself, so if it is meaningless, it is *also true*. This is, pretty clearly, conceptually impossible, since we can only meaningfully attribute truth or falsity to meaningful sentences. Something, somewhere in our reasoning, has gone horribly wrong.

The problematic assumption in the last paragraph was the utterly unmotivated claim that sentences must ground out in "extra-semantic reality" in order to be meaningful. This assumption should be rejected, not *just* because it lands us in the Babblers Paradox, but because there is no reason whatsoever why we should suppose that it is true. The independent motivation sketched out above for the claim that truth-value-ascriptions like the Liar and the Truth-Teller need to inherit their meaning from the meanings of the sentences to which they ascribe truth or falsity was *specific* to the concept of *truth*. There is no reason to suppose that a similar analysis should hold for *all* semantic concepts without exception, and excellent reason (quite independent of our desire to avoid the Babblers Paradox) to suppose that it does *not* apply to the concept of meaningfulness.

¹⁰⁹ Remember that I'm not using "ground out" in any *formal* sense, but merely referring to the intuitive core of the idea. Since, when it came to truth, I said that grounding out meant being paraphrasable in a way that did not use 'true,' the counterpart for grounding out in 'extra-semantic reality' would be paraphrasability in a way that doesn't use *any* semantic concept. This would be an extremely hard standard.

The meaningfulness predicate is *nothing* like the truth predicate. The truth predicate functions as a shorthand device for expressing agreement with claims without having to waste time repeating the claim. It's particularly useful for long lists of claims ("yeah, that's all true") and more useful yet for claims with which the speaker wants to express agreement even without knowing their content (i.e. "blind truth endorsements" like "everything the Pope says is true"). By contrast, meaningfulness-attributions indicate absolutely nothing about whether the speaker agrees with or disagrees with or reserves judgment about the sentence that he is attributing meaningfulness to except that they think that there is something there to agree with, disagree with or reserve judgment about. To apply truth-talk to a sentence is ultimately to talk about the subject of that sentence. By contrast, when we attribute meaningfulness or meaningless to a sentence, the subject of our meaningfulness-talk *is the sentence*, rather than whatever the sentence is about. When someone asserts that "'snow is white' is meaningful," they are no more talking about snow than is the speaker who asserts that "'snow is white' has three words in it."

Since the content of truth-attributions is inherited from the content of the statements to which truth is being attributed (e.g. the content of "it's true that 'snow is white'" is the attribution of whiteness to snow), truth-attributions are meaningful when directed to meaningful statements and meaningless when directed to meaningless statements. Meaningfulness-attributions aren't like that, and couldn't be like that if the meaningfulness predicate was to serve any kind of informative function in our language *at all*. After all, the point of talking about meaningfulness is to sort out the meaningful sentences from the meaningless ones. As such, attributions of meaningfulness to

meaningless statements had better be themselves meaningful and, in fact, *true*.

Otherwise, what's the point?

When we say of a sentence that it is meaningful or meaningless, we are not expressing agreement, disagreement or anything of the kind with the subject matter of the sentence. We are, rather, making a substantive, meaningful claim about the sentence. No fact about the sentence to which meaningfulness or meaninglessness is being attributed could ever render an attribution of either meaningless.

With this in mind, let's go back to the Babler Paradox. What should we say about (10)? The answer is simplicity itself. (10) is false. Asserting of any sentence that it is meaningless is always a meaningful claim, and a sentence that says of a meaningful claim that it is meaningless is always false. No contradiction, strangeness or conceptual absurdity follows from (10) being false.¹¹⁰

One might object here that (10) has the stench of trickery about it that, when it was present in the Liar case, was the motivating reason for the meaninglessness solution. A crude way to express an initial suspicion of meaninglessness about the Liar might be, "wait....*what* statement is false?" One might have the same gut reaction to the Babler. "Wait....*what* statement is meaningless?" The problem with this objection is that, before we decided that the Liar *really was meaningless*, we had to back up that vague, diffuse gut reaction with a well-motivated philosophical story about *why* sentences like the Liar were meaningless. No such story seems to be available about meaningfulness for the Babler, and the story we told about truth for the Liar was specific to truth.

¹¹⁰ Similarly, "this sentence is meaningful" is true, "this sentence poses a revenge problem for the correct view of the Liar Paradox" is false and "this sentence does not pose a revenge problem for the correct view of the Liar Paradox" is true. I see no reason to think that *any* of these sentences are meaningless.

It is of course possible that there are other semantic concepts about which we can tell a story similar to the one we told about truth, such that attributions of that concept to a sentence are only meaningful if they “ground out” in statements about something other than that concept. There is no reason to rule out *a priori* the possibility of other cases like this, and some reason to suspect that there may turn out to be other such cases. However, the reason we imposed this requirement on truth was *specific* to the way the truth predicate seems to function, and any other allegedly analogous concepts would have to be evaluated on a case by case basis, on their own merits. In the case of the meaningfulness-predicate, we’ve seen that, for reasons quite independent of avoiding the Babler Paradox, we’ve seen that it can’t be like truth in this respect, and once that realization has been made, the paradox disappears.

Now, a further question naturally arises at this point. Although (10) has been shown not to be any kind of troubling revenge paradox for the meaningfulness solution the Liar Paradox, it might be claimed that the Strengthened Liar (20, above) still is. Earlier, we dismissed (20) as a challenge to the meaningfulness solution on the grounds that when we say that (20) is meaningless, we aren’t asserting the negation of the claim that it is true and the negation of the claim that it is false, we are asserting that attempts to apply the concepts of “truth,” “falsity,” “negation,” etc. to (20) are themselves meaningless.

That said, the challenge to the meaningfulness solution from (20) could re-assert itself here for new and more sophisticated reasons. Let’s start with the obviously-true-sounding claim that for any sentence α , if α is true, it’s meaningful. (After all, since our solution hinges on the claim that we are talking nonsense if we apply a word like “truth”

to a meaningless statement, it would seem strange if we rejected this claim.) We might formalize this claim as the “meaningfulness schema” (MS).

$$\text{MS: } T\langle\alpha\rangle \rightarrow M\langle\alpha\rangle$$

The problem should be immediately clear. Let α be (20), which, remember, was

(20) The sentence marked (20) is not true.

We assert $\neg M\langle\alpha\rangle$, by way of expressing our solution to the paradox. By Modus Tollens, $\neg T\langle\alpha\rangle$which is exactly what α asserts of itself! This looks like a serious problem.

By way of getting a handle on what has gone wrong here, notice that, by treating (20) as the sort of thing properly symbolized by a Greek letter, the kind of thing we can perform logical operations on, we’ve committed a category mistake. Both premises of our little syllogism....

1. $T\langle\alpha\rangle \rightarrow M\langle\alpha\rangle$
2. $\neg M\langle\alpha\rangle$
3. $\neg T\langle\alpha\rangle$

....are meaningless, and to deny this is to beg the question against the meaningfulness solution.¹¹¹ That said, dismissing the problem on these grounds alone might smell like letting off a guilty man on a technicality. After all, even if we are doing something illegitimate by expressing the claim that (20) is meaningless as $\neg M\langle\alpha\rangle$, there

¹¹¹ One might be concerned that self-referential incoherence looms large here. By saying that 2 is meaningless, aren’t I saying that my own stance on ungrounded truth talk is meaningless? After all, aren’t I saying the same thing that 2 says? The answer is that I don’t take attributions of meaningfulness to meaningless sentences to be meaningless. I do not, however, think that this attribution can be meaningfully conveyed by formalizing such sentences with Greek letters and trying to apply logical operations to them. To try to do so is to commit a nonsensical category mistake. Similarly, one can meaningfully (and correctly!) say of a cough that it is not a sentence, but if one tries to do so with a negation symbol, a “is a sentence operator” and a Greek letter symbolizing the cough, the result is meaningless nonsense.

must be some good way of expressing the claim....after all, we've shown above that there are no sentences about which attributions of meaningfulness are not meaningful. So let's re-phrase it in terms that don't involve symbolizing or performing operations on (20) itself. In our new, simpler formulation, α as the claim that (20) is true and β as the claim that it is meaningless, so the whole thing now simply reads:

1. $\alpha \rightarrow \beta$
2. $\neg\beta$
3. $\neg\alpha$

Now, there can be no claim that the second premise is meaningless. It is clearly meaningful. However, the first premise is still meaningless, because it applies truth talk to a meaningless sentence, and treating *that* claim as if it were a meaningful declarative sentence, symbolizing it with a Greek letter and trying to perform logical operations on it is *still* nonsensical.

This time, no charge of wiggling out of a serious problem by technical nit-picking is possible. There is no way to formulate any sort of "if it's true, it's meaningful" conditional for a meaningless sentence without the antecedent (and hence the conditional as a whole) being meaningless. We can even meaningfully (and correctly) announce as a general principle that for any statement, if it's true, it's meaningful, but when we try to formulate a conditional to that effect in any case for which the consequent is false, the antecedent (and hence the conditional as a whole) will be meaningless.¹¹²

¹¹² To head off a possible concern, this does *not* mean that the statement can be both meaningful and correct even if some of its instances are not. The point is just that, like any logical principle starting with "for any statement," it can only quantify over meaningful declarative sentences. Thus, *for all of its instances*, the antecedent will be true. Compare to: "For all statements, if the statement is the kind of thing which a meaningful logical generalization can quantify over, the statement is the kind of thing that we can symbolize with a Greek letter and write a 'T' or an 'F' under in a truth table." The idea that there could be

At this point, one might still have a lingering feeling that something has gone wrong here, a vague sense that, however clear the reasons why we can't in this case, we *should* be able to reason from the falsity of the consequent of that conditional to the falsity of the antecedent, and any fancy philosophical story about why we can't must be misleading in some way. The underlying feeling might be a gut reaction that for *any* X and *any* Y, if the conditional "if something is an X, it's a Y" is meaningful in all cases where the consequent is true, it should also be meaningful in cases where the consequent is false.

Any such gut reaction, whatever its sources, should be defused by thinking about a closely related predicate where the situation seems much clearer. What about "being the kind of thing which meaningful conditionals can be constructed about"? Now, pretty clearly, all all propositions, or precissified sentences, or sentences in contexts, or whatever one takes truth-bearers to be, are the kinds of things that we can construct meaningful conditionals about, regardless of whether they are true or false, so "if Sentence X is true, Sentence X is the kind of thing about which meaningful conditionals can be constructed" seems like a meaningful statement for any truth-bearing Sentence X. However, equally clearly, for any sentence about which "Sentence X is the kind of thing about which meaningful conditional can be constructed" is false, we can't construct a meaningful conditional about it. So here again we have a case in which, for all cases where the consequent is true, the conditional is meaningful, but, when it is false, it isn't.

instances of this generalization such that the antecedent of the conditional was false is absurd, but the generalization seems quite correct in any case. In a way, the situation is the parallel to the one that arises for existence talk if one sides with figures like Russell and Quine against Meinong and takes it that there being a value of a bound variable is a matter of *reference to objects*, and that 'reference to non-existent objects' is an incoherent notion. Any time 'X exists' is true, we can, if you like, express that with an existence predicate. Any time, we try to say 'X doesn't exist' with an existence predicate, however, a contradiction ensues...as it should, given the assumption that there is no object X to refer and attribute anything to unless X exists!

Given this case, there's no denying that *in some cases*, conditionals of a certain form can only and always be meaningful when they have true consequents.

Once this has been established, any feeling that *in the case of the simple meaningfulness predicate*, we should be able to infer from "Sentence X is meaningless" to "Sentence X is not true" comes down to simple disagreement with the claim that attributions of truth need to inherit their meaning from the statements they attribute truth to. As such, no (non-question-begging) objection to the meaningfulness solution can be derived from this line of thought. The whole point of the solution is that the sort of mistake one makes when one attributes truth to a meaningless sentence isn't a mistake about which truth-value is the right one, but a category mistake about whether it is the sort of thing to which truth-talk applies.

Of course, this observation could very naturally lead to a new and even more troubling revenge paradox for the meaningfulness solution. What about (30)?

(30) It would be a mistake to call this sentence true.

One is making one sort of mistake in calling meaningless sentences true, and another sort of mistake in calling false sentences true, so all the conceptual options seem to be bad here. If it's true, it would be a mistake to say so, which is absurd. If it is false, then it would be a mistake to say that it's true, so it is true. If it is meaningless, then it would be a mistake to say that it's true. We have a problem on all options.

Now, we could plausibly argue here that (30) is actually ambiguous between being a disguised version of (31) or a disguised version of (32).

(31) One would be making a *category mistake* if one called this sentence true.

(32) One would be mistakenly picking the wrong truth-value if one called this sentence true.

Now (32) is just a form of the Babblers, and (33) of the Strengthened Liar, and in both cases, we've shown that neither works as a revenge paradox against the meaninglessness solution. If (31) is ambiguous between the two, what we can say about it is fairly straightforward. "Well, it depends what it means. If you mean one thing, then the sentence is simply false. If you intend to mean another thing, the sentence is not the kind of thing that we can meaningfully apply truth-talk to."

However, this might get us off the hook too easily. What if (30) can be read, not as *ambiguous*, but as *disjunctive*? We can re-phrase it slightly, as (33), below, to capture this.

(33) It would be a mistake *of some sort* to call this sentence true.

Now, there's no ambiguity. If I say that the ink in a pen "is a color that's also on the American flag," what I'm saying is not *ambiguous* between saying that the ink is red, saying that the ink is white and saying that the ink is blue. Rather, I'm making a perfectly clear, non-ambiguous statement that would be made true by any of those possibilities. And again, if it's true, it would be a mistake to say it, which seems unacceptable, if it's false, it's true, and if it's meaningless, it's true, so we have absurdity all around.

To get a handle on what's gone wrong here, the first step is to notice that instead of being ambiguous between being (31) in disguise and being (32) in disguise, (33) seems to be perfectly biguous as being something like their disjunction. Whatever we say about (33) should be the same as what we say about (34), which is doing the same thing in slightly different language.

(34) This sentence is either false or meaningless.

Now, I say “something like the disjunction” because (34) is not *actually* meant to be a disjunction of a Babblor referring only to itself, and a Liar attempting to refer only to itself. Rather, it’s supposed to be a disjunction between the claim that the whole sentence is meaningless and the claim that the whole sentence is false. To try to get a better handle on this point, think about other sentences with similar forms, like (35).

(35) This sentence is either false or composed of ten words.

(35) is undeniably composed of ten words, but that doesn’t make it true. In fact, it doesn’t even make it the kind of thing we can meaningfully apply the language of truth and falsity to. To see why, consider (36).

(36) Either blork glork melork or this sentence sentence is labelled (36).

The fact that the second disjunct is true hardly suffices to make the whole disjunction meaningful, much less true. Now, if the analysis of the Liar argued for above is the right one, the first disjunct of (35) is as meaningless as the first disjunct of (36). One may be convinced or unconvinced by that argument, but one way or the other, that’s hardly relevant to any discussion of whether *this particular objection* to the meaningless solution goes through. To take seriously that question, one has to take seriously at least for the sake of argument the claim that the former is no more meaningful than the latter.

One might object that the analysis of the meaningfulness of Liar sentences doesn’t apply to the first disjunct of (35) on the grounds that it isn’t an entirely *empty* attribution of falsity. After all, one is saying something meaningful (and false) by calling the second disjunct false. The problem is what “extra content” the first disjunct of (35) is supposed to have, above and beyond calling the second disjunct false. (After all, if that *all* it was

doing was calling the second disjunct false, the whole disjunction would be a fairly uninteresting tautology. That's pretty clearly not what's going on there.) That extra content is supposed to come from its attribution of falsity to itself, and there, for precisely the same reason as all other Liar-like sentences, it fails to be meaningful, since there's no statement about something other than truth from which the truth-value attribution can inherit its meaning. As such, the first disjunct of (35), no less than the first disjunct of (36), is meaningless, and disjunctions with meaningless disjuncts are themselves meaningless.

If I point to a rock and say "either that or snow is white," and I ask what you mean--"do you mean either the rock is white or snow is white?"--and you say, no, no, I'm not making a statement *about* the rock, I'm using the rock as a disjunct, then I'm well within my rights to accuse you of making a nonsensical category mistake. The "disjunction" doesn't come out true, because if we use the word "disjunction" not just to mean uttering a string of words with "or" in the middle but as a word for compound statements of a certain form ($\alpha \vee \beta$) on which we can perform further logical operations, etc., there *is no disjunction* in the rock case.

Similarly, neither (35) nor (36) is a real disjunction, and hence there isn't one in (34), never mind that, like the other three cases being discussed, the "second disjunct" is true. (To dispense with an obvious question, "either this sentence is false or this is not a real disjunction" is not a real disjunction, but there's no reason for that to be more of a problem than in the case of any of the other nonsense "disjunctions" with true "second disjuncts.") Having gotten straight on all of this, we can go back to (33).

If (33) is trying to say precisely what (34) is, and, as just argued, there's no problem or implausibility involved in saying that (34) is meaningless, then there shouldn't be any problem or implausibility involved in saying that (33) is meaningless. We put the "of some sort" into (33) to rule out any claim--like we made about (32)--that (33) is ambiguous, and to make it clear that it is instead (trying to be) disjunctive. Well, the implied "disjunction" is precisely the same in (32) as it is in (33). Making a mistake *of some sort* means that either we are mistakenly picking the wrong truth-value or it is a mistake to think that truth-talk applies at all to (33). As before, the "second disjunct" is meaningful and in fact true, but that hardly suffices to make the whole "disjunction" meaningful.

Due to the fact that the attempted "disjunction" is left implicit by the surface grammar, it may seem harder to swallow this analysis for (33) than it was for (34), but the logical structure of the considerations for and against taking it as meaningless, for and against taking that meaninglessness to pose some sort of revenge problem, are *identical* in both cases. What makes (33) different from (30) is that (33) tries to make a disjunctive claim rather than an (ambiguous) simple one. If we don't read it disjunctively, there's no more problem there than there was for (30). If we do, then there's no more problem with (33) than there is with (34). In both cases, "disjunctions" of meaningless pseudo-statements on the one hand and meaningful statements on the other cannot be "made true" by the truth of the meaningful disjunct, since talk of truth or falsity simply doesn't meaningfully apply to such "disjunctions." If one rejects this analysis for (33), they must reject it for (34) as well, and rejecting it for (34) surely commits one to either rejecting the meaninglessness analysis of Liar-like sentences *on grounds logically independent of*

this particular objection to the meaningless solution, or committing the severe implausibility of rejecting it for (35), (36) and the rock case as well. Since the second possibility seems inadvisable, the only way to rationally reject our analysis of (33) (without, of course, begging the question) is to find some other grounds to reject the meaningfulness solution to the Liar Paradox. That's fine, and of course, as with any theory, there's no way to rule out its merely possible refutation in advance, but the fact that this is what it takes to get around this analysis of (33) pretty clearly shows that there is no good objection to the meaningfulness analysis *from the difficulties raised by sentence (33)*. Since that was what was at issue, that's certainly good enough for now.

Moreover, we can take a long step back and look at the state of play in the larger arena of defending the meaningfulness solution against alleged revenge problems. The solution has stood up against the Strengthened Liar, against the Babblers, against sophisticated attempts to re-introduce the Strengthened Liar, and against all sorts of attempts to formulate revenge paradoxes based on the language used to discuss meaningless sentences, including a variety of disjunctive paradoxes. None have made any headway.

Of course, it's always possible that some future, yet-to-be-formulated revenge paradox could sink the meaningfulness solution, and it would be necessary to go back to the drawing board. In principle, there's no way that can ever be ruled out, and that's fine. All theories should be believed fallibilistically. (Including, of course, the theory that all theories should be believed fallibilistically.) However, based on the state of things so far, we seem to have ample grounds for at least extreme probabilistic confidence in the following claim: Any revenge paradox that posed a real problem for this solution would

be radically unlike any type of standard revenge paradox discussed in the literature. Most revenge paradoxes for solutions to the Liar Paradox can be posed by a clever objector after five or ten seconds of thought, using general patterns of revenge-taking to which it's fairly easy to get a feel for after consideration of a dozen or so examples of initially plausible but failed solutions. Those sorts of revenge attempts have been extensively surveyed and seen to fail.

If the meaningfulness solution is brought down, it won't be brought down by more of the same.

One final note about this before we move on: someone could object to our analysis of sentences with the grammatical form of disjunctions, but which rely on an attempt to apply truth-talk to meaningless statements, on the grounds that the truth-functionality of disjunctions make them true if they have one true disjunct, *regardless* of the status of the other disjunct. "Either what Cheech said yesterday afternoon at 4:20 is true or the sky is blue" is well-formed, and as such, we can be sure that it is true just by virtue of the blueness of the sky and the relevant logical rules, even if we don't have the slightest idea what Cheech said at 4:20. Maybe Cheech said something true, maybe he said something false and maybe he even said something meaningless, like "whoah, dude, what I'm saying right now is, like, totally true!" It doesn't matter. If we write " $\alpha \vee \beta$ " on a chalkboard, symbolizing "what Cheech said yesterday afternoon at 4:20 is true" as α and "the sky is blue" as β , then once we've written a little "T" under β , we don't need to write anything under α before we go ahead and write a little "T" under the " \vee " to signify that the whole disjunction is true, since it's irrelevant to the outcome whether α gets a T, an F, or whatever.

Now, if the meaningfulness analysis of Liar-like sentences was meant as a version of gap theory, such that meaningless sentences were the sorts of things we could meaningfully apply truth-talk to, but they had a third truth-value, such that we wrote little M's under them in truth tables, this would be an extremely serious objection. Not necessarily a decisive one, but certainly not one that could be easily brushed aside. The reason that it wouldn't necessarily be decisive is that it involves importing rules regulating disjunctions in classical logic over to the new, three-valued context, and it's not immediately obvious that we should do so. In classical logic (and other logical contexts where Bivalence is assumed), $\alpha \vee \beta$ is always true if α is true, because both of the combinations of truth-values that are assumed to be possible at that point (α and β both being true, or α being true and β being false) make $\alpha \vee \beta$ true. Whether to regard the combination of one disjunct getting the third truth-value and the other being true as making the disjunction true or not (and hence, whether to keep the law of disjunction addition in the non-classical, three-valued logic one is adopting) is a complicated question, and (assuming for the sake of argument that gap theory is correct) it's far from clear to me what the right answer is.

Fortunately, this is a question that does not arise on my view, since the meaningfulness analysis that I'm proposing is *not* a version of gap theory. In fact, I've explicitly rejected the possibility of gaps, and I regard the fact that the meaningfulness analysis does not force us to abandon classical logic as one of its chief virtues. With this point rather firmly in mind, let's go back and re-examine the Cheech case. When trying to frame the objection in the most devastating possible way, we said that once we knew that β (the sky is blue) got a T, we could go ahead and assign a T to $(\alpha \vee \beta)$, since "it's

irrelevant to the outcome whether α gets a T, an F, or whatever.” The “or whatever,” however, totally misrepresents the nature of truth-functionality in classical logic. The whole point is that there are only two possibilities (α gets a T, or it gets an F), and, in either case, we have a combination of truth-values of the parts of $(\alpha \vee \beta)$ that guarantees that the whole compound statement gets the truth-value “true.”

As I argued above, applying talk of “well-formedness” to ordinary language sentences involves making an idealized, unhelpful and poorly motivated link between the formation rules of formal logics and the grammatical rules of natural languages. Being grammatically “well-formed” is neither necessary nor sufficient for being a meaningful declarative sentence, the kind of thing we can meaningfully apply truth-talk to. Moreover, even in the context of formal logic, uninterpreted wff’s like $(\alpha \vee \beta)$ aren’t the kind of thing that can be true or false. (We can construct a truth-table for an uninterpreted wff, but there’s only a “correct line” of that truth table when we interpret it.) Only *interpretations* of wff’s can be true or false, and the whole point of the meaninglessness solution to the semantic paradoxes is that, when you try to take a meaningless pseudo-statement like the Liar or the Truth-Teller as the interpretation of a wff or of any part of one, you are making a greivous category mistake. Such meaningless utterances simply aren’t the kind of thing that can be true or false, that can have logical operations performed on them, or that can be meaningfully taken as the interpretation of a Greek letter appearing in a logical formula. As philosophers of the Vienna Circle were so fond of pointing out, the negation of nonsense is nonsense. There is no reason that disjunction, conjunction and the rest should be any different from negation in that regard.

Crucially, the same goes for conditionals, for the same reason. This means that sentence (37), below, is unproblematically meaningless for the exactly the same reason that the Liar and the Truth-Teller are meaningless.

(37) If the sentence labeled (37) is true, then absolutely everything is true and reasoning is impossible.

As we will see, this represents a distinct advantage for the meaningfulness analysis over the dialethic analysis. The former is a *unified* solution to all of the semantic paradoxes involving self-referential truth-talk. The latter is not.

Chapter Eight: There Are No True Contradictions

As we've seen, the dialetheist analysis of change, of obligations and of set theory fails. Even the best, strongest argument for true contradictions, the argument from the Liar Paradox, relies on the mistaken impression that empty truth-value descriptions can somehow be meaningful. The mere fact, however, that Priest's case for dialetheism fails does not give us a good reason to reject the existence of true contradictions, any more than the failure of Descartes' anti-skeptical arguments in the *Meditations* gives us a good reason to reject belief in the existence of the external world. In the absence of a better reason, what we've arrived at, at best, is an impasse between monaletheism and dialetheism.

Let's see if we can do better than that. To begin with, we need to clearly differentiate between four possibilities:

- (1) There are no true contradictions.
- (2) True contradictions are possible, but very few contradictions are true.
- (3) Not all contradictions are true, but a great many of them are.
- (4) All contradictions are true.

(1) is monaletheism, and (2)-(4) are the dialetheist options. (4) is trivialism, the position that *everything* is true (and false). For obvious reasons, much of the project of trying to carve out a remotely plausible version of dialetheism is about trying to stop the slide from the rejection of (1) to an embrace of (4). For reasons that are less obvious but no less important, another large part of that project is about the rejection of (3). Embracing (3) might not amount to the sort of psychotic break from reality that an

embrace of (4) would be, but it would have disastrously counter-intuitive consequences for our practices of reasoning. We'll come back to that question in a moment. For now, let's start with the seemingly easier problem of resisting the slide from dialetheism to trivialism.

We've already seen (back in Chapter Two) that one can't use the explosion proof to reduce dialetheism to trivialism without begging the question. This does not by any means, however, settle the question of how the two views are related. I will argue that, first of all, that a partisan of Liar-based dialetheism is in the awkward situation of not being even being able express their disagreement with trivialism, and secondly that they have no good, non-question-begging way to block the argument from Curry's Paradox to trivialism. (By "Liar-based dialetheism," I simply mean any version of dialetheism in which sentences like the Simple and Strengthened Liars are held to be both true and false. In principle, of course, one could be a dialetheist without believing those particular sentences to be dialetheic.) Thus, in order to avoid the slide to (4), dialetheists must give up something that is supposed to be one of the central attractions of their view, its allegedly simple, unified and satisfying analysis of "spandrels of truth" like the Liar.

Worse yet, we'll see that they need to do the same thing even to express their rejection of (3). Finally, I will end the chapter by arguing that, once we start to think through the epistemic consequences of granting that contradictions can be true, we find that *even if* dialetheists are willing to abandon their analysis of the Liar Paradox, they still don't, and can't, have a good, principled reason to stop short of (3).

I. Failures of Truth?

Intuitively, the distinction between Priest-style dialetheism on the one hand and trivialism on the other is that, according to the latter, *everything* is true (and false), while, according to the former, some sentences are both true and false but others are just true or just false. Just as Priest calls sentences that are both “dialetheias,” we can call sentences that are only true, or only false “monaletheias.” Since Priest’s position is (2), he believes that the overwhelming majority of sentences are monaletheias, while a small minority are dialetheias. So far, so good.

The problem, it seems to me, is that the (Liar-based) dialetheist has nothing plausible to say about sentences like (38) and (39).

(38) This sentence is (just) false.

(39) This sentence is a false monaletheia.

Recall that Priest (quite correctly) rules out truth-value gaps. That leaves him with three possible alethic statuses to assign to these sentences. They could be (just) true, (just) false or both true and false. By standard Liar reasoning, if (38) is (just) true, then it’s also (just) false, since that’s what it says of itself. If it’s (just) false, then it’s also true. If it’s both true and false, then it’s true, so it’s (just) false. On all three of the three possible answers for a Priest-style dialetheist, “(just) false” turns out to overlap with true, and we’re left with the uncomfortable conclusion that some statements are both (just) false and true. Similar results hold for (39). If (38) is true, then it’s both true and a false monaletheia. If it’s false, then either (a) it’s a false monaletheia or (b) it’s a dialetheia. On option (a), since that’s what it says of itself, it once more turns out to be *both* a false

monaletheia *and true*. On option (b), it's true, so it's *still* both true and a false monaletheia. So when a dialetheist says that "some sentences are (just) false" or "some sentences are false monaletheias," she hasn't succeeded in expressing the notion that some sentences have a status that's incompatible with truth, or that she rules out their truth, or that such sentences fail to be true or that nothing makes them true.

Of course, even the previous sentence *expressing the problem* can't do what it intuitively seems to be doing if we accept Liar-based dialetheism. After all, consider a few more Liars.

(40) This sentence has a status such that having that status is incompatible with being true.

(41) This sentence has a status such that having that status rules out the possibility that it is true.

(42) This sentence fails to be true.

(43) There is nothing that makes this sentence true.

(44) This sentence is one that is properly assigned the value 0 in Priest's logic LP.

As before, standard Liar reasoning will get us the result that some sentences are true *and* have a status that's incompatible with truth, that some sentences are true *and* fail to be true, and that some sentences are true even though nothing makes them true. The general lesson should be getting painfully clear by now.

Along the lines of the suggestion about classical and paraconsistent negation from Koji Tanaka quoted in Chapter Two, a desperate dialetheist might argue that the contrast

class to dialetheias can be indicated by saying of some sentences that if they are both true and false, then everything is true.¹¹³ For several reasons, that won't work.

First of all, note that for the alethic status of the antecedent to be enough to guarantee the truth of the conditional, the conditionals in question *must* be the simple truth-functional conditionals of classical logic.¹¹⁴ After all, consider sentence (45), below.

(45) If the moon is made of green cheese, then everything is true.

If we read the “if...then” as a relevant conditional, where the standards for the conditional coming out as true are higher than the standards for the classical conditional coming out as true, then (45) is false. Similarly, the counterfactual version, (46), is obviously false, on any recognizable (and, of course, on any remotely plausible) account of counterfactual conditionals.

(46) If the moon were made of green cheese, then everything would be true.

¹¹³ Obviously, a Liar sentence can be constructed that says of itself that it false and a member of the contrast class to dialetheias.

¹¹⁴ One might think that classical logic could get the truth conditions of conditionals wrong, but that conditionals could still be truth-functional, and in particular that on some other truth table could still deliver the result that all conditionals with antecedents that are (just) false are true, but, given that the conditional is truth-functional, it's hard to see what other truth table conditionals could plausibly have. (Of course, if there are more truth-values than “true” and “false,” conditionals could be truth-functional without the classical truth table telling the *whole* story, but the story the classical truth table tells would still be right as far as it goes.) If Modus Ponens is going to be universally truth-preserving, the first line of the classical truth table must be right. Given that everyone agrees that at least *some* conditionals with false antecedents and false consequents are false, if the conditional is going to be truth-functional, the fourth line of the classical truth-table must be right. The second and third lines are guaranteed by the intuitive asymmetry of conditionals—for example, given that John McCain, Barack Obama and Ralph Nader all ran for President in 2008, and that only one person can win, “if Nader won, McCain lost” seems to be true, but “if McCain lost, Nader won” seems to be false. Again, given truth-functionality, if *some* conditionals with false antecedents and true consequents are true, then *all* of them must be true, and the same point of course goes for conditionals with true antecedents and false consequents being false. Of course, if dialetheism is right, it *can't* always be true that conditionals with true antecedents and false consequents are *always* false (remember our discussion of the failure of Modus Tollens in dialethic contexts in Chapter Two), but given the fact that everyone would agree that *some* such conditionals are false, so much the worse for the availability of truth-functional conditionals for dialetheists.

Consideration of the rest of the options should quickly show that only the truth-functional conditional of classical logic fits the bill. The problem is that the conditional of a dialetheist's logic *can't* be the conditional of classical logic.

After all, in classical logic, $(\alpha \rightarrow \beta)$ has the same truth table as $(\neg \alpha \vee \beta)$. Some critics of classical logic are fond of the rhetorical flourish of saying that the conditional of classical logic just *is* $(\neg \alpha \vee \beta)$, but this seems like a considerable over-statement, at least in the absence of a persuasive argument for universally equating sameness of meaning with sameness of truth conditions. To borrow an example from Quentin Smith, it's hard to see how "the sun has a shape" could be true without "the sun has a size" being true and vice versa, but intuitively, the two statements *mean* very different things.¹¹⁵ Moreover, given our previous endorsement of the deflationary "nothing above and beyond" principle about truth, it's hard to make sense of the idea that truth could play an explanatory role in discussions of meaning.¹¹⁶

In any case, however that may be, the truth-functional conditionals of classical logic are *true* under the same circumstances as the disjunction of their consequents with the negations of their antecedents, and this alone makes them unavailable to the dialetheist, at least the dialetheist who wants to continue to accept that Modus Ponens is universally truth-preserving. After all, given that equivalence of truth conditions, plus Double Negation, all circumstances under which α and $(\alpha \rightarrow \beta)$ are both true are *also*

¹¹⁵ Smith (1993), p. 4

¹¹⁶ Interestingly enough, JC Beall rejects the claim that the meaning of connectives is given by their truth conditions on precisely such deflationary grounds, but nonetheless insists that "the hook" (i.e. the conditional of classical logic) just *is* $(\neg \alpha \vee \beta)$. (See Beall (2009), p. 118 on the former point, and pp. 25-26 on the latter.) If he has an argument for this identification independent of a general assumption that the meaning of connectives is given by their truth conditions, he keeps it to himself.

circumstances under which $\neg\neg\alpha$ and $(\alpha \vee \beta)$ are true.¹¹⁷ Thus, the price of jointly accepting that conditionals are truth-functional and that Modus Ponens is valid is accepting that Disjunctive Syllogism is valid as well.¹¹⁸ Given non-triviality, this is something that the dialetheist cannot do.

Given that (45) only comes out as true if the “if...then” is understood truth-functionally, and that dialetheist’s *can’t* understand “if...then” truth-functionally, such conditionals *can’t* be used to express the exclusive falsity of any claims. One might try to save the basic outline of the suggestion by using entailment instead of conditionals, thus claiming that “the moon is made of green cheese” *entails* everything.

The problem, of course, is that it *doesn’t*. Since “the moon is made of green cheese” is symbolized by a single Greek letter, thus giving it the same logical form as, say, “the moon is a physical object,” no rule of either classical or paraconsistent logic allows us to derive everything from it. Of course, in *classical* logic, the entailment goes through when *combined with the background premise that “the moon is made of green cheese” is false*, but this is just explosion, the one rule that, by definition, no paraconsistent logic can include.

¹¹⁷ Not only does the assumption that dialetheism is true give us no good principled motivation for rejecting Double Negation, but Double Negation plays a starring role in certain arguments for dialetheism, particularly in reducing gaps to gluts.

¹¹⁸ One might think the dialetheist could compromise here and have two conditionals in their language, a classical one \rightarrow such that $(\alpha \rightarrow \beta)$ is logically equivalent to $(\neg\alpha \vee \beta)$ and, perhaps, a relevant one \rightarrow^* . Modus Ponens would be valid (but only for \rightarrow^*) and Disjunctive Syllogism would still be invalid. The “non-detachable” classical conditional would be kept around for one purpose only, which would be as a way of expressing the exclusive falsity of sentences by tying them to triviality. Even this suggestion, however, falls flat. After all, for *every* true contradiction $(\alpha \wedge \neg\alpha)$, there will be a true disjunction $(\neg\alpha \vee \beta)$ where β is “everything is true.” (For example, if it’s both the case and not the case that the Russell Set is a member of itself, it’s not the case that the Russell Set is a member of itself, and if it’s not the case that the Russell Set is a member of itself, then either the Russell Set is a member of itself or everything is true.) Thus, on this two conditional scheme, dialetheists would be constantly asserting $(\alpha \rightarrow \beta)$ where β is “everything is true,” and they believed the antecedent to be both true and false. Thus, such conditionals *couldn’t* be used to express exclusive falsity.

Of course, given the validity of Modus Ponens, the inference would go through even in paraconsistent logic if we added the premise that “if the moon is made of green cheese, everything is true,” but, for reasons we’ve just been through, that premise simply can’t be true in a dialethic logical setting where Modus Ponens is still valid. I can only see one remaining way of making the inference go through, which is to add something like a “super-falseness” operator to the dialetheist’s logical language, $F!$, such that, given $F!(\alpha)$, α *would* entail everything. The problem, obviously, is that given the assumption that Liars are dialethic, we could always construct a Liar sentence L that asserted of itself that it was super-false, $L=F!(L)$. Thus, the price of extending the dialetheist’s logic in the way just considered is triviality. Without thus extending it, however, the claim that “the moon is made of green cheese” even *could* entail everything in a dialethic logic falls flat.

So, to sum up the results of our discussion so far, (a) the suggestion that the dialetheist express the exclusive falsity of exclusively false claims by tying them to triviality by using conditionals fails unless the conditionals are truth –functional, (b) these *can’t* be the conditionals of the dialetheist’s logic, (c) the suggestion that the tie to triviality can be pulled off with entailment instead of conditionals only goes through if the dialetheist can express exclusive falsity in some other way (i.e. if they can do it with conditionals, or with a super-falseness operator), and (d) they can’t. All of that seems bad enough, but it gets worse.

The most serious problem with the idea that the dialetheist can express exclusive falsity by tying exclusively false statements to triviality, the one which absolutely squashes this suggestion, is that it only works if we *start* by being able to do precisely the

thing that we've been trying and failing to do this whole discussion, which is to provide the Liar-based dialetheist with a way of *expressing the claim that trivialism is wrong*.¹¹⁹ In other words, the suggestion that the dialetheist should use the implication of triviality to rule out the truth of certain sentences only works if the dialetheist has some way of ruling out triviality to begin with, and they can only do *that* if they have a way of saying that some false sentences aren't also true that expresses what that phrase intuitively seems to express. This is a circle from which there is no escape.

One of the chief attractions of dialetheism is its analysis of Liars. Indeed, for some authors, this seems to be the overwhelming attraction of dialetheism. For the dialetheist to so much as *express the thought* that their stance differs from trivialism, however, they must first abandon the claim that Liars are dialetheias.

II. Negation vs. Rejection?

Or so it seems. JC Beall has provided what, on first blush, looks like an intriguing way for the Liar-based dialetheist to have her cake and eat it too. He grants that it's technically impossible on his account to make sense of a distinction between the meanings of "true" and "just true," given that one can always construct a Liar L such that L says that $\neg L$ is just true. He suggests, however, that the turn of phrase "just true" can still express the information that he needs it to.

...we can, if need be, appeal to a pragmatic difference. In the context of transparent truth theories, rejection is not simply acceptance of negation. In dialethic theories, one may accept $\neg\alpha$ without rejecting α . (In non-dialethic theories, such as Field's, one may reject α without accepting $\neg\alpha$.) In this way,

¹¹⁹ One might be tempted to think the dialetheist can just say *that*, but that won't get it done. After all, there's always, "Someone who said that this sentence was true would be mistaken about its truth-value."

perhaps ‘just true’ carries pragmatic implicatures not carried by (an utterance of) ‘ttrue’, something to the effect that the speaker rejects the given negation. For example, if Max says that α is *just true*, he says nothing more or less than that α is ttrue; however, his assertion carries the implicature that Max rejects $\neg\alpha$. *Just true* and *ttrue*, on this account, remain equivalent; it’s just that an utterance of the former carries ‘autobiographical’ implications that the latter doesn’t.¹²⁰

This looks like a promising thought, given the problems we’ve been discussing.

Unsurprisingly, then, Beall puts rejection to extensive use as a sort of substitute for a way of saying of sentences that they’re just false. He says, for example, of Curry sentences like “if this sentence is true, the moon is made of green cheese” that, “On my account, Curry sentences are false. I reject that they’re ttrue.”¹²¹

He also uses rejection talk to differentiate his version of dialetheism from gap theory. He considers the objection that the “gluts” he postulates are really just gaps, because for any α such that he wants to say that it is both true and false ($\text{Tr}\langle\alpha\rangle \wedge \text{F}\langle\alpha\rangle$), given his logic and the intersubstitutivity rules governing the truth and falsity predicates, one can always infer that it is neither true nor false ($\neg\text{Tr}\langle\alpha\rangle \wedge \neg\text{F}\langle\alpha\rangle$).

The point is correct. Any glut is also a gap, *so understood*. Of course, traditional gap theorists, whatever else they might (or might not) say about ‘gaps’, are those who *reject* (what they want to call) ‘gaps’. The gaps in my theory are ones that I fully embrace...¹²²

Similarly, and more importantly for the point at hand, here’s Beall on trivialism:

Trivialists are dialetheists, but trivialism—unlike dialetheism—is not a rational view. Rational dialetheists maintain that some (actually, many) truths are *just true*; they reject that all or even most claims are gluts. Indeed, on my account it is only the spandrels of *ttruth* (or related notions) that are gluts...¹²³

¹²⁰ Beall (2009), pp. 51-52. On the issue of “ttrue” and “transparent truth theories” allegedly needing to be non-classical, see Appendix B.

¹²¹ Beall (2009), p. 33

¹²² Beall (2009), p. 105

¹²³ Beall (2009), p. 48

Note that the phrase “just true” appears in his passage, but, having previously admitted that he can’t have a notion of “just true” that doesn’t give rise to “spandrels,” and thus that doesn’t overlap with “false,” Beall is at pains to clarify that he’s using to signal rejection in just the “pragmatic implicature” way discussed above, hence the semi-colon. To belabor the point only a little bit, by Beall’s own lights (given his discussion of “just true”), with *just* the phrase “some (actually, many) *ttruth* are *just true*,” he would not have succeeded in expressing a difference between “rational dialetheists” and trivialists.¹²⁴

Now, this use of acceptance/rejection talk as a somewhat awkward stand-in for missing non-overlapping categories of “just true” and “just false” *only works* if one accepts that the categories of “acceptance” and “rejection” never overlap. This might seem plausible enough when we’re considering whether they overlap in the psychological state of any given agent at any given time. If one takes the truth of “an agent A doesn’t accept claim C at time T” as a conceptually necessary condition for the truth of “an agent A rejects claim C at time T,” then, given a materialist theory of minds, at least, simultaneous acceptance and rejection would constitute exactly the sort of contradiction in the concrete physical world that even most dialetheists would want to avoid postulating. Of course, there’s an awkward issue of how the dialetheist can express their stance on dialetheias-about-acceptance without *using rejection to do it*, which might seem

¹²⁴ Of course, *by the trivialist’s lights*, everything any non-trivialist ever says to differentiate their position from trivialism—whether the monaletheist saying that trivialism is “false” or the “rational dialetheist” saying that trivialism is “just false”—is *always* compatible with the truth of trivialism. That’s a trivialism problem that everyone has. A trivialism problem that only the dialetheist has is that, *even by the dialetheist’s own standards*, they haven’t succeeded in differentiating themselves from the trivialist when they say that trivialism is “just false.” Hence the acceptance/rejection talk.

to put the cart before the horse in this context. For the moment, however, let's charitably put that aside and assume that this much, at least is true.

The further problem is that what we've granted so far is grossly insufficient for the *purposes* to which acceptance/rejection talk is being put by the dialetheist who wants to use it as a stand-in for "just true"/"just false" talk, or for that matter by the paracomplete theorist who may wish to use talk of "rejecting" the claim that a paradoxical sentence is true, and "rejecting" the claim that it is false, as a stand-in for the revenge-paradox-generating claim that it is "neither true nor false." If Beall is just reporting his own psychological states when he talks about rejecting trivialism, rejecting the truth of Curry sentences, accepting the "gaps" generated from his gluts and so on, then what he's saying is philosophically uninteresting.

Remember his statement that all "rational" dialetheists reject that most claims are gluts. Imagine that only three dialetheists who Beall would consider "rational" exist in the world at any given time. (No doubt, historically, there have been times when this has been the case.) Now, assume that they all liked to drink tea. Moreover, assume that two self-described trivialists existed in the world, and they preferred coffee. If *all* that's going on in Beall's acceptance/rejection talk is psychological reporting, then "rational dialetheists reject that most claims are gluts, while trivialists take all claims to be gluts" and "rational dialetheists like to drink tea, while trivialists prefer coffee" would both be true in precisely the same sense, and both would be precisely as philosophically uninteresting. The point, surely, is that Beall takes their rejection of most possible claims-of-gluttiness to be a necessary condition for their rationality. What he's saying, in other words, is that rationally, dialetheists *should* reject that most statements are gluts.

Moreover, even if, charitably, one takes “most statements” to just refer to the majority of the statements that have ever been made by anyone (rather than say, the majority of a hypothetical limiting frequency of all possible statements that *could* ever be made), if the point is a matter of reporting on their psychological states, it’s wildly implausible that Beall’s “rational dialetheists” actually have *any* particular psychological state to most statements that have ever been made by anyone. Worse yet, when Beall talks about rejecting that Curry sentences are true, given that there’d be little enough point if he was just talking about whatever Curry sentences have been formulated up to this point in the history of arguments about paradox, the “psychological reporting” interpretation would have him claiming to have the attitude of rejection to every single member of the infinite set of possible Curry sentences, including a presumably infinite subset with consequents that Beall has never heard of.

Similarly, on the paracomplete side, Hartry Field talks of rejecting the claim that a given paradoxical sentence *S* is true, rejecting that it’s false, rejecting that it’s either true or false, rejecting that it’s neither true nor false, rejecting that it’s either true or false or neither true and false, and so on all the way up the whole transfinite chain of such claims. Not only does this not make sense as a claim about his current attitudes, it’s ludicrous even as a *dispositional* claim. After all, couldn’t Field be tricked or simply mistaken about a particularly complicated claim somewhere in this infinite chain whose relationship to the original paradoxical sentence wasn’t obvious? Moreover, what about the members of this chain too long and complicated for mere human minds to process? Not to mention, of course, that when Beall talks about things that are rejected or accepted

“on an account,” the psychological reporting interpretation becomes nonsensical. Accounts, presumably, aren’t the kinds of things with psychological states.

This particular dead horse wouldn’t be worth beating so severely if not for the fact that (a) instances of acceptance/rejection talk in the literature on paradoxes often read as somewhat ambiguous on this point, and (b) it is crucial to what follows. In any case, the point is this—the uses to which dialetheists (and paracomplete logicians) put acceptance/rejection talk *only* make sense if they aren’t talking about what the author of any given book or article or anyone else *does* reject, but rather what everyone rationally *should* reject.

Now, earlier we granted that acceptance and rejection never overlap. However, given the discussion of the last several paragraphs, we can see that for Beall’s move to work—for acceptance/rejection talk to do the work for which we intuitively want non-overlapping notions of “just true” and “just false”—we don’t *just* need acceptance and rejection to overlap. We also need the categories of “should accept” and “should reject” to be non-overlapping, and, I think, *these* non-overlapping categories are no more available to someone who takes Liar-type sentences to be meaningful, truth-evaluable and so on that are non-overlapping categories of “just true” and “just false.”

Everyone (whether one is a classical logician, a paracomplete logician or a dialetheist) thinks that true statements ought to be accepted. The dialetheist also wants to put statements that are *both true and false* in the “to be accepted” box, while putting the rest of the false statements in the “to be rejected” box. The classical and paracomplete logicians want to put all of the false statements in the “to be rejected” box, and the paracomplete logician wants to put some statements into the “to be rejected box” which

they wouldn't use either the word "true" or the word "false" to describe. Putting all that together, we can say that someone (say, an omniscient God) with perfect acceptance/rejection behavior would accept all true statements, reject everything else (and, for clarity, since we don't want to leave the door open for the possibility that perfect acceptance/rejection behavior would include "also rejecting some of the accepted statements" or anything of the kind) not act on any further rules about acceptance and rejection.¹²⁵ Given that, however, what do we do with the following sentence?

(47) This sentence would be rejected by a being who accepted all true sentences, rejected all sentences that they didn't accept and failed to act on any rules regarding acceptance and rejection beyond the ones listed in this sentence.

If one takes Liar-like sentences to be meaningful, truth-evaluable and so on, then it seems to me that one has no choice but to say that this sentence goes in *both* the "to be accepted" box *and* the "to be rejected" box.¹²⁶ After all, if (47) is true, it should be accepted *and* rejected, whether it's (just) true or both true and false. If it's (just) false, it should be rejected, which means that it's true and it should *also be accepted*. Thus, application of standard Liar reasoning to precisely the sort of sentence Beall would call an "inevitable spandrel" generates the result that rational acceptance and rational

¹²⁵ Depending on what one takes to go into "acting on rules," we might need to clarify that to "not acting on any further rules *not already contained in the rules just listed*." For example, the rule "accept every statement that follows from 'grass is green'" would be contained within the rule "accept all true statements."

¹²⁶ Similarly, it's not clear to me just what routes of escape are open to the paracomplete logician here. While standard Liar reasoning would not generate the dilemma for them—since their default move with regards paradoxical sentences is not to try out possible truth-values but to reject all the possibilities. On the other hand, whether they accept or reject (47)—in the interpretation of acceptance/rejection talk about paradoxes discussed above—an obligation to regard it as true (and hence to accept that acceptance and rejection overlap) still lurks in

rejection *sometimes overlap*.¹²⁷ It should be clear enough that for any attempt to construct a category of “should (just) be rejected” or “should be rejected and not accepted” or anything of the kind, a strengthened/“revenge” version of (47) can always be constructed. Thus, as a dialetheist with standard attitudes towards Liars, unless he wants to make an *ad hoc* exception for (47), when Beall says that dialetheists should rationally reject that all (or even most) claims are gluts, he’s giving us no information whatsoever about whether they *should also accept* that all or even most claims are gluts.

It is just not possible to simultaneously take a dialetheic attitude towards standard paradoxical sentences and express one’s overall position in a way that differentiates from trivialism. If anything, careful exploration of Beall’s “pragmatic implicature”/rejection gambit only digs the dialetheist deeper.

III. Curry

That said, it gets much, much worse.

Like the Tarskian who cannot express the thought that their hierarchy of artificial languages accomplishes the job for which it was designed, the dialetheist who can’t express their opposition to trivialism is in an awkward and embarrassing position. Still, not being able to express α is one thing, and being forced into embracing $\neg\alpha$ is quite

¹²⁷ Note that this is one Liar that’s potentially less troubling to a monaletheist than to a dialetheist. After all, it doesn’t generate a *contradiction*, but merely a special kind of normative dilemma. The problem is that it’s a kind of normative dilemma that destroys the last best hope of the dialetheist to express their position in a way that (by their own lights) differentiates it from trivialism. While I take (47) to be adequately solved by my account of Liars in general, given that it clearly fails to ground out in a subject other than truth, I don’t *need* non-overlapping categories of rational acceptance and rational rejection to express by my own standards my exclusion of statements from the category of truth. I can do that just fine with the predicate “false.”

another. As it turns out, a dialetheist who regards Liars as dialetheias is actually in the latter position here.

To start to get a handle on the issue of how the dialetheist does or doesn't succeed in having a principled reason to reject trivialism, we should take a step back and consider how *anyone* can have a principled reason to avoid trivialism. To begin with, a justification for rejecting *any* view had better not begin and end with stamping one's feet and insisting that it can't be right. When we beg the question against our opponents, we don't do anything to rationally justify our rejection of their views.

Now, one could argue that it's impossible to beg a question against a trivialist, because the trivialist accepts everything, necessarily including the wrongness of trivialism. This would, however, be entirely too hasty. The epistemic reasons that begging the question is barred from rational debate don't cease to apply in the strange case of our opponents cheerfully granting whatever is throw at them. Priest himself has some very sensible things to say about precisely this subject in *Doubt Truth To Be A Liar*, and when looking at how he fares on these grounds, it will be sufficient to compare his actual performance with the standards he himself lays out in that book. What he asks us to do in that book is to imagine, when arguing against the trivialist, that we are not trying to convince the trivialist (who, after all, is already convinced of everything) but that we are trying to convince a neutral arbitrator who has cautiously not yet made up her mind about the issue of trivialism, and who will accept or reject arguments and evidence brought up one way or the other, according to normal rules applicable to any inquiry about anything. Convincing the trivialist that trivialism is wrong is a trivial matter.

Convincing the imaginary arbitrator to rule in our favor is not. It is the latter, harder standard to which we should set ourselves.

Convincing the arbitrator is...not as easy as it might first appear... For example, an obvious argument to use is that the trivialist's position is inconsistent, and so ought not to be believed. (Naturally, a dialetheist is not going to make this objection; but a classical logician might.) It is clear that this argument fails, though, since it begs the question. The trivialist affirms...all contradictions. Simply to claim that these cannot be true is to take for granted part of what is at issue.¹²⁸

Things are not so hopeless as they may appear, however. In fact, there is excellent reason to be confident that a nontrivialist with normal beliefs about the world could make a lot of good, convincing anti-trivialist arguments of the kind that the hypothetical neutral arbitrator *would* allow.

The trivialist believes many strange things. They believe, for example, that you are a scrambled egg. This is objectionable since there is just no evidence to justify it. The rational person should not believe anything for which there is no reason. The arbitrator must agree. And if the trivialist argues that there is a reason, namely that it follows from trivialism, then it is they who now beg the question. The trivialist will also claim that there is independent evidence. (They will claim everything.) But claiming does not of course make it so. The evidence needs to be produced for the arbitrator....¹²⁹

This should be enough to get across Priest's general idea about how to argue against trivialism, and it is an idea that seems absolutely correct to me.¹³⁰ Moreover, the anti-trivialist argument just quoted seems to be a good one, and many similarly good arguments can be made. Before concluding that we have a good, principled reason to reject trivialism, however, we must not just consider the case *against* trivialism but the case *for* it as well.

¹²⁸ Priest (2006), p. 66

¹²⁹ Priest (2006), p. 66

¹³⁰ Once this framework is established, Priest goes on to introduce various complexities to the scrambled egg argument, and to make other arguments against trivialism. Whether or not all of his arguments go through is beside the point for now. The important part of the discussion from *Doubt Truth To Be A Liar* for our purposes here is not Priest's anti-trivialist argumentation, but his conceptual framework for how anti-trivialist argumentation should proceed.

Just as the best argument for dialetheism is the Liar Paradox, the best argument for trivialism is the Curry Paradox. Liars say of themselves that they are false, or untrue, or (just) untrue, or something else of the same general kind. Curry sentences, by contrast, say of themselves that *if* they are true, some α is true as well, where α can be any arbitrarily chosen sentence that you like. For example, take sentence (48).

(48) If the sentence marked as (48) is true, then Graham Priest is a scrambled egg.

Symbolize (48) as α . The relevant instance of the biconditional truth schema for (48) is “‘If the sentence marked as (48) is true, then Graham Priest is a scrambled egg’ iff if the sentence marked as (48) is true, then Graham Priest is a scrambled egg.” If we name the sentence α , and the claim that Graham Priest is a scrambled egg β , then we can symbolize this as:

$$1. \text{Tr}\langle\alpha\rangle \leftrightarrow (\text{Tr}\langle\alpha\rangle \rightarrow \beta)$$

Plugging in the definition of \leftrightarrow , we get:

$$2. [\text{Tr}\langle\alpha\rangle \rightarrow (\text{Tr}\langle\alpha\rangle \rightarrow \beta)] \ \& \ [(\text{Tr}\langle\alpha\rangle \rightarrow \beta) \rightarrow \text{Tr}\langle\alpha\rangle]$$

By Conjunction-Elimination, we get:

$$3. \text{Tr}\langle\alpha\rangle \rightarrow (\text{Tr}\langle\alpha\rangle \rightarrow \beta)$$

Now, from here, an easy conditional proof, using only Modus Ponens (MP), will get us to $\text{Tr}\langle\alpha\rangle \rightarrow \beta$.

4. $\text{Tr}\langle\alpha\rangle$	A for CP
5. $\text{Tr}\langle\alpha\rangle \rightarrow \beta$	3, 4, MP
6. β	4, 5, MP
7. $\text{Tr}\langle\alpha\rangle \rightarrow \beta$	4-6, CP

Now, applying Conjunction-Elimination once again to 2, we get:

8. $(\text{Tr}\langle\alpha\rangle \rightarrow \beta) \rightarrow \text{Tr}\langle\alpha\rangle$

From 7, 8 and Modus Ponens, we get:

9. $\text{Tr}\langle\alpha\rangle$

Now, from 7, 9 and Modus Ponens, we get:

10. β

Graham Priest is indeed a scrambled egg. In fact, since Curry sentences can be constructed with any arbitrary consequent that you care to pick, we can use this argument to prove *all* of the trivialist's claims without exception, or just cut to the chase by having the consequent be "trivialism is true."

This seems to be as good an argument for trivialism as the argument from the Liar Paradox is for dialetheism. In the latter case, in Chapter Four, we considered Gilbert Harman's attempt to block the derivation of inconsistency by a rejection of the biconditional truth schema and rejected it as question-begging. Priest's hypothetical neutral arbitrator would surely rule out a parallel rejection of the relevant schema instance in order to block the derivation of triviality for precisely the same reason. Similarly with *any* otherwise unmotivated tinkering with the formal machinery we used to derive triviality from Curry.

Now, the solution to Liar-type paradoxes advanced and defended in Chapters Six and Seven applies equally well to Curry. (Nor should this be surprising. Curry *is* a Liar-like paradox.) Truth talk can only be meaningfully applied to sentences that ground out in subjects other than truth, and Curry sentences do not. In a Curry sentence like "if this sentence is true, then the moon is made of green cheese", the "this sentence" attempts to

refer not just to meaningful “consequent,” but to the whole “conditional”, and like the meaningless statements with the grammatical form of disjunctions and meaningful “second disjuncts” embedded in them that we discussed in Chapter Seven, conditionals with meaningless antecedents are meaningless. “If blook nork geblork, then the moon is made of green cheese” does not say anything. No special problem is posed for the meaningfulness solution by conditionals.

By contrast, the Liar-based dialetheist is in a tough spot here. They are clearly committed to the claim that self-referential truth talk is meaningful, and, in fact, truth-evaluable. Curry sentences must, then, be meaningful declarative sentences, as would the relevant instances of the biconditional truth schema, and no one who admits this can have a good, principled, nonad hoc and non-question-begging reason to avoid granting that these schema instances give the correct truth conditions for Curry sentences. Once that point has been granted, all we need is Modus Ponens and conditionalization (the rule that we can use conditional proofs to prove the truth of conditionals). Rejecting Modus Ponens is extremely implausible. In fact, no one is more adamant about this point than Priest, who discusses this point in *In Contradiction*.

Any conditional worth its salt, \rightarrow , should satisfy the *modus ponens* principle... This is, indeed, analytically part of what implication is.¹³¹

Now, whatever one thinks of the analytic/synthetic distinction, or the complicated issue of the relationship between the *meaning* of logical constants and the rules governing their behavior that we looked at in Chapter Two, Priest’s basic point in this passage is a good one, which we can re-phrase in weaker terms that are neutral on those other debates by saying that if someone denies that $\alpha, \alpha \rightarrow \beta \vdash \beta$ is a truth-preserving inference, this

¹³¹ Priest (1987), p. 83

counts as pretty good evidence that they don't mean the same thing by \rightarrow that the rest of us do.¹³² We can extend (this charitably weakened version of) Priest's point by saying that if someone denied conditionalization, that would be equally good evidence that they didn't mean the same thing by \rightarrow as the rest of us. The two rules are conceptually parallel to each other, a fact recognized by introductory logic texts that introduce conditionalization as " \rightarrow -introduction" and Modus Ponens as " \rightarrow -elimination." In one case, the fact that α implies β is enough to guarantee that it entails it, and in the other case, the fact that α entails β is enough to guarantee that it implies it. This mutual relationship between entailment and implication seems obvious enough that no one thinks much of anything of talking about the "rule form" and the "conditional form" of various logical rules, and it's not at all clear to me why a denial of one half of this relationship should count as better evidence than the denial of the other for the claim that the denier doesn't mean what the rest of us mean by \rightarrow .

Denying either of these rules in order to escape the inference from Curry sentences seems like a desperate, adhoc maneuver. This is, however, precisely the route that Priest takes. In *In Contradiction*, he rejects "absorption" (also called "contraction")—the inference from $[\alpha \rightarrow (\alpha \rightarrow \beta)]$ to $(\alpha \rightarrow \beta)$ —on the grounds that "entailment must not fall foul of Curry paradoxes."¹³³ Absorption, however, just *is*

¹³² Notice that this is weaker than saying that it necessarily counts as *decisive* evidence in every situation, regardless of what other evidence we have available to us, and it's much weaker than saying that it's *impossible* for people who mean the same thing by \rightarrow to disagree about Modus Ponens. I would reject the latter claim, for the same sorts of reasons given in the negation discussion in Chapter one.

¹³³ Priest (1987), p. 83. Of course, this passage appears in the first edition of *In Contradiction*, published nineteen years before publication of *Doubt Truth To Be A Liar*, so one could suppose that in that earlier stage of the evolution of Priest's thinking, he had not yet decided that it was important to avoid begging the question against the trivialist, but that he later saw the light. The problem is that in the extra material and commentary on the second edition of *In Contradiction*, published the same year as *Doubt Truth To Be A Liar*, Priest gives no indication whatsoever of backing away from this view about absorption or from his

conditionalization plus Modus Ponens, so, given Priest's endorsement of Modus Ponens, his rejection of absorption boils down to a rejection of conditionalization.

He just barely gets to keep Modus Ponens itself (in its rule form), but his anti-Curry reasoning has the further unattractive consequence that he must reject the conditional form of Modus Ponens-- $[\alpha \wedge (\alpha \rightarrow \beta)] \rightarrow \beta$, sometimes called "assertion"—which, like absorption, is the product of accepting the validity of both Modus Ponens and conditionalization. Priest lays out a view about implication that avoid all of this, but the important point for the present discussion is that he justifies this theory by the need to bring about these counter-intuitive consequences, and that he justifies *that* with the need to get around Curry.

It must be admitted that the assertion principle *looks* acceptable enough, but...a criterion of adequacy for a solution to the problem of formulating an account of implication is that it must not validate assertion (or absorption).¹³⁴

Priest, in other words, is tinkering with otherwise intuitively plausible rules of inference in order to avoid being committed to triviality. If this is the best he can do when the trivialist brings her argument from the Curry Paradox to the neutral arbitrator, then the arbitrator will have no choice but to rule out this objection on the grounds of flagrant question-begging and, ultimately, rule in favor of the trivialist against the nontrivialist dialetheist. Given that Priest needs to accept that ungrounded truth talk is meaningful as a prerequisite for his claim that Liars are truth-evaluable (and, indeed, true, and also false), it looks like this is in fact the best he can do.

earlier justification of it. In fact, in *Doubt Truth To Be A Liar*, Priest refers the reader to *In Contradiction* for arguments against absorption.

¹³⁴ Priest (1987), p. 84. This passage comes at the end of Priest's discussion of Curry, which he treats as a reductio argument against assertion and absorption.

As such, not only does the dialetheist have no way to express their opposition to trivialism if they are unwilling to grant that Liars are not dialetheias, but they have no principled, non-question-begging, non-*ad hoc* way of avoiding trivialism. As such, a dialetheist who does not want to grant that *all* contradictions are true had better admit that the category of true contradictions does not include any contradictions about the alethic status of Liar sentences.

IV. The Epistemic Consequences Of Dialetheism

It's important for Priest's project that he not only reject (4), but that he also reject (3) as well, and as we will see below, even the vital concession that Liars are not dialethic will not be sufficient to get the job done. To get a grip on why it's important that Priest have a reason to reject (3), recall from the discussion in Chapter Three that (a) it can be shown quite easily that Disjunctive Syllogism (DS), Reductio Ad Absurdum (RAA) and Modus Tollens (MT) are invalid if dialetheism is correct, and (b) it would be patently absurd to claim that we are *being irrational* when we reason according to these rules. From a mathematics professor consciously reasoning according to RAA to prove to his class that there is no largest prime number to dogs instinctively reasoning according to DS when they follow a scent down a forking path (and, failing to pick up the scent on the first fork, immediately start scampering along the second fork without bothering to check it first), we all employ some or all of these rules on an extremely routine basis. Nothing seems more obvious about reasoning than that reasoning according to these rules

is rational, and, if dialetheism were incompatible with that claim, that fact would constitute an intuitively powerful reason to reject dialetheism.¹³⁵

Priest attempts to sidestep this problem by claiming that the statistical frequency of true contradictions is very low, so we should assign a very low epistemic probability to any particular contradiction which we do not have independent reason to believe is true.¹³⁶ This in turn is supposed to allow us to “re-capture” those rules of classical logic, like DS, MT and RAA, that are not universally truth-preserving if dialetheism is true. Even though they are, strictly speaking, invalid, they are “quasi-valid,” meaning that, all else being equal, we are probabilistically justified in using them.¹³⁷

Priest has two arguments for his claim.

One is that we use rules like DS all the time without getting into trouble, and (as we have seen) they would fail when applied to claims that are both true and false. Now, we do use it all the time, because we take it to be universally truth-preserving, but Priest is being far too quick in his reasoning here. After all, in situations, both theoretical and garden-variety, where we think we have a good reason to believe $(\alpha \vee \beta)$ and a good

¹³⁵ Readers who start out feeling radically unsympathetic to dialetheism might ask whether this consequence is more counterintuitive than the bare claim that there are true contradictions. My response would be that it doesn't matter whether or not this is true. The fact that some view V sounds counter-intuitive when we first consider it, take as an argument to convince those who have already convinced themselves of V and provided us with all sorts of arguments for V, doesn't carry much dialectical weight. On the other hand, a demonstration that V has a *further* counter-intuitive consequence C does carry a good bit of weight, all the more so if the partisans of V start out by agreeing that C is intuitively unacceptable, and put enormous argumentative effort into trying to show that V can escape C.

¹³⁶ This argument is developed in Priest (2006a), pp. 115-117. Now, I would argue that, for reasons we saw in section (I), above, for Priest to even be able to express the claim that the rate of true contradictions is very low—i.e. that the great majority of statements are true or false but don't combine the two statuses—he needs to abandon the claim that Liar-like sentences are both true and false.

¹³⁷ One might be concerned that, if we are wondering about the proportion of claims that are true and false, or of contradictions that are true, we will face problems about performing statistical calculations on categories with transfinite numbers of members. It's not clear to me, however, that this is what's going on. First of all, if one does not accept Platonism about propositions—the claim that claims that no one has ever made and will never be made still, in some sense, exist—then the problem goes away, and even if it doesn't, there may be a fix in terms of looking at the hypothetical limiting frequency of arbitrarily selected members of that transfinite set of claims, or of contradictory claims, or whatever one takes the relevant category to be.

reason to believe $\neg\alpha$, we quite frequently end up discovering that β is not the case. The normal response to this is not that “Disjunctive Syllogism has gotten us into trouble,” but that our premises were wrong, even if we can’t figure out what in particular was wrong with them. The normal procedure is to conclude—even if, as often happens in cases of garden-variety reasoning, we never go back and come to a new view about the matter—that we were wrong in supposing that $(\alpha \vee \beta)$ was true, and that there must be some undiscovered third option which we had not considered. This is the normal procedure even if we have no idea what the third option might be. Alternately, if we end up later finding excellent evidence for α , we assume that we were being too hasty in inferring $\neg\alpha$. I believe that these are normal procedures, not because we have independent reasons in each case to think that the premises rather than the argument form is at fault, but because these are the habits of thought formed as a result of the universal assumption of monaletheism. If Priest’s arguments from the paradoxes of self-reference, motion, change and normative conflicts are sound, and we must reject monaletheism, then these habits will all have to be unlearned.

His second argument for the low frequency of true contradictions is the “head count” argument. In *In Contradiction*, Priest invites his readers to think of all the propositions they have encountered in the last day or two, and to count up how many could be “reasonably thought of as” dialethic.¹³⁸ In most cases, of course, especially if they haven’t spent most of the last day or two reading his books, the implied answer is “not very many.”

¹³⁸ Priest (2006a), p. 116

Epistemically, this procedure for establishing how many true contradictions there are (once we've granted for the sake of argument that there are *any*) is grossly insufficient. People who grew up in an intellectual environment informed on every level by the assumption that the truth must always be consistent must, if there are indeed true contradictions, be forced to see them one by one. If we are to accept the possibility of true contradictions, a quick unscientific survey of our *current* beliefs about consistency won't do. What we need is a rigorous re-examination of our current beliefs in light of the expanded range of logical possibilities we must now force ourselves to take seriously to see if any of them might *now* be reasonably seen as dialetheic.¹³⁹ As we'll see, the answer to *that* question, given the relevant assumption, is be "a great many of them."

After all, the normal human epistemic situation is one in which we tend to be confronted with incomplete, mixed and ambiguous evidence, not only in higher-order cases like conflicts between quantum and relativistic physics, but even (and especially!) in garden-variety cases where the threshold for epistemic justification is very low. For example, if Robin tells me that she saw Jane at the library at 4:20 PM on April 20th, 2009, then in the normal course of things, no one would accuse me of falling down in my epistemic duties if I believe on the basis of that eyewitness testimony that Jane was at the library. Similarly, if Kristen tells me that she saw Jane at a bar on South Beach at 4:20 PM on April 20th, 2008, then no one would accuse me of neglecting my epistemic duties

¹³⁹ In this context, the dialetheist who believes that, given the possibility of true contradictions, the number of them we have reason to believe in is very low, would be begging the question by invoking the need to avoid 'multiplying contradictions beyond necessity' as a consideration against believing in any particular contradiction. The 'information argument' sometimes given for this principle, according to which, if we take any given domain to be consistent, this allows us to use resources derived from classical logic to reason about it, and thus to infer a lot more information about it, does nothing to motivate any sort of epistemic principle. The information argument is a "would be nice" argument, an argument that good things would come to pass if P turned out to be true, but this sort of argument does nothing to give us a slightest good, rational reason to believe that P is actually true.

if I believed *that*. What if I receive both reports? In the normal course of things, since our habits of thought are informed by the universal pre-philosophical assumption of monaletheism, I assume that they cannot both be right, and reserve judgment, assuming (although I would not make any such assumption if I had only heard only one or the other of the two reports) that one or the other of them must be mistaken, or lying, or joking, or whatever. After all, on the assumption that it is physically impossible for Jane to be in two places at once, it follows from the claim that she was in the library that she was not in a bar on South Beach, and vice versa.

Why, however, should the dialetheist not take this case, or *any* case of mixed evidence in which evidence for each of two contradictory conjuncts would (in the absence of the evidence for the other) be considered sufficient for justified belief in that conjunct, as a case of evidence for the conjunction of the two contradictory beliefs?

The point can be expressed formally in a language with both modal and epistemic operators by saying that, even on an extremely weak reading of the justification operator J , the following epistemic principle seems enormously plausible:

$$[(J\alpha \wedge J\beta) \wedge \diamond(\alpha \wedge \beta)] \rightarrow J(\alpha \wedge \beta)$$

On a standard of justification as strong as may be normally assumed in the use of the justification operator, the inclusion of “ $\diamond(\alpha \wedge \beta)$ ” in the antecedent may seem redundant, but it’s worth noting that even on an extremely weak reading of J , this is an intuitively compelling principle. The reading I have in mind is something along the lines of “for any claim α , $J\alpha$ iff there is enough evidence for α that, provided that there is not at least equally good evidence for some claim β whose truth would be incompatible with

the truth of α , it is rational to believe α .” Given this reading, of course, $(J\alpha \wedge J\beta)$ is quite compatible with $\neg\Diamond(\alpha \wedge \beta)$.

By definition, to be a dialetheist is to believe that the fact that some β is the negation of some α is not enough to rule out the possibility of α and β being jointly true, and to believe that, in fact, some conjunctions of this kind are not only possibly true but actually true. It could be argued that particular contradictions are impossible on some grounds weaker than logical impossibility, but this would require some further argument. Presumably for any dialetheist who, like Priest, believes that moments of change are contradictory, the mere fact that a statement is about the physical world would not be sufficient to bar it from being a true contradiction. Of course, Priest does make the weaker claim that the “observable world” is free from contradictions, and one might think that in the multiply-located-Jane case, this gives a good reason to regard the relevant $(\alpha \wedge \neg\alpha)$ as a sort local impossibility.

There are two general problems with this claim. The first is that Priest’s argument for the consistency of the observable world seems to be, basically, that if there were any contradictions in the observable world, we would observe them, and we don’t, so there aren’t any.¹⁴⁰

When we perceive, we can see that something *is* the case. Can we also perceive that something is *not* the case? Some have thought not. We can perceive only that it is red. Any judgment to the effect that it is not green has to be added to what we see by inference. This, as we now see, is false. I can see directly that something is not green. Or consider another example: you enter a room; the whole room is visible from where you stand; there is no one there. You can see that Pierre is not in the room. No Pierre-shaped objects meet the eye....

¹⁴⁰ Interestingly enough, Priest’s fellow dialetheist David Ripley disagrees with the claim that we don’t (veridically) observe contradictions, and has produced experimental evidence for the claim that we do in fact observe some. We’ll come to *that* claim in the next chapter.

Next: if α and β are actual observable states of affairs, then so is their conjunction. We can see that something is a cat; we can see that the thing is black. Hence, we can see that it is a black cat...

To apply it: let α describe an observable state of affairs, that is happening right here and now. Then $\neg\alpha$ describes another observable state of affairs, that x is not happening here and now. But then the conjunction $\alpha \wedge \neg\alpha$ also describes an observable state of affairs. That is, if such a state of affairs were ever to be the case, it could be seen...

...But, apart from the odd visual illusion...our perceptions of the world are entirely consistent. Hence, the observable world is consistent.¹⁴¹

The first of several problems with this argument is that its form is Modus Tollens, and, as we've seen, it doesn't look like a dialetheist like Priest is entitled to regard Modus Tollens as truth-preserving. Consider arguments like the following:

1. If the sentence named Liar asserts that Liar is untrue, then Liar is true, and Liar is untrue, so it is not the case that the sentence named Liar asserts that that Liar is untrue.

2. If naïve set theory is correct, the Russell Set is a member of itself, but the Russell Set is not a member of itself, so it is not the case that naïve set theory is correct. If Object O is currently changing from having Property P to lacking it (i.e. the moment of change is happening right now), it has Property P right now, but it is not the case that it has Property P right now, so it is not the case that Object O is currently changing from having Property P to lacking it.

3. If all property-owners have a right to vote and no women have a right to vote, then Sally the property-owning woman has a right to vote, but she does not have a right to vote, so it's not the case that all property-owners have a right to vote and no women have the right to vote.

It would be extremely implausible for Priest to claim that any of these conditionals are false, given his arguments about all of these subjects, and in each case he

¹⁴¹ Priest (2006), pp. 62-63

pretty clearly thinks the antecedent is (just) true even though he regards the consequents as dialetheias. Modus Tollens is to implication what Reductio Ad Absurdum is to entailment.

So that's the first problem.¹⁴² Priest is arguing in a way that is invalid if he's right about dialetheism. The second problem is that the relevant counterfactual conditional (if there were true contradictions about the observable world, our observations wouldn't be so reliably consistent) is not obviously true. What if our perception of the state of affairs described by α always occludes our perception of the state of affairs described by $\neg\alpha$? Priest acknowledges the possibility of such a "consistency filter", but argues, reasonably enough, that if we started to take the mere possibility of a thing as real evidence for its actuality, "we would collapse into total skepticism."¹⁴³ There's a bare logical possibility that Graham Priest is a tentacled being from beyond the stars dreaming about being a human logician, but we're not well-advised to take the possibility as establishing the actuality.

Now, one could reasonably argue that our perceptive faculties having a consistency filter is a good lot more likely than Priest being a tentacled being from beyond the stars. The thought that our distant ancestors only needed to be able to perceive one or the other half of an inconsistent situation in order to find food and mates and avoid predators on the plains of Africa is certainly a plausible one. That said, Priest's basic

¹⁴² Of course, one could argue that, even if it can't have deductive force for a dialetheist like Priest, Modus Tollens is still probabilistically "quasi-valid," so Priest is entitled to reason according to it. The dialectical context gets a bit tricky here, since we're considering this argument *in the context of trying to decide whether Priest is rationally entitled to regard the monaletheism-based rules of Classical Logic as probabilistically quasi-valid*. To be entitled to *that* conclusion, Priest needs to show us that there aren't very many true contradictions, as part of the process of showing us *that*, he needs to establish that the observable world is consistent, and to show us (using this argument, at least) that the observable world is consistent, he needs to reason according to Modus Tollens.

¹⁴³ Priest (2006), p. 63

point is well-taken. Without any particular evidence to settle the question one way or the other, perhaps we should be agnostic about whether we have consistency filters.

Of course, for Priest's Modus Tollens argument to go through, however, we need to move from agnosticism to the positive conclusion that we don't have a consistency filter. Our perceptions *can* be inconsistent. Priest argues that perceptual illusions establish exactly this. A couple of cases should give the flavor of all of Priest's examples.

...if one consumes enough alcohol, then before one passes out, one's environment—the room—appears to “swim”, that is, move—usually, spin. If one focuses one's attention on a point in the room, it seems to be stationary. The rest of the visual field, however, appears to continue moving. But the internal spatial relations between the fixed part and the moving part do not seem to change...we have a perception of stationary motion.

...It is easy to construct a pair of glasses that have a red filter on one lens and a green filter on the other. (It is possible to buy glasses of this kind for viewing old-fashioned 3-D films.) If I put such glasses on, and allow perception to stabilize, everything appears a uniform brownish hue. But for a short space of time, until things stabilize, I have the very strange experience of seeing everything as red and green, although the red and the green seem, somehow, to be at different depths in the visual field.

It might be said that being red and green is not a contradiction. But it is: red and green are not complementary colours.¹⁴⁴

These examples, however, are far from sufficient to establish that we don't have a consistency filter. They are emphatically *not* the sort of contradictory observables suggested by Priest's use of Sartre's Pierre example. We directly, non-inferentially perceive that Pierre is not in the café, just as, if he were in the café, we would directly, non-inferentially perceive *that*. As such, so guess the implication, if Pierre were simultaneously in the café and not in the café, we would perceive that. We can call this kind of direct observation a “Pierre-level-observation” for short.

Now, if you perceived that the room was spinning around the fixed part *and* directly perceived that the room was not spinning around the fixed part, or perceived that

¹⁴⁴ Priest (2006), pp. 60-61

the internal spatial relations were changing *and* that they were not changing, *that* would a (non-veridical) Pierre-level-observation of a contradiction. Similarly, if Priest put on his 3-D glasses, and, before everything settled into the uniform brownish hue, it were entirely red and entirely green *at the same depth of his visual field*, that would be a (non-veridical) Pierre-level-observation of a contradiction, and it would establish that, if there were observable-level contradictions, we could have Pierre-level observations of them. All Priest has established is the much less interesting point that α can describe a situation of affairs in which x is the case and β can describe a situation in which y is the case, and one might *infer* that x being the case would only be possible if y is not the case. Presumably, no one claiming that we have a consistency filter would deny *that*.

Of course, the fact that Priest fails to establish the non-existence of the consistency filter is hardly enough to establish its existence, but given that the correct position on this point seems to be agnosticism, it looks like we shouldn't assent to Priest's claim that if there were observable-level contradictions, we would observe them. So that's the second problem with Priest's Modus Tollens argument.

The third problem piggy-backs on the second. Direct observation of the state of affairs described by a claim α about the observable world is hardly the only way to become justified in believing α . Priest acknowledges this, point out that perception is fallible and that if "it turned out, for example, that supposing grass in Australia to be red and green all over allowed us to explain and predict every fluctuation of the Australian dollar, but had no other untoward consequences, we would have strong evidence that our senses deceive us in this case..."¹⁴⁵

¹⁴⁵ Priest (2006), p. 62

This, however, is still entirely too restrictive. The total list of ways in which we can become evidentially justified in believing a claim about the observable world is not exhausted by direct first-person perception and the sort of high-level theoretical confirmation that Priest is playfully gesturing at in his Australian grass example. What about situations in which the epistemic bar for accepting α is normally and reasonably set very low? Once again, what about the Jane example?

Any dialetheist, rejecting the possibility that Jane could be at the library and, at the same time, in a bar on South Beach and hence not in the library, on the grounds that this concerns the “observable world” and the observable world is free from contradiction, would be putting the cart before the horse and assuming precisely what is at stake when we consider the relevant evidence from the testimony of Kristen and Robin. In order to determine whether the observable world is free from contradictions, we need to examine precisely the sorts of cases to which our principle $[(J\alpha \wedge J\beta) \wedge \Diamond(\alpha \wedge \beta)] \rightarrow J(\alpha \wedge \beta)$ would be applied. Denying $\Diamond(\alpha \wedge \beta)$ where α is “Jane is at the library”, β is “Jane is in South Beach” and α entails $\neg\beta$ because the consistency of the observable world would be to utterly and transparently beg the question against the skeptic about the classical recapture. Given the weak, minimalistic reading of the justification operator stipulated above, to deny either $J\alpha$ or $J\beta$ (or both) would be to suggest an ad hoc and unmotivated deviation from normal, applicable standards of evidence...or one motivated only by the (question-begging) assumption that even if there are true contradictions, there aren't very many. As such, it looks like the dialetheist has no good, principled way to block the conclusion $J(\alpha \wedge \beta)$.

Although “garden variety reasoning” like the Jane case forms the most fertile domain for the generation of contradictions whose truth we have no good way to avoid granting, given dialetheism and $[(J\alpha \wedge J\beta) \wedge \Diamond(\alpha \wedge \beta)] \rightarrow J(\alpha \wedge \beta)$, it’s certainly not the only one. Pairs of ultimately incompatible scientific theories are often both supported by various kinds of evidence (conflicts between relativistic and quantum results only being the most obvious example), mathematicians have good reasons for and against adopting various axioms, and both sides of major debates in all sorts of areas of philosophy are supported by intuitively compelling considerations.

Again: the normal human epistemic situation is that of being surrounded by incomplete, ambiguous and conflicting evidence, and, in any given domain, the normal project of rational inquiry is, to a great extent, about finding ways to decide *between* α and $\neg\alpha$ despite $(J\alpha \wedge J\neg\alpha)$. Once we grant $\Diamond(\alpha \wedge \neg\alpha)$, we have no good way to avoid inferring $(\alpha \wedge \neg\alpha)$ very frequently indeed, in every domain of inquiry that there is. Taken to its natural conclusion, dialetheism in logic means chaos in epistemology, and the classical re-capture fails.

Chapter Nine: Monaletheism and Negation

In Chapter Two, we considered and rejected the claim made by Quine and others that classical logicians mean something different by \neg than dialetheists and other paraconsistent logicians. This was important, because it establishes that, when considering the dialetheist critique of classical logic, the two sides aren't merely talking past one another. Despite this, it's obviously true that dialetheism embodies a *theory of negation* that all monaletheists by definition reject. The bone of contention between the two camps, after all, is the question of whether the truth of any statement can overlap with the truth of its negation.

In Chapters Three through Seven, we considered arguments for this theory of negation—from the “paradoxes of motion and change,” from inconsistent obligations, from naïve set theory and, most importantly, from the semantic paradoxes. In each case, we found that the arguments failed, and there are good, compelling, nonad hoc and non-question-begging ways to block the contradictions. Along the way discovered good reasons to reject another theory of negation, the claim that the *failure* of the truth of a statement can overlap with the *failure* of the truth of its negation. (Or, in other versions of truth-value gap theory, just that the failure of the truth of a statement can overlap with the failure of its falsity. Since, in consideration of these issues, we endorsed the theory of falsehood whereby the falsity of a statement *just is* the truth of its negation, this distinction can for the moment be put aside.) We rejected this both because of Graham Priest's intuitively compelling case that it's impossible for there to be no fact that makes

a statement α either true or false, because the fact that there's no fact that makes α true is sufficient to make it false, and because, despite the fact that paradox-avoidance is often billed as one of the biggest motivation for truth-value gap theory, when we came to the Strengthened Liar Paradox, we saw that applying gap theory *entails* dialetheism. Thus, we rejected monaletheism-with-gaps in favor of monaletheism-without-gaps.

Finally, in Chapter Eight, we went beyond blocking the various arguments in favor of dialetheism to giving positive arguments for monaletheism. Given the meaninglessness solution to the Liar Paradox endorsed in Chapters Seven and Eight, the monaletheist can block the inference to triviality from the Curry Paradox without begging the question against the trivialist, and the dialetheist cannot. Worse yet, despite the fact that the simplicity of its (bullet-biting) approach to Liar sentences is supposed to be one of the greatest virtues of dialetheism, we saw that a dialetheist who takes Liars to be dialetheias is incapable of even *expressing their rejection* of trivialism. Finally, we saw that the dialetheist claim that there are true contradictions (or even the weaker position that *true contradictions are logically possible*) wreaks havoc in epistemology, since accepting this claim entails accepting a great many contradictions about pretty much every domain of inquiry, and leaves the dialetheist with no good way to “re-capture” rules of classical logic that constitute a huge (and hugely intuitively correct) part of normal reasoning. As such, not only do we not have a good reason to accept dialetheism, we have excellent reason to reject it.

This finally brings us back to the subject we tackled in Chapter Two: negation. Having argued extensively against nonstandard claims about negation (that α and $\neg\alpha$ can both be true, and that they can both fail to be true), our final task will be to lay out a

positive theory of what negation *is* that captures the intuitive features of the concept and motivates our claim that truth is neither gappy nor glutty.

The analysis of negation I defend will be very simple and commonsensical. It is just this: *negation is failure*. To accept that $\neg\alpha$ is true is to be rationally committed to reject α and vice versa, because there is nothing to the truth conditions of any statement $\neg\alpha$ above and beyond the failure of α 's truth conditions to hold. Extending Priest's point that the fact that there is no fact that makes α true is sufficient to make α false (and, equivalently, to make $\neg\alpha$ true), my claim is that it is also *necessary* to make α false. $\neg\alpha$ is true iff there is no fact that makes α true, and that's the end of the story. Thus, I reject both "partially-defined predicates" and "over-defined predicates," because no matter how vague and muddled natural languages are, no matter how eccentric and complex the history of its development may be, it's still the case that *whatever* the conditions may be for some predicate P mapping onto some feature of the world, the conditions for $\neg P$ mapping onto the same bit of the world are simply that bit of the world's failure to meet the conditions for P doing so.

The immediate intuitive case for this view, and from the view to the exclusion of gaps and gluts, is straightforward. When explaining the distinction between necessary and contingent statements to introductory philosophy students, it's common to use examples like "it will rain tomorrow." This is a contingent statement that might end up being true and might end up being false depending on complicated meteorological factors that are sometimes difficult to predict, but "either it will rain tomorrow or it will not rain tomorrow" is necessarily true. Whatever happens, we can be confident that one or the other of these two disjuncts will be true. At this point in the discussion, every now and

then a dissatisfied student may raise her hand and ask, “What if it drizzles?” At this point, in the normal course of things, her instructor might say that, of course, it could be that *in one sense of “rains,”* the drizzle will count as rain, and in another sense, it won’t, but that’s not the middle option that the Law of the Excluded Middle is meant to exclude. The point is that, if we agree on exactly how many drops of rain it takes for “it’s raining” to be true, then there isn’t a third option besides exactly that many drops of rain falling from the sky or fewer drops than that (perhaps even no drops) falling.

The idea of both of these things would simultaneously fail to occur is intuitively absurd, in fact intuitively absurd in *exactly the same way* as the idea that there could *both* be exactly that many drops of rain and fail to be that many drops of rain. Considering this point, it’s tempting to say that the fact that so many analytic philosophers are gap theorists and so few are dialetheists, indeed that many philosophers who are friendly to gap theory consider dialetheism too ridiculous to be worth arguing against, has less to do with gaps being innately more plausible than gluts than with the culture of academic philosophy being such that philosophers regard some logical laws with more superstitious awe and reverence than others. In any case, where α is “it’s raining” in some sense of “raining,” $\neg\alpha$ is true iff there isn’t as much rain as we required for it to be raining in that sense, and having dispensed with arguments for an expanded range of logical possibility from odd cases like the Liar Paradox, it’s overwhelmingly intuitively clear that this view of negation entails that there either will or won’t be that many drops of rain (but not both).

I. Folk Dialetheism?

Interestingly enough, the “common sense” intuitive credentials of this simple negation-as-failure view have been challenged by recent experimental work. At least one study seems to show that non-philosophers are surprisingly willing to assent to explicit contradictions in certain cases. Before getting too deeply into the details of the study, it’s worth making a few preliminary things clear. To begin with, *of course*, the majority of non-philosophers *could* be rabidly committed to a false theory of negation, and there’s no way that any experimental result could *by itself* settle this or any other substantive philosophical debate. Nor, contrary to the straw man version of experimental philosophy popular in certain quarters, has even the most enthusiastic “x-phi” practitioner ever claimed otherwise.

This point does not, however, make the experimental results philosophically irrelevant. They could add up to at least a powerful bit of *prima facie* evidence for the claim that the monaletheist, gap-intolerant, negation-as-failure view actually represents a departure from pre-reflective common sense, and the sense in which negation-as-failure may seem intuitive to some philosophers is really a matter of their intuitions having been corrupted by their immersion in a tradition that assumes classical logic. Not overwhelming or indefeasible evidence, mind you, but not nothing either. At the very least, if we take the pre-reflective common sense of ordinary speakers of the language to embody a mistaken theory of negation, we should be able to formulate a convincing error

theory, a story about how it is that the folk came to be mistaken in such an important way and why it is that we aren't fooled by whatever misled them.

Fortunately, I don't think the existing results support even this limited conclusion about the disconnect between negation-as-failure and pre-reflective common sense, even in the limited way that any one, very early study can support any substantive claim about social psychology. To see why not, let's dig into the details.

The relevant results were presented by David Ripley in his paper *Contradictions at the borders*. The sorts of contradictions Ripley is interested in are contradictions involving borderline cases of the application of vague predicates. To test ordinary speaker's intuitions about these cases, Ripley designed an experiment involving slides of circles and squares.

Participants were 149 undergraduate students at the University of North Carolina. They saw a slide (projected onto a screen) with seven circle/square pairs on it, labeled 'Pair A' to 'Pair G'. In Pair A, at the very top of the slide, the circle was as far from the square as it could be, while in Pair G, at the very bottom of the slide, the circle was touching the square. In between, the remaining five pairs moved the circle bit-by-bit closer to the square....

Participants were randomly assigned to one of four conditions. In each condition, participants were asked to indicate their amount of agreement with a particular sentence as applied to each of the seven circle/square pairs. The four conditions involved four different sentences; each participant, then, saw only one sentence and rated it seven times, once for each pair. Ratings were on a scale from 1 to 7, with 1 labeled 'Disagree' and 7 labeled 'Agree'. The four sentences were:

Conjunction, Non-elided: The circle is near the square and it isn't near the square.

Conjunction, Elided: The circle both is and isn't near the square.

Disjunction, Non-elided: The circle neither is near the square nor isn't near the square.

Disjunction, Elided: The circle neither is nor isn't near the square.

...note that each of the sentences has the form of a contradiction. The conjunctions wear their contradictoriness on their faces, while the disjunctions are a bit disguised; but one application of a DeMorgan law reveals them to be contradictions as well.¹⁴⁶

¹⁴⁶ Ripley (2009), pp. 3-4

Now, one preliminary point to make about all of this is that the disjunctions actually have the form of *gaps* (gaps of an extreme sort, such that they are presented not just as exceptions to the Principle of Bivalence but to the Law of the Excluded Middle as well), and even granting the soundness of the gaps-entail-gluts argumentation endorsed in Chapter Five, Ripley is considerably over-stating his case by saying that the relevant disjunctions have the *form* of contradictions. After all, there are gappy logics in which the inference to contradictions from these disjunctions can be blocked. For example, in his essay *Transparent Disquotationalism*, JC Beall argues for a “speckled” logical framework, in which these are two types of negation—“choice” negation and “exclusion” negation—such that, on Beall’s account, Strengthened Liars formed with “choice” negation come out as gappy and those formed with “exclusion” negation come out as glutty.¹⁴⁷ A speckled logician could read the “neither” at the beginning of Ripley’s disjunctions as an instance of “choice” negation and the “isn’t” as an instance of “exclusion” negation, and thus have a good reason by their lights to reject the inference from these disjunctions to contradictions.

Less esoterically, Jamie Tappenden and Scott Soames, who are decidedly both *not* dialetheists, have both argued that natural languages contain “partially defined predicates,” where to call a predicate P “partially defined” means that the cases where P applies and the cases where $\neg P$ applies don’t jointly exhaust all cases.¹⁴⁸ Edwin Mares who *is* a dialetheist and who also believes in “overdefined predicates” (an issue we’ll return to in the next section), but who agrees with Tappenden and Soames that some

¹⁴⁷ For more on this, see Beall (2006), pp. 7-22

¹⁴⁸ See Tappenden (1999) and Soames (1999).

predicates are neither normally defined nor overdefined but rather partially defined, gives a nice real life example.

This has happened to me several times. I am playing a board game with some friends. At one point, one of the players finds himself in financial trouble. A second player takes pity on him and lends him some play-money. A third player complains about the actions of the second player saying ‘that’s not allowed’. The rules on the inside of the box-top are consulted, but no mention of lending money is made. An argument ensues and a fourth player is called upon to make a ruling. Until this ruling is made, there is no fact of the matter about whether the lending of money is allowed in the game. Hence, until this point in time, the predicate ‘is a legal move in the game’ is partially defined.¹⁴⁹

Surely, any gap theorist who regarded “near the square” as partially defined would agree with the claim that it was neither the case that the circle was near the square nor not near the square in the borderline cases, and *disagree* with the claim that it was both near the square and not near the square. Moreover, the rules of their gappy logic of choice would presumably not license anyone to DeMorgan away this distinction.

Of course, I’m not *endorsing* this line of thought. I don’t believe in partially defined predicates any more than I believe in overdefined predicates (where the conditions for P and \neg P applying sometimes overlap), I’ve already rejected the notion that different accounts of negation track different phenomena, and I disagree with Beall and agree with Ripley that gaps entail gluts. My point in bringing up Beall, Tappenden, Soames and Mare is just that there are philosophically sophisticated issues at work in the debate about whether this entailment is valid, and that we shouldn’t read any particular position on this debate into the survey answers of undergraduates.

That said, while this is an important issue that we’ll come back to, it should be noted that *even if* survey results Ripley interprets as agreement to contradictions are really “just” agreement to gaps, this by itself gives no aid and comfort to my position.

¹⁴⁹ Mares, p. 267

Gappy and glutty analyses of borderline cases of the application of vague predicates are equally incompatible with the simple negation-as-failure view, and while the differences between them are conceptually important, pre-philosophical “common sense” being friendly to one is no more or less of a problem for my view than it being friendly to the other. So, what were the results?

Notably, the pattern formed by the mean responses to each pair was a hump pattern... The highest mean occurred in response to Pair C; there the mean response was 4.1, slightly above the midpoint of the 1 to 7 scale. In other words, participants exhibit higher levels of agreement to these apparent contradictions when they are about borderline cases; they do not reject what appear to be borderline contradictions.¹⁵⁰

Ripley goes on to consider various ways of explaining away the data according to which either (a) the undergraduates in the study weren't really assenting to logical contradictions, or (b) they were, but they were mistaken. He considers a variety of explanations of type (a), but he's fairly unsympathetic to explanations of type (b), arguing that “it seems a bit odd to suppose that speakers are mistaken about what's near what, when they can see the relevant objects clearly, are deceived in no way about the distance between them, and are not under any time pressure to come to a judgment.”¹⁵¹

Now, one might object that Ripley is playing on an ambiguity between senses in which it really would be odd to think ordinary observers were mistaken about nearness under those conditions (e.g. which circle-square pairs were closer together and which were further apart from each other, approximately how many inches were in between a given circle and a given square and so on), and a sense in which the relevance of these factors is far less clear (i.e. the conceptually complicated question of *how to most accurately describe* the spatial relationship). Put that to one side. The most striking thing

¹⁵⁰ Ripley (2009), p. 6

¹⁵¹ Ripley (2009), p. 12

that's wrong Ripley's gloss on this issue is that, while the *mean* response to the borderline Pair C was 4.1, presumably the responses even to that pair varied enough for quite a few to be of the sort that we'd have to classify as *rejecting* the contradiction. Even if most responses were above the midpoint, and we were willing to put all type (a) reservations aside and describe all such responses as "assenting to a contradiction" and the monaletheist has to explain how it is that most participants in the study came to be mistaken about what's near what, despite their clear view, lack of time constraints, etc., a dialetheist who shares Ripley's interpretation faces an *exactly parallel* task of explaining how it is that those of the participants whose responses to Pair C fell into the "rejection of the contradiction" category came to be mistaken about what's near what, despite their clear view, lack of time constraints and so on. If Ripley is right about (a), such that the undergraduates participating had an excellent sense of what the logical upshot of their answers were, and the upshot is just what Ripley thought it was (which, remembering our discussion of the relationship between gaps and gluts above, is far from clear even if we put away all other reservations about methodology and interpretation), both sides are obliged to provide error theories.

It might not, however, come to that. What about type (a) objections? Ripley considers indexical contextualism, nonindexical contextualism and noncompositional theories. The indexical contextualist says that different uses of vague properties like "near" express subtly different properties, so the undergraduates in the survey were agreeing that the circle and the square had one relational property and lacked another when they agreed that it was near the square and not near the square. Ripley takes this suggestion to be scotched by the fact that participants were approximately equally willing

to assent to the “elided” conjunction “the circle both is and isn’t near the square” as they were to the non-elided “the circle is near the square and not near the square.” The word “near” can refer to two different properties in the latter sentence, but since it only appears once in the former, this isn’t an option there.¹⁵²

The nonindexical contextualist allows that “near” refers to the same property, but thinks that the surveyed undergraduates were giving that property subtly different extensions when they said it applied and when they said that it didn’t apply. Although this is not Ripley’s favored explanation (Ripley believes that the undergraduates are in fact assenting to outright logical contradictions), he allows that something like this might work. He does, however, impose pretty strict conditions on what it would take for it to work, conditions that are spelled out in a way that seems designed to leave the reader with the impression that filling out the nonindexical contextualist analysis is a fairly tall order, and it involves really stretching the data to make it fit.

In order to avoid contradiction, then, she must say that one property has two different extensions with regard to two different contexts. Importantly, those contexts must both be at play in the interpretation of the single sentence. If the context is ephemeral, dependent on, say, a transitory mental state of the judge... then this should be possible. On the other hand, if the context is coarser grained, dependent on only things like world, approximate time, location, speaker and the like, then we can see that context could not have changed mid-sentence, and so a contextualist explanation couldn’t get off the ground... For a nonindexical contextualist explanation to work, the context relevant to determining the extension of ‘near the square’ cannot be the context in which ‘near the square’ is read by the participant. After all, there is only one such context, but the contextualist appeals crucially to a change in context between two extension-determinations. I see two options for the nonindexical contextualist: (1) it may be that participants process the sentence into some form that includes two occurrences of ‘near the square’ or something (conceptual material, presumably) corresponding to ‘near the square’—then each separate occurrence can be affected by the context in which it occurs—or (2) it may be that participants evaluate the conjuncts one at a time, retaining only the truth-value of each conjunct after its

¹⁵² Ripley (2009), p. 8

evaluation—then each evaluation can be affected by the context in which it occurs.¹⁵³

So, summing up the discussion of contextualism, Ripley argues that participants can't be agreeing to the non-elided conjunction just because they take "near" to refer to two things in its two appearances of the sentence, because there's no statistically significant difference between participants' agreement with the non-elided conjunction and their agreement with the elided conjunction in which the word "near" only appears once in the sentence. It's just barely possible that they could be assigning it different extensions when they assent to the claim that it applies and the claim that it does not apply, but to get that off the ground in the elided case, you need to appeal to claim that the surveyed undergraduates are processing the sentence through the filter of baroquely complex psychological processes, and the claim that they *really are assenting to what they seem to be assenting to* looks like the simplest, cleanest explanation of the data.

A remaining type (a) objection is the noncompositional option, whereby the meanings of the contradictory-sounding sentences the surveyed undergraduates are assenting to aren't really a function of the meaning of "near", the extensions of "near" and "not near" and so on, but rather serve to idiomatically express something else entirely, perhaps simply "the circle is a borderline case of 'near the square.'" ¹⁵⁴ While he devotes a lot of time to grappling with contextualism, Ripley brushes off the noncompositional option in a couple of paragraphs.

The problem with such an account is that its difficult to see why such apparent contradictions would express borderline-case-ness. How would such an idiom get off the ground? Presumably because some other explanation canvassed here was at one time correct...then language learners, for whatever reason, might have mistaken their elders' compositional utterances for direct claims of borderline-

¹⁵³ Ripley (2009), pp. 9-10

¹⁵⁴ Ripley (2009), p. 9

case-ness. This fills in the story, but it does so compositionally...and it seems to me that it will then take on the pros and cons of whatever compositional theory it chooses.

There will be a few extra cons, however. A noncompositional theory must explain why there is no significant difference in the frequency of observed hump responses between the four experimental conditions, and why there is no significant difference between the maximum responses given by participants in these conditions. Do we have four closely-related idioms? If so, why? In addition, this strategy invokes an additional step: learners coming to acquire noncompositional uses of these once-compositionally-used expressions. Without further evidence, a noncompositional theory involves needless complication; better to stick with a compositional story.¹⁵⁵

So, now we have four options. Either (1) the surveyed undergraduates were really attributing some relational property to the square and the circle attributing the negation of some relational property to them, but they weren't the *same* property, (2) they were attributing one property with one extension to the circle and the square but attributing the negation of the same property with a different extension to them, (3) they were expressing something entirely different from attributing a relational property and the negation of that property (or even the negation of a second property) to the circle and the square, but using a misleading idiom, or (4) they were really attributing a relational property to the circle and the square, and attributing the negation of the same property with the same extension to them. (1) is ruled out because the participants were as willing to agree with the elided conjunction as the non-elided one, (2) might be squeezed into conformity with the evidence, but only if we introduce all sorts of strange complications into the story we tell about the surveyed undergraduates' conceptual processing procedures, and (3) is a non-starter, because the most likely story for how such an idiom arose would have to piggy-back on (1) or (2) and inherit their problems, in any case, it's unclear why it would arise, and the four different experimental conditions would have to

¹⁵⁵ Ripley (2009), pp. 9-10

represent four substantially different idioms that all somehow functioned to express the same thought. As such, (4) still stands as the best explanation.

I take this entire discussion to be entirely too literalistic about how to read the sentences and the likely meaning gleaned from them by the participants, and entirely too generous to logically untrained undergraduates' ability to capture their thoughts in a sufficiently fine-grained way, particularly with as few cues as they were given in this study, to capture the most plausible reading of the thought processes even of those participants who's responses *can* be accurately labeled as "agreement" with any of the four sentences, but before coming to that point, it's worth pausing to consider some extremely basic questions of methodology. Remember that participants were asked (for whichever one of the four conditions they'd been assigned to) to rank their response to each circle-square pair from 1 to 7, with 1 labeled "disagree" and 7 labeled "agree." Given that 2, 3, 4, 5 and 6 were unlabeled, there's a nontrivial question about how to the responses of those who bubbled in anything *except* for 1 and 7. If 1 and 7 had been labeled "Strongly Disagree" and "Strongly Agree," rather than simply "Agree" and "Disagree," then we could charitably interpret 2, 3, 4, 5 and 6 as something like "Somewhat Strongly disagree," "Disagree, But Just Barely," "No Idea One Way Or The Other," "Agree, But Just Barely," and "Somewhat Strongly Agree" respectively, but without actual labels on the middle five options, this would still be something of a stretch, and it might pay to be more cautious and less confident about how to interpret even *that* version of the study. This applies even more strongly to the study we actually have, with the only labeled options being given the binary-sounding labels "Agree" and "Disagree." Perhaps we can *very* tentatively think of the unlabelled 4 as exact

uncertainty, since it's exactly halfway in between the options labeled "Agree" and "Disagree," but that still leaves the interpretation of 2, 3, 5 and 6 completely up in the air.

This problem becomes even more severe when we note that the mean responses for *all seven circle-square pairs* were greater than 2, in fact, that there was not a single circle-square pair for which the mean response wasn't further above 2 than the mean response for Pair C was above 4. (This is an important point, because the mean response for Pair C is, on Ripley's interpretation, supposed to amount to willingness to at least some extent to agree to a contradiction, even though 4 is exact uncertainty.) The point is worth underlining: Pair A and Pair G were supposed to be the absolutely clear cases which no one would be tempted to think of as being at the border between "nearness" applying and not applying. In Pair G, the circle was actually *touching* the square, and in Pair A, the circle was "as far away from the square as can be projected on the screen."¹⁵⁶ The fact that the mean responses to the absolutely clear, unambiguous cases were past 2, on their way to 3, counts as fairly powerful *prima facie* evidence that the surveyed undergraduates were at least somewhat confused about how to interpret the unlabelled numbers.

One obvious way to interpret this last point is that they weren't really indicating their degree of acceptance or rejection of the sentences (in which case, presumably, most of them would have bubbled in 1 for at least Pairs A and G), but rather something more like their degree of confidence in their acceptance or rejection of them. If so, the fact that they were less confident about their rejection of contradictions (or gaps) when it came to borderline cases than they were about clear cases of the application of a predicate or its negations is fairly uninteresting. Presumably, even if David Lewis and W.V. Quine were

¹⁵⁶ Ripley (2009), pp. 3-4

taking the survey, their answers *about the confidence question* would be humped in the same way. One can firmly reject some idea I, and still admit that Case 1 is a better fit for I than Case 2, and thus while, on consideration, still rejecting I across the board, still be less confident about the matter when considering the case that provides more ammunition for the partisans of I than a case that provides less ammunition for them.

Moreover, the confidence gloss explains why the mean responses were above 2 even for the crystal-clear cases. The students could have reacted with confusion to the assertions of true contradictions (and truth-value gaps)—“what do you mean, it’s near and it’s not near? What a strange thing to say...”—and, being uncertain about why anyone would think that and where this was going, or even quite what the import of the claim was, and suspecting that they were missing something important, not been entirely confident about their rejection of the claims even in the clear cases. When it came to the least clear case, they had absolutely no idea what to make of it, and the mean response (right on the edge of absolute uncertainty, so close to it as to be statistically insignificant) reflected this.

This can be buttressed by the discussion earlier about the relationship between truth-value gaps and true-contradictions. I argued above that, while I agree with Ripley that the former entail the latter, there are nuanced and philosophically sophisticated issues at stake in the debate about whether this entailment goes through, and it’s a bit much to read logically and philosophically untrained undergraduates as being implicitly siding with either side of this debate in their answers, so it’s far too hasty to read the (relatively) gap-friendly responses as correspondingly (relatively) friendly to the glutty answers they could be DeMorganed into in logics like classical logic and Graham Priest’s logic LP. As

such, the fact that there was no particularly statistically significant difference between the responses to the disjunctions and the conjunctions stands as fairly powerful evidence *against* Ripley's interpretation of the results. A far more plausible and cautious reading (especially given that no undergraduate answered about more than one sentence, so they never got a chance to evaluate *between* the sentences) would be that the surveyed undergraduates took the (gappy) disjunctions and the (glutty) conjunctions, elided and non-elided, as complicated, ambiguous and strange descriptions of the spatial relationships between the circles and the squares, and, as that spatial relationship became less and less clear, they used their degree of assent to whatever complicated, strange description they'd been handed as a blunt instrument to express that lack of clarity. This is, note, a far cry from it being a literalistic matter of the students recognizing some particular idiom (or four of them) that expressed a particular, well-defined claim about the borderline-status of the nearness of the circle and the square. Nor does this explanation involve any kind of complicated and subtle context-shift in the middle of the sentence being evaluated that affected their evaluation of complicated, fine-grained conceptual matters like exactly what property the word "near" referred to or exactly what its extension was.

In ordinary language contexts, both gappy and glutty-sounding formulations are casually used to express ambiguity all the time. Consider some conversational examples of a banally familiar sort:

"Did you ask Suzy if she wanted to go out with Johnny?" "Yeah." "And?" "Well, she didn't say 'yes', but she didn't exactly not say 'yes' either, if you know what I mean."

“Did he bring it up again after I left?” “Well, he did and he didn’t. What he said was...”

To illustrate that the point being invoked here applies more generally than the specific context of formulations that, taken literally, sound like they commit the speakers to truth-value gaps or true contradictions, consider examples like:

“Are you excited about tomorrow?” “Well, no, but see, I’m not unexcited either. It’s weird. I feel more like...”

Like the first two, the third snatch of conversation would be one that no one would bat an eye at over-hearing. It’s a fairly boring example of a completely, banally normal way of using language imprecisely to express complicated thoughts. If a logic professor with orthodox views overheard the third snatch of conversation between two janitors on his way into his office, he presumably wouldn’t think, “Good lord, the folk don’t regard the Law of Double Negation as truth-preserving! I guess I need to formulate an error theory about how ordinary speakers of the language come to be mistaken about Double Negation...” Nor should he.¹⁵⁷

An alternate version of Ripley’s survey that might more accurately screen for these normal, unremarkable uses of logically misleading language as an indicator of complexity, ambiguity and confusion might, instead of assigning each undergraduate one sentence and having them rate their agreement or disagreement from “Agree” to “Disagree” through an ambiguous maze of unlabelled options of unclear import, instead

¹⁵⁷ Simon Blackburn sums up this line of thought very nicely in his book *Think*. “In the short story ‘The Lady With The Pet Dog’ by Chekhov, Anna Sergeevna tells her husband that she is going to Moscow every so often to visit a doctor, ‘and her husband believed her and did not believe her.’ Formal logic does not tell us to jump up and down on Chekhov for this blatant contradiction. We know that Chekhov is suggesting something else.” (Blackburn (1999), p. 210)

ask them to pick which sentence out of a menu *best describes* their attitude to the spatial relationship between the circle and the square for each pair:

A-I think circle is near the square.

B-I think circle is not near the square.

C-I think the circle is near the square in one sense of “near the square” and not near the square in another sense of “near the square.”

D-I think the circle is near the square and not near it, in exactly the same sense of “near the square.”

E-I think the circle is neither near the square nor not near it, in exactly the same sense of “near the square.”

F-I have no idea what to say about whether the circle is near the square or not.

It’s too hard to tell.

Although there’s no way to be sure in advance of the evidence, my very confident prediction would be that, while C- *might* get more votes when it came to Pair C than when it came to, say, Pair G (the pair where the circle and square were touching), I’m quite sure that it would not be the most popular option of the six for *any* of the seven circle-square combinations. At the very least, until a better study, with more carefully disambiguated options, has been done, we have absolutely no evidence *against* the claim that the simple, clean negation-as-failure view accords with pre-philosophical common sense.

Moreover, *even if* we accept Ripley’s interpretation of his results (which we shouldn’t), and *even if* the project of interpreting the responses as anything other than assent to contradictions is as complex and difficult as Ripley takes it to be (which it is),

we're still left with another problem. As Ripley acknowledges, even on his interpretation, a dialetheist-about-borderline-cases-of-vague-predicates like him has to provide an account of why the mean response even to Pair C was as low as it was. Remember, 4 is the midpoint between "Agree" and "Disagree," probably best interpreted as absolute uncertainty, and the mean for Pair C was 4.1, so close to 4 that it's fairly statistically implausible to make a big fuss about the difference between the two numbers.

Ripley mentions some studies that seem to indicate that people in Asian cultures are more willing to agree to contradictions than Westerners. (Before we go any further, it's important to note that there's a world of difference between saying that Asians are *more* willing, *comparatively*, to assent to contradictions than Westerners, and saying that most or even a significant minority of Asians actually believe some contradictions to be true, or unambiguously agree to certain contradictions, or anything of the kind. The latter claim is one that, to the best of my knowledge, no one has ever explicitly made or presented any evidence for.) Following the authors of one such study, Ripley speculates that Westerners may hold a "cultural norm against agreeing to contradictions."¹⁵⁸

Suppose this to be true. Then, despite their linguistic competence pushing them to accept the borderline contradictions, subjects in the present experiment... may well have had their assent reduced by cultural norms. The effect would be much the same if we were to ask participants for their grammatical (rather than semantic) intuitions about sentences like 'Which table did you leave the book on?'; although ending a sentence with a preposition is perfectly grammatical in English, the cultural norm against it may well drive participants to reduce their judgments of grammaticality.¹⁵⁹

The problem with this, of course, is that the study of borderline contradictions that we *have* is about Westerners, and (even interpreted the way that Ripley does) it shows, at best, extreme uncertainty about such contradictions. The fact that it *might* turn out to be

¹⁵⁸ Ripley (2009), p. 15

¹⁵⁹ Ripley (2009), p. 15

the case that a study in another culture would show something different does nothing to add to the evidential weight of the existing study.¹⁶⁰ New studies could always change this picture, but at present, given the preceding discussion about the methodological flaws of the study, and the prevalent use of contradictory-sounding language as a way of gesturing at larger, more complex or ambiguous thoughts in ordinary contexts, the claim that ordinary speakers feel any sort of pull to actually agreeing with outright logical contradictions, interpreted as such, is far from the most natural interpretation of the results.

Now, rigorous empirical research always trumps anecdotal impressions, but right now, given that Ripley's study doesn't seem to show us much of anything one way or the other about non-philosophers intuitions about the truth of contradictions, anecdotal impressions are all we have, and even if they hold less weight than rigorous empirical research, it would be over-stating the point to say that anecdotal impressions, especially strong and confident ones, hold no weight whatsoever. On that score, at least, until these impressions can be corrected by evidence, it looks like pre-philosophical "common sense" and the simple, clean negation-as-failure account are a very good fit indeed.

II. Semantic Dialetheism

Of course, like any other view, negation-as-failure could be popular, intuitively compelling and ultimately wrong. We've already taken a quick look at how the

¹⁶⁰ Of course, if, say, it turned out that the mean result for Pair C in a study of Chinese undergraduates was 6.9, that would retroactively lend some credence to Ripley's speculation about linguistic competence and cultural norms. The question, however, is about the *current* state of evidence for the claim that ordinary speakers are inclined to believe borderline contradictions, and the current state of it is not good.

Tappenden/Soames/Mare “partially defined predicate” view as one challenge to it. As a dialetheist, Mare believes not only in gaps arising from “partially defined predicates,” but from gluts arising from “overdefined predicates.”

What is going on in cases of partially defined predicates is that the linguistic community somehow specifies a set of cases in which the predicates can be successfully applied and a set of cases in which its negation can be successfully applied. In the law and in rules of games, this specification takes the form of a sort of stipulation. But the form of the stipulation is not straightforward. What are stipulated are almost always sets of conditions that should be met, not a list of individual cases. But these conditions might not be exclusive. We could imagine, for example, a legal code being so complicated as to include overdefined predicates.

The legal case (and the case of rules giving rise to inconsistent obligations and permissions in general) is one we’ve already met and addressed, in section (II) of Chapter Three, but this illustrates the general point.¹⁶¹ Mares advocates “semantic dialetheism,” a view that differs from the “metaphysical dialetheism” advocated by Priest in that it is not committed to the claim that the most accurate total description of the word is one that contains contradictions.¹⁶² Rather, the semantic dialetheist thinks that the way the normal

¹⁶¹ The alleged “partially defined predicate” discussed earlier, about the permissibility of lending play-money to other players, falls rather neatly into the solution offered for over-defined rules before. If we take “genuine permissions” to be defined by practice (i.e. whether the lender gets away with it, granted that people know what he’s done, etc.), and the “according to the rules printed on the box top” permission to be something else, and we can indicate that we’re talking about the latter with an “according to the rules” operator R , then of course, Pa or $\neg Pa$ will always be the case, but $R(P\alpha) \vee R(\neg Pa)$ can easily fail without violence to the Law of the Excluded Middle. Anyone who’s ever spent a significant amount of time playing pool in bars will immediately recognize the justice of this distinction. The “Official Rules” for 8 Ball put out by the Billiards Congress of America stipulate a very specific arrangement for the balls when they are racked before the first break, but in most settings where pool is played in America, these matters are decided by a quick, drunken glance at the way the balls have been racked and a muttered a “yeah, that looks fine to me.”

¹⁶² I follow Mares in calling Priest’s position “metaphysical dialetheism,” but the issue isn’t actually quite as clear as all that. I’d argue that Priest’s own discussion of semantic dialetheism in the second edition of *In Contradiction*, in which he fails to take a definite position on it, confuses the issue by taking it to be primarily a matter of whether there are in some sense ‘negative facts’ that make true the negation of true statements made true by ‘positive facts,’ rather than the real nub of the issue, which is whether *any* conceptual tools we use to describe the relevant facts (even if none of them are ‘negative,’ whatever exactly that’s taken to mean) would, if they were to give us a complete, satisfying theory of the phenomena, necessarily be inconsistent. (See Priest (2006a), pp. 299-303, for his discussion of this.) Priest sometimes *talks* very much like a semantic dialetheist, as in his statement in *In Contradiction* that “natural language

concepts inherent in ordinary language map onto the world create dialetheias, but is open to the possibility that the same phenomena could be completely redescribed with consistent concepts.

Of course, it could be argued that this description of the distinction is incoherent, that if there's overlap between the cases in which English predicates apply and the cases where they don't, then even in a future language with all the muddledness of English sifted out of it—let's call it Futurese—there will still be true contradictions, at least assuming that we're allowed to use Futurese to *refer* to English sentence constructions. After all, even if one relativizes truth to languages, it will still be true-in-Futurese that some English-language predicate mapped on to some bit of the world and that it didn't do so.

There may be ways for the semantic dialetheist to get around this, but we can put that issue to one side and note that *even if* this objection goes through, there's still a way of cashing out the distinction between the two types of dialetheism. By analogy to the philosophy of mind, we can see monaletheism, semantic dialetheism and metaphysical

being what it is, we should not necessarily expect the pieces of the language to fit together neatly, like some multi-dimensional jigsaw puzzle. There may well be mis-matches. In particular, the conditions of application of a sentence may well mis-match those of the application of its negation, especially if the world arranges itself in an unkind fashion. At such spots in the weft and warp of language, we have dialetheias.” (Priest (2006a), p. 67) Moreover, his central candidates for the status of true contradictions in most of his work concern semantics and set theory, which are each in their own way arguably mechanisms for *talking about* and *categorizing* the world, and he has at least casually claimed (see Chapter Two) that the best evidence for dialetheism comes from inconsistencies about legal obligations, which Mares cites as an example of overdefinition. On the other hand, given the intimate connection between the case for partially-defined predicates and the case for overdefined predicates, this picture of Priest as a semantic dialetheist is awfully hard to square with his polemics against truth-value gap theory. Moreover, even if we stretch definitions to the point of somehow excluding sets and sentences from membership in “the world,” this gloss on Priest's position totally breaks down when we consider his analysis of change (see Section (II) of Chapter Two, above). There, pretty clearly, is a domain of unambiguously extra-linguistic reality with reference to which Priest is on record as thinking that consistent explanations are necessarily inaccurate. That said, there are too many textual ambiguities for me to be entire sure about the metaphysical vs. semantic classification, and I think it's entirely possible that at one point in his career, Priest was a semantic dialetheist, and he slowly evolved to his current position as a metaphysical dialetheist.

dialetheism as the counterparts to eliminative materialism, reductive materialism and dualism. The dualist claims that mental and physical states are quite different things, the reductive dualist claims that mental states are *real*—when we use terms like “pain,” “pleasure” and so on, we are talking about things that exist, so “I’m in pain” is a true statement and so on—but that our mental talk can (in principle) be reduced to physical talk, and the eliminative materialist (like, say, the Churchlands) takes mental talk to be as systematically in error as witchcraft talk. Similarly, the monaletheist takes intentional statements of assent to an explicit contradiction (unlike, say, the contradictory-sounding language used to express ambiguity, referenced in section (I) of this chapter) to be uniformly untrue. Meanwhile, the semantic dialetheist thinks some such statements *are true*, but that they can (in principle) be replaced with consistent descriptions of the same bits of extra-linguistic reality with no loss in accuracy. On the other hand, the metaphysical dialetheist thinks some such statements are true and *can’t*, even in principle, be paraphrased in any sort of complicated consistent way without loss of meaning, hence Priest’s frequently repeated claim that the consistency of theories of various domains can only be purchased with the loss of expressive power.

Of course, for obvious reasons, semantic dialetheism is in deep conflict with the view of negation I’m pushing here. Regardless of the issue of paraphrase with other, consistent predicates, if even *some* predicates are such that they and their negations both map onto the world, then negation is not failure. Mares tackles this issue head on.

Second Objection

Semantic dialetheism violates our intuition that ‘negation is failure’. The intuition behind classical negation is that a negation statement *not-A* is true if and only if *A* fails to be true.

Reply. This is true. But if one accepts that there are partially defined predicates, one must reject the view that negation is merely failure. Now, one might elaborate on the view and use a semantic theory, such as supervaluations, to make classical negation compatible with partially defined predicates. This is quite a coherent position, but if one admits that, at least in some cases, the meaning of predicates is given by a set of conditions under which the predicate is applied and a set of conditions under which the negation is to be applied, then there seems no reason why there cannot be possible cases in which those conditions overlap. Thus, I maintain that one who accepts partially defined predicates should reject classical negation and admit the possibility that there are overdefined predicates as well.

Moreover, following Tappenden (1996), I think that the classical semantics for negation conflates negation and denial. We deny statements that we think are not true. Some of these are not false either. For example, we deny statements because their presuppositions fail (e.g. I would deny that I have stopped beating my dog). The intuition that negation is merely failure, I think, comes from a tendency to make this conflation.¹⁶³

Mares' first argument here simply begs the question against the negation-as-failure view, since (although there are ways, as he indicates, to make them compatible with the rules of *classical logic* as a formal system), partially defined predicates are just as obviously incompatible with negation-as-failure as are overdefined predicates. The whole point of the partially defined predicates view is that the conditions for the negation of some predicates applying are at least somewhat *independent* of the conditions for the predicate itself applying, and the former are not (always) simply the failure of the latter.¹⁶⁴ Someone who objects to overdefined predicates on the grounds that they fly in the face of the negation-as-failure intuition would hardly be likely to be friendly to partially defined predicates in the first place.

His second argument is more interesting, although less clear. The idea that identifying the truth of $\neg\alpha$ with the failure of α to be true conflates negation and

¹⁶³ Mares, pp. 269-270

¹⁶⁴ Of course, a glut-intolerant gap theorist could postulate that the failure of the conditions for P mapping onto some bit of the world is a necessary but not sufficient condition for $\neg P$ mapping onto it, but this would still be incompatible with my claim, which is simply that negation *is* failure--for every statement α , $\neg\alpha$ is true iff α fails to be true, and that the failure of α to be true is all that's meant by " $\neg\alpha$ is true."

rejection seems wrong on its face. Denial is a doxastic state, a psychological attitude one takes towards a claim, while the truth of the negation of the claim is a property it has quite independently of anyone's or everyone's psychological attitudes towards it.¹⁶⁵ Moreover, even if one accepts partially-defined predicates and overdefined predicates into one's scheme of things, presumably *accepting something* requires thinking that it's true is true (even if one also think that it's false, as in, e.g. Graham Priest's acceptance of the claim that "the Russell Set is a member of itself"). Of course, one can accept something that *isn't* true, and presumably this happens all the time as the result of all the usual causes of human error and confusion, but *accepting something, and being right* is co-extensive with the thing one accepts being true, and no conflation between the semantic state of "being true" and the doxastic state of "being one of the things Person X accepts" seems to be going on in that description. Why not the same for denial? When one says that denying something *and being right to deny it* is co-extensive with the thing being denied being *not true*, no conflation is going on. Rather, we have a cleanly divided and intuitively appealing picture of how the doxastic notion of denial and the semantic notion of negation are related to each other.

Moreover, his example is a very strange one. Presumably, Mares never beat his dog. Given this, on a normal reading of the structure of the claim, "It is not the case that Mares stopped beating his dog" is true. Of course, this is conversationally misleading because it might be (hastily, and falsely) taken to mean "Mares once beat his dog, and it is not the case that he stopped beating his dog", which would entail that he currently beats his dog, but the extra conjunct appeared out of thin air. Of course, if we re-framed things to concern the misleading question, "Has Mares stopped beating his dog yet?",

¹⁶⁵ Unless of course, a claim is *about* those attitudes.

then it might seem like we have a case where denial can be (justifiably) applied that has nothing to do with negation, because questions can't be negated. The problem, of course, is that questions also can't be denied. Their *presuppositions* can be denied, though, and those are declaratives that can also be negated. Indeed, if someone asked Mares' the question in casual conversation, a likely way for him to express his denial would be by saying, "But I never beat my dog!" This statement, in turn, can be paraphrased without loss of meaning as "it is not the case that Mares once beat his dog," i.e. the negation of the presupposition. So, once again, without conflating a semantic concept with a doxastic one, we can see that denial and negation are indeed intimately connected. In this case, it seems flatly impossible for Mares' denial of the presupposition to make any sense (if, of course, he's *correctly* denying it) if the negation of it is not true.

Of course, all of this just gets us to the conclusion that Mares has not made his case against negation-as-failure. For obvious reasons, this doesn't add up to a case against Mares' view. To get a better handle on what we can say in objection to semantic dialetheism, above and beyond the question-begging objection that it conflicts with our view of negation, let's turn to Mares' supporting examples of semantic dialetheias.

It would seem that our notion of a shadow supports three principles.

1. If *X* casts any shadow, then there is some light falling directly on *X*.
2. *X cannot* cast a shadow through an opaque object.
3. Every part of a shadow is itself a shadow.

These three principles do not contradict each other in a straightforward way, but they can lead to our classifying things in an inconsistent way. Suppose that there is a barn with sunlight falling on one side of it. It casts a large shadow on the ground in the opposite direction. Suppose also that a bird flies beside the barn on the shaded side. Now consider the area of the ground that would be the bird's shadow if light were falling directly on the bird. Principle 3 forces us to claim that this area is in shadow, i.e. that there is shadow falling on it. But, by principle 1, it cannot be the shadow of the bird, since the bird itself is in shadow. And, by principle 2, it cannot be the shadow of the barn, since then it would have to be cast through an opaque object. Since the bird and the barn are our only

choices for ownership of the shadow, it would seem that we should also say that this is a nonshadow....¹⁶⁶

Now, an initial point about this is that it is not at all clear to me that “shadow” is a disguised description rather than something like a Kripkean rigid designator for a natural kind. Beyond that, it’s not clear how much of a problem is really posed *even for the shadow conditions listed*, by this particular case. Why, after all, couldn’t the bird and the barn jointly be casting the shadow? It could be objected that no light is falling on the bird, but my claim is not that the bird is casting the shadow and that the barn is casting the shadow, but that the bird-and-the-barn-taken-together are casting the shadow. It could be objected that the bird and the barn are distinct things, but this in itself clearly doesn’t matter. After all, if the bird were perched on the top of the barn, so that the shadow’s shape and length were impacted by the presence of the bird, that’s precisely how we’d be normally inclined to describe the situation. One, unified shadow, cast by the-bird-and-the-barn. One could press this objection and say that Principle 1 should be interpreted in an extremely strong way, so that light has to be falling on *all parts* of X for X to be casting the shadow. This, however, doesn’t capture Mares’ (or the common-sensical) claim that “the barn” is casting the shadow. Not all of the barn has light falling on it, and moreover, presumably, parts of the barn could disappear, and if they were thin enough parts, the shadow would be unchanged. So it’s not at all clear to me that one couldn’t hold onto all three conditions in the bird case without embracing any dialetheias about shadows.

More importantly, these conditions don’t actually seem to be definitional to the notion of being a shadow. They seem, rather, to be reasonably compelling observations

¹⁶⁶ Pp. 267-268

about shadows. For example, imagine a tribe whose religious belief system involved not just the notion that all objects had souls (as in various real life animist belief systems), but also the notion that most objects (perhaps excluding certain particularly holy objects, like the sun) have (or, perhaps, are followed around by) demonic anti-souls, and that shadows are caused by these “anti-souls” rather than by light directly falling on objects. Perhaps the tribe would believe that, in the dark, we can’t *see* the shadow a person casts, being occluded by all the darkness around it, but that it’s still there, hiding. Or, more fancifully, they might believe that anti-souls sleep at night just as people do, which is why they are never seen during the hours of darkness. One way or the other, members of this tribe would certainly disbelieve Mares’ Principle 1. It seems implausible, however, to say that this tribe’s belief system wasn’t *about shadows*, that they weren’t applying the predicate “being a shadow” (and asserting all sorts of false things about it).

As such, *even if* Mares is right that the three conditions listed can come into conflict with each other (which is far from clear to me), the correct response would be to say that (like the hypothetical tribe with supernatural beliefs about shadows) we have a *false theory of shadows*.

Of course, it may not be immediately clear quite what’s at stake when it comes to the question of how to revise our beliefs about shadows. Fortunately, Mares has supplied us with two more cases, both of which are about more obviously philosophically interesting issues.

... Consider the story of the ship of Theseus. In this story, a ship is rebuilt slowly over time replacing one plank at a time. The ship eventually has no original planks and another ship is constructed from the original planks. The question is ‘Which ship is the original ship?’. Our notion of identity supports the principle (a) that if two objects are made of exactly the same materials, they are identical. It also supports the principle (b) that if the material an object is made of changes

slowly over time, it remains the same thing. And it supports the principle (c) that if two material things are both complete and are wholly separated at a given time, they are distinct, and (d) that identity is transitive. Thus, by (a) the original ship is identical to the ship rebuilt from the original material. By (b), the ship built from the new material is identical to the original ship. Thus, by (d), the two rebuilt ships are identical to one another, but by (c) they are distinct. I claim that our notion of identity is overdefined. Both ships are the original ship. This raises an interesting problem, because the two end ships are not identical to each other, which in turn raises a standard question (luckily outside the scope of this chapter): should we reject the transitivity of identity? What is not beyond the scope of this chapter is the question of whether our standard means of determining whether two time slices are of the same object can lead to contradictions. I maintain that the Ship of Theseus case indicates that they can and that it seems likely that our concept of identity over time is overdetermined.

A second philosophical example comes from Kripke (1979), and his famous story of puzzling Pierre. Pierre is French and has not visited London. He comes to believe that London is pretty, but speaks no English. He is disposed to and often does say ‘Londres est jolie.’ Later he visits London, but is unaware that it is in fact the referent of ‘Londres.’ He now speaks English and is disposed to say ‘London is not pretty,’ since he has seen the worst side of the British capital. And he is not disposed to say ‘London is pretty,’ and would in fact deny it.... It would seem to me that belief reports are governed by the following principles:

1. If a person is disposed to sincerely utter a translation of the sentence S in any language that he or she understands, then we can say that he or she believes that S.
2. If a person is not disposed to utter a translation of the sentence S in any language that he or she understands or if he or she denies it, then we can say that he or she does not believe S.

In the Pierre example, applying these two principles leads us to the claim that Pierre believes that London is pretty and that Pierre does not believe that London is pretty.

Philosophers who have tried to get rid of the paradox about Pierre have, by and large, altered one or the other of these two principles. They would claim, I think, that they are attempting to discover the true principles that underlie our notion of belief. Another way of construing what they are doing is to say that they are devising alternative concepts to our overdefined concept of belief. Admittedly, it is difficult to choose between these two approaches. But the fact that these two principles are so intuitive does place the burden of proof on philosophers who claim that we do have a concept all the applications of which are consistent.¹⁶⁷

Let's start with the Ship of Theseus and the conditions (a)-(d) that are allegedly licensed by our notion of identity. Now, the denial of (d) *does* represent a genuinely radical metaphysical (and logical) position, although it has its defenders. It's interesting

¹⁶⁷ Mares, pp. 272-273

that Mares takes *that* to be the one that's open to dispute, because denial of (a), (b) or (c) is far less radical, and in fact none of those principles command anything like consensus agreement among major thinkers in contemporary metaphysics. Many philosophers claim to find it counter-intuitive that, for example, statues are identical to the clay they are made out of, so they would deny (a). The principles of classical extensional mereology, which have been defended by, for example, David Lewis, and according to which, for any objects A and B, there is a third object (A + B), flatly contradicts (c), at least if "distinct" means "they aren't part of the same larger object." Moreover, there's a lot to be said for those classical principle of unrestricted mereological composition, given its simplicity and elegance and the notorious difficulties involved in nailing down more restrictive answers to the question, "When do two things constitute parts of a larger thing that is wholly composed of them?", without leaving out all sorts of cases where it seems intuitively obvious that multiple objects *are* parts of a larger object wholly composed of them.

Of course, there may be a lot to be said *against* unrestricted mereological composition as well—standard objections involve the counterintuitive nature of "mereological monsters" like the four-dimensional object composed by the big toe on Julius Ceaser's right foot, the last bottle of Coca Cola that President Kennedy enjoyed before he was assassinated, the ring of Saturn and the pinky finger on my left hand—but that just serves to show that there are difficult, non-obvious issues at stake in the debate about mereological composition, with intuitively plausible considerations being brought to bear on both sides. It seems to be a bit much to say that *the very notion of identity* commits us to one or another position on this hotly contested issue.

If (a) and (c) are subjects of philosophical controversy, (b), as stated, looks flatly wrong. A pool of water that changes its material, very slowly, over time, from being filled mostly with water and a little bit of dirt, to gradually being filled up with more and more dirt and less and less water, until it is finally completely filled with dirt, is surely not “the same pool of water” at the end of the process. Rather, we’d most naturally want to say that “there used to be a pool of water there, but it’s gone now and been replaced by that patch of ground.”¹⁶⁸

Of course, the point I’ve been arguing for—the non-obviousness of (a)-(c)—doesn’t entail that there aren’t sticky philosophical problems about identity at stake in cases like the Ship of Theseus, or, say, Derek Parfit’s fission and fusion cases. There are. And the fact that at least three of his four principles are at least somewhat intuitively plausible, and each have many philosophical defenders—and even the considerably overstated (b) can perhaps be re-stated, with many qualifications and reservations, to become intuitively plausible, in accord with many extant metaphysical views, etc.—is relevant to the stickiness of these problems. However, since we’ve seen that each of the principles (a)-(c) is philosophically controversial for reasons quite independent of contradiction-avoidance, it seems far more reasonable to say that when some initially intuitively plausible methods for determining whether two time slices are of the same object come into conflict with each other, that should be part of a case for rejecting at least one of those principles, and that indeed this process can carry a valuable role in sharpening, clarifying and winnowing down originally vague and diffuse intuitions about identity problems.

¹⁶⁸ Of course, trivially, if one accepts classical extensional mereology, the spatio-temporal sum of the pool of water and the patch of ground existed intact throughout the whole process. This at best gets us to the point where, if we’re willing to deny (c), we can hold onto a slightly-beside-the-point version of (b).

Put more starkly, it looks like we've come to an important objection to semantic dialetheism about identity, that can be generalized into an overall objection to semantic dialetheism: When principles that initially seem intuitively plausible come into conflict with each other, if we simply embrace the contradiction instead of trying to resolve it, we're not only depriving ourselves of what initially seems like an important parameter of rejection (a particularly important one in an area as abstract and distant from empirical checkpoints as this one, where there are very few parameters of rejection for any view), but we're retarding conceptual progress by not taking the opportunity to go back and re-examine, re-consider and sharpen our original intuitive reactions to these principles.

This objection is roughly analogous to a standard objection to "Intelligent Design" (ID) creationism made by defenders of neo-Darwinian biological orthodoxy. The ID theorist, by saying that various things (e.g. in Michael Behe's case, biological cells) are "irreducibly complex," and that there is no way they could have arisen naturalistically from more basic components, is putting God into the gaps in his understanding and thus shutting down future research. If we settle for "God did it," this leads to an incurious attitude where we stop searching for further physical explanations (which are the only sort of explanations we're in any position to investigate), and we deprive ourselves of future knowledge, in a certain sense putting our own ignorance on a pedestal and worshipping it. Similarly, despite the considerable differences between the process of belief-revision in as directly empirical a field as biology and as airily abstract a field as the metaphysics of identity over time, there's a legitimate worry that embracing the contradiction when the principles that initially sounded intuitively appealing coming into conflict with each other can stunt future intellectual progress in the latter field in a way

that's at least somewhat analogous to the way that using God to paper over explanatory gaps can stunt intellectual progress in the former field.¹⁶⁹

This is an important point as part of an overall case against semantic dialetheism. We'll come back to the project of building that case shortly. Meanwhile, what about Pierre and the allegedly dialetheic state of his beliefs?

To begin with, I think similar points could be made about Mares' belief conditions to his shadow conditions. They both seem like initially intuitively plausible claims about beliefs, but it would surely be possible to deny them while still talking about beliefs, and this can be shown with a closely related example. Presumably, there have been at various times in history cultural groups so isolated that they haven't known about the existence of other cultures, with other languages. That said, members of such isolated culture understand each other to believe things, to disagree about their beliefs and so on, and it seems extremely implausible that they aren't talking about the same thing we are when they use (whatever the equivalent word in their language is to) the word "belief." (If one finds the historical speculation implausibly distracting, one could simply consider children in homogenous areas who can speak their own language fluently enough to know about "beliefs" but who aren't yet aware of the existence of other languages.) Now, it could be argued that, if and when this isolated culture discovered the existence of other

¹⁶⁹ In a sense, some semantic dialetheists might find even this objection question-begging. If embracing a contradiction is the right move, then we aren't depriving ourselves of future knowledge that would be gained by going back and re-examining our original intuitions, sharpening them in the process of choosing between them and making finer-grained conceptual distinctions and so on. Note, however, that the ID theorist could make *precisely* the same claim, that the question is being begged against them, because, if postulating divine agency to explain away gaps is the right move, we are depriving ourselves of no future knowledge that would be gained in more (fruitless) research. In both cases, the right reply is the same: ye shall know us by our fruits. Past cases in biology where phenomena have seemed irreducibly complex have been successfully reduced, and past cases in philosophy where intuitions have come into conflict have led to sharpened intuitions, finer-grained distinctions and promising new ideas. Moreover, it's surely at least epistemically virtuous in some way to keep trying to come up with new breakthroughs in the future, and it seems plausible that, all else being equal, the epistemic vice of discouraging people from trying is a serious point against a maneuver like the ones we've been discussing.

languages, they would assent to Principles 1 and 2, that the principles were all along in some slightly mysterious sense implicitly buried in whatever principles they held about belief, such that even though it pretty clearly couldn't be part of the consciously held principles of the linguistic community (because it involves a concept, "other languages," with which they are not familiar), it is still in some sense unconsciously part of them. This seems *ad hoc* and implausible to me, but in any case, larger objections are in the offing, so let's put it to one side.

People surely have inconsistent beliefs ($B\alpha \wedge B\neg\alpha$) all the time.¹⁷⁰ If Pierre just assented to "London is pretty" and "London is not pretty," that would be the situation, and no threat to monaletheism would be in the offing. The problem is that, given 1 and 2, we seem to have a case for the outright contradiction ($B\alpha \wedge \neg B\alpha$).

One initial way of objecting to this might be to say that this involves the sort of claim that even a (merely semantic) dialetheist like Mares should shy away from, which is that Pierre simultaneously has and lacks precisely the same psychological state (belief in α) at the same time.¹⁷¹ That sounds like the sort of deeply contradictory situation that couldn't be completely described with consistent successor concepts. Moreover, it seems to mis-describe the situation. Pierre's psychological state of not being disposed to assent to the English sentence "London is pretty" is *not* a state of not being disposed to assent to

¹⁷⁰ As I've indicated previously, I think it's plausible that no belief set is ever totally consistent, and that a reasonable way to think about the process of rational belief revision is as an attempt to make ones beliefs as consistent as possible, by eliminating inconsistencies when they are discovered... e.g. the inconsistency involved in thinking that some theory is true on the one hand and taking new experimental evidence that conflicts with it to be accurate. Of course, all this is quite consistent with saying that it's never rational to *knowingly* believe a contradiction, and even that the most rational kind of belief set would, among other things, be a totally consistent one, although the "among other things" is important, because the easiest way for a belief set to be completely internally consistent would be for it to be empty. In any case, total consistency can be important as a regulative ideal of maximal rationality without being totally achievable in practice.

¹⁷¹ Hartry Field, in another context, confidently asserts that he doesn't "believe any dialetheist would suggest that we might simultaneously believe and not believe the same thing." (Field 2008, p. 388)

the French sentence “Londres est jolie.” The former seems to be a state he *just* has and the latter seems to be a state he *just* lacks, and nothing in the description lends credence to the claim that, once we’ve disambiguated everything, he simultaneously has and lacks any state.

As stated, though, this objection doesn’t quite go through. Mares would be within his rights to accuse me of begging the question against him with this line of thought, by simply assuming negation-as-failure when I gloss $(B\alpha \wedge \neg B\alpha)$ as Pierre having and lacking precisely the same psychological state.¹⁷² For the semantic dialetheist, one aspect of Pierre’s psychological makeup can make true $B\alpha$ and a wholly separate aspect can make true $\neg B\alpha$, because the conditions for the truth of the negation of a statement can be quite independent from the conditions for the truth of the statement itself, and not simply the failure of them to hold.

Consideration of this point, however, leads us to a further argument against “overdefined predicate” theory, one that not only dovetails nicely with our first argument against it, above, but which quite naturally extends into an argument against “underdefined predicate” theory as well. Allowing for “overdefined predicates” (*or* “underdefined predicates”) robs our negation-talk of expressive power, sapping our resources for disambiguating between closely relating concepts. If I’m using the same formulation to describe two different aspects of Pierre’s psychological makeup (one of them the one that, for Mares, $B\alpha$ is about, and one of them the one that he takes $\neg B\alpha$ to be about), not just as a matter of initial sloppiness in getting my ideas across in a loose way, but on considered reflection when expressing things in what I take to be an accurate

¹⁷² To be more precise, if this gloss doesn’t quite assume negation-*as*-failure, it assumes the weaker view that failure is a necessary condition for negation.

way (even if not *the only possible* accurate way), then I'm failing to illuminate the distinction between them.

To put this in a slightly snarky way, imagine that I'm at the shipyard, the ship rebuilt from the original planks of the Ship of Theseus is sitting there, and some curious Greek comes up to me and asks, "Is that the Ship of Theseus?" I respond, "it is and it isn't." Thinking I'm using this contradictory-*sounding* formulation in the normal conversational way discussed in section (I) of this chapter, he waves his hand and says, "OK, what'd'ya mean by that?" What should I do?

A-Say, "no, I actually, literally mean that it is and it isn't. I'm not using expression that as a shorthand for some other idea."

...or...

B-Say, "well, if what you mean by 'being the Ship of Theseus' is that the planks in it are all the original planks from when the Ship of Theseus was made, then yeah, it is, but if what you mean by 'being the Ship of Theseus' is that it's the ship, made out of the new planks, that was the end result of all those planks being replaced one at a time, then no, it isn't."

Now, to let the example play out a bit beyond its immediate usefulness in illustrating my point, it is of course true that if I took option A, the option that semantic dialetheism would entitle me to, one of two things would happen. Either the curious Greek would either throw his hands up in frustration and wander off to find someone who would talk sense to me, or, if he were philosophically curious (as, after all, Greeks sometimes are), he would grill me about why I said such a strange thing to the point where, once he got my original arguments *for* the claim, in an indirect, circuitous sort of

way, as a side-product of this separate inquiry, he'd find out what he wanted to know about the ship. By taking the semantic dialetheist approach, however, I rob myself of my ability to express it in the simple, normal way, *using negation*, because I'm using the same formulation to express what is the case and (something else) which is not the case.

In general, if, as in “partially defined” cases, two *independent* sets of truth conditions and a situation are such that neither of the sets of truth conditions are met by the situation, and I express this fact by using one formulation, and refusing to assent to either it or its negation, or if, as in the “overdefined” case, two *independent* sets of truth conditions and a situation are such that both of them hold, and I make the same formulation do double duty and say that it holds and its negation holds, I'm missing an opportunity to *express more and clearer information about what's going on* by disambiguating, using two formulations and asserting the negations of both in the “partially defined” case, or using two formulations, assenting to one and assenting to the negation of the other. Graham Priest often complains that various solutions to the paradoxes purchase consistency at the expense of expressive power. I'd argue that semantic dialetheism purchases *inconsistency* at the expense of expressive power.¹⁷³ Given that the epistemic arguments against dialetheism advanced in Chapter Eight apply to at least some extent to semantic dialetheism as well as the metaphysical variety, and that semantic dialetheism is epistemically objectionable in an *additional* sense (the one brought out by the Intelligent Design analogy, above), that sounds like a very strange bargain to make, like meeting the devil at the crossroads and selling your soul in exchange for *not* being able to play the fiddle.

¹⁷³ Also, of course, the gap theorist who believes in “partially defined predicates” is purchasing incompleteness at the expense of expressive power.

As such, semantic dialetheism (like semantic gappiness) should be rejected.

III. Keeping Score

Since the discussion of semantic dialetheism represents the last anti-dialetheist argument in the present work, and my overall project here is to defend classical orthodoxy against the dialetheist challenge, this looks like a good time to go back and see how we've done, relative to the criteria for success for that project laid out at the end of the Introduction. Remember, at that point, we noted that the *main* argument for dialetheism (although, as we've seen, hardly the *only* argument for it) is the argument from the Liar Paradox and related semantic antinomies, and spades of books about the Liar are *already* occupying an awful lot of shelf space in University libraries. Why one more? What particular contribution can we make? I laid out five standard flaws in the Liar-solving argumentation presented in those books, and indicated that if I could do better, the present work would have earned its right to exist. Here they are again:

- (a) They fail to take seriously the idea that the paradoxes constitute *arguments for dialetheism*, and as such fail to shoulder their rational obligation to avoid begging the question.
- (b) They ignore all arguments for dialetheism *other* than the arguments from the semantic paradoxes.
- (c) Even as solutions to the semantic paradoxes, they ultimately fail to save consistency, because they are vulnerable to various strengthened and

revenge paradoxes, whereby the very concepts used in the solution are used to forge new and more virulent paradoxes.

- (d) They purchase consistency only at the expense of expressive or inferential power, for example by giving up on natural language or classical logic.

Now, let's take these one at a time. Start with (a). I've presented reasons to reject crucial premises of all of the arguments for dialetheism I've considered that go beyond mere contradiction avoidance, and I've provided arguments against it, like the epistemic argument in Chapter Eight, that, rather than relying on the raw counterintuitiveness of true contradictions, draw further counter-intuitive consequences from it, consequences that, in the case of the epistemic argument, Priest himself has been at pains to try to get around.

I've pretty clearly done better on count (b) than any extant book of paradox-solving, addressing three of Priest's strongest non-Liar-related arguments for metaphysical dialetheism (the argument from the paradoxes of motion and change, the argument from deontic dilemmas and the argument from naïve set theory), as well as Mares' case for semantic dialetheism and Ripley's closely related argument about ordinary speaker's observation-reports about allegedly inconsistent situations. Of course, there are other arguments for dialetheism out there, and more are generated all the time as an idea that was, not so long ago, held by perhaps a half dozen philosophers in obscure corners of the globe begins to have a certain cutting-edge attractiveness to some up-and-coming graduate students and junior professors. (I've met a few of these new wave dialetheists, and if they nevertheless represent a small fringe relative to philosophy of

logic as a whole, I strongly suspect that their numbers will increase over time.¹⁷⁴)

Addressing every argument for dialetheism would be fruitless in a Tristan Shandy sort of way in any case, and at any rate I haven't tried, but I think I can claim to have at least taking a long step towards correcting the almost-universal silence from our side of the debate about almost all of the dialtheist's arguments other than the argument from the Liar Paradox.

While on the subject of the relationship between the Liar and other motivations for dialetheism, it's worth pausing to make something explicit that we've touched on before. Graham Priest has argued that all of the "paradoxes of self-reference" (a family he takes to include the Liar and most similar semantic paradoxes, even ones like Yablo that plainly lack self-reference, and also Russell's Paradox for naïve set theory) are of the same type, and that paradoxes of the same type should be given the same type of solution. They all fall, he thinks, into the Inclosure Schema, whose elements are existence, closure and transcendence...i.e. that something (a set, or a sentence) exists, that it falls under some category, and that it doesn't. One could object that cases where closure and transcendence seem to be mixed (which is a fancy way of saying "things that seem to be inconsistent") need not have anything in common except arising from error, and I'm sympathetic to that, but I think two further points can be made:

1-It would be very difficult for Priest to show that his solution to the solution to (most of) the semantic paradoxes on the one hand and the paradoxes of naïve set theory

¹⁷⁴ This is, moreover, something that a lot of traditional philosophers of the sort who take challenges to basic logical principles to be too ridiculous to dignify with a response will simply have to get used to. No matter how convincing we might find the case for atheism from problem of evil, some bright philosophers will always find theism appealing, new Plantingas will be born and new defenses against the problem will always crop up. Philosophical theism, and the debate about the philosophy of religion, are here to stay. Similarly, I think, at this point dialetheism, and the debate about non-contradiction, is here to stay. Any theorist who takes the goal of philosophical inquiry to be the securing of universal agreement to their particular ideas is setting themselves up for a lifetime of disappointment.

on the other hand, which is biting the bullet (and thus accepting that all three elements hold) is more unified *at the level of abstraction at which the Inclosure Schema operates* than my solution of denying the existence condition in both cases by denying the existence *meaningful* sentences, the sorts of things we can apply truth talk to, which attribute untruth to themselves, and denying that scientifically superfluous, ontologically un-parsimonious abstract mathematical objects like the Russell Set exist. Of course, I'm denying the existence of one for reasons specific to semantic considerations, and the existence of the other for reasons specific to considerations about the relationship between set theory and external reality, but the same could be said for the specificity of Priest's arguments in each domain for embracing all three elements of the Inclosure Schema. For example, thinks sets exists, and that the cumulative hierarchy fails to exhaust the total universe of sets, and he justifies that with a discussion of the debate about these matters in the philosophy of mathematics, considerations that will not come into play if he ever dignifies what he calls he "heroic" solution with a response and he needs to justify his belief in the existence of *meaningful* sentences that say of themselves that they are untrue.

2-Curry is a semantic paradox of the same type as the Liar, whose paradoxical properties result from what it attempts to say about its own truth-value, and Priest's solutions to the two are flagrantly disunified. This seems obvious enough that, if a consequence of the Inclosure Schema is that the Liar and Curry are not of the same type, so much the worse for the Inclosure Schema as a diagnostic tool that helps us sort out paradoxes into relevantly similar types.

On (c), I've argued extensively that the meaningfulness solution that I present, rooted in the "nothing above and beyond" principle, is *not* vulnerable to standard strengthened and revenge paradoxes. There are good, principled reasons why "this sentence is meaningless," "this sentence is either false or meaningless," "it would be a mistake of some sort to call this sentence true" and other standard-type revenge paradoxes fall completely flat as efforts to re-introduce inconsistency to the framework of the meaningfulness solution. Of course, it's always possible that new and unforeseen revenge paradoxes of a radically different type will create a problem for the view...that can't be ruled out in advance. (Everything, even including the statement which you are now reading, should be believed fallibilistically.) That said, it looks like we can be confident that standard revenge-making from standard revenge procedures isn't going to have any traction given the meaningfulness solution.

As far as (d) goes, this is, I think, one of the great virtues of the meaningfulness solution. It allows us to hold onto classical logic, with its intuitively appealing rules, so deeply embedded in our ordinary reasoning practices, and its tremendous inferential power. It doesn't force us to abandon natural language with its rich expressive resources in favor of relativizing truth to languages and retreating to some artificial language or artificially regimented fragment of natural language. Of course, while it doesn't ban anyone from uttering whatever they like for fear that inconsistency will result, it does tell us that some things that people utter that might initially seem meaningful are not (which, at that level of generality, seems obviously true in any case, as can be shown with many other examples), but at any rate, nothing is lost if we have good, solid reasons, quite independent of consistency-maintenance, to suppose that the relevant utterances are not

meaningful, and we can tell a plausible story about why some people might be mistaken about them, and no further counterintuitive consequences about meaningfulness result (e.g. we are not forced to say that “the quoted passage has eight words in it” is meaningless, or anything of the sort).

We can sum up the importance of (a)-(d) in the following way.

In his more extreme moods, Graham Priest likes to call the position I’m calling monaletheism a dogma inherited from Aristotle.¹⁷⁵ One level, this might seem like rhetorical over-reach, the equivalent of calling belief in the existence of the external physical world a dogma inherited from Descartes, but on another level, it’s a challenge worth answering.

If we merely accept monaletheism because it’s what we’ve always assumed it to be true, and it sounds crazy to question it, then it is just a dogma. It is only when we take seriously the idea that it is as open to the possibility of rational revision, and address the case for rejecting it head-on without straw-manning the opposition or resorting to question-begging arguments or *ad hoc* maneuvers with nothing but consistency-maintenance going for them, that it can become more than just a dogma. I hope that I’ve made a contribution to bringing that about.

¹⁷⁵ E.g. Priest 2006, p. 7. Of course, he says ‘The Law of Non-Contradiction,’ not monaletheism, but that is sloppy, since, as discussed in the Introduction, Priest himself non-dogmatically imports the LNC into his own logic LP, simply affirming that certain ‘meta-contradictions’ $(\alpha \wedge \neg\alpha) \wedge \neg(\alpha \wedge \neg\alpha)$ are entailed by every ‘raw’ contradiction he endorses.

IV. A Final Thought About Truth

While my main project here has been to defend classical orthodoxy against the challenge from gluttness, I've made an effort along the way, when we've encountered the issue, first in discussion of the Liar Paradox and then in discussion of over- and partially-defined predicates, to defend it against the gappy challenge as well. I've leaned on some of Graham Priest's arguments for assistance there, and his discussion of gaps in the new material added to the second edition, he puts the point underlying those arguments very nicely, saying that truth is the goal of assertion and that "[e]ssentially, since assertion is a one-player game, anything less than truth is falsity: there is no middle ground, such as drawing, in a two-player game."¹⁷⁶

Now I have no desire to complicate things here by getting into the sticky details of actually evaluating the teleological theory of truth that Priest endorses. It is far from clear to me just in what sense truth can be said to be the goal of assertion—certainly, it's one of the goals of *sincere* assertion, but I'm not optimistic about the prospects for cashing out the notion of sincerity here without making things hopelessly circular—or, if it was in some clear sense the most important goal of assertion in general, whether this fact would in some sense be the nature of truth, or would illuminate the nature of truth in some special way. In fact, I lean towards the view that truth isn't the sort of thing that "has" a nature, and we're better off saying something broadly deflationary about it. Fortunately, hacking through that thicket of issues would take us well off-course, and we can just say this: *Whether or not* Priest is right about the nature of truth, his analogy is a good one, and it helpfully illuminates the issues at stake in the argument between gap

¹⁷⁶ Priest (2006a), p. 267

theorists and partisans of Bivalence and/or the Law of the Excluded Middle (depending on which principle is being denied by any given version of gap theory). The gap theorist takes truth to be a two-player game, fought out between the truth of a statement and its falsity and/or the truth of its negation, and takes it that sometimes the two sides can tie.¹⁷⁷ The rest of us think that this is impossible, because truth is one player game and ties are conceptually impossible.

I think that, as is often the case, Priest's best insights outrun his dialetheist theoretical commitments, and can in fact be used to undermine them. This is a perfect case in point. The one-player game metaphor cuts just as nicely against dialetheism as it does against gap theory. *Both* views depart from classical orthodoxy in seeing truth as being, in some sense, a two-player game where the two sides can sometimes tie. They differ on the question of how to conceptualize these ties. Do we think of a tie between truth and falsity as a situation where both players lose (gap), or where both players win (glut)? Either way, the question of falsity's fortunes is a *further* question that we have to answer after we've answered the question of how truth fared in the game.

As we've seen, JC Beall believes in both gappy ties and glutty ties.¹⁷⁸

Interestingly enough, he also endorses the sort of broadly deflationary program about truth that I've indicated my sympathy for, and the spirit of which is certainly at work in my solution to the Liar Paradox. He's provided a nice, vivid analogy for how he takes

¹⁷⁷ For the sake of simplicity, for the rest of this discussion I will speak only of falsity, since I take falsity and truth of negation to be equivalent. I think that for present purposes I can do that without begging any questions.

¹⁷⁸ For the sake of simplicity, unless otherwise noted, all references to Beall's vies in the main text are to the gap-tolerant position argued for in Beall (2005). In Beall (2009), he switches positions on the issue of gaps, and claims that all "spandrels" of the truth predicate are either glutty or (just) false. I critique Beall's new view in Appendix B.

truth to work that he thinks captures the relevant features and makes plausible his deflationary endorsement of dialetheism.

Suppose that, for one reason or another, we decide to revise the syntax of English. In particular, we decide to remove ‘is true’ (and ‘is false’) from the vocabulary, and we add the unitary predicates “‘Aiehtela accepts’ and ‘Aiehtelanu accepts.’

With syntax modified, we venture to teach our children the language, the proper usage of ‘Aiehela accepts’ and ‘Aiehtelanu accepts,’ in effect, the target semantics. Of course, children learn best through pictures, and so we teach them the target semantics by purporting to tell them about Aiehtela’s (-nu’s) acceptance behavior...

...For example, if grass is green, then Aiehtela accepts ‘grass is green’. If Aiehtela accepts ‘grass is green,’ then grass is green. If grass is not green, then Aiehtelanu accepts that grass is green, and if Aiehtelanu accepts that grass is green, then grass is not green....

...After all, the children are familiar with agents *accepting* various claims, and perhaps (to make things vivid) the children picture Aiehtela raising his/her hand to indicate acceptance...

...Our children, of course, might immediately ask after Aiehtela (-nu)—what’s s/he like, where does s/he live and so on. Inasmuch as we are teaching them the language (as opposed to teaching them philosophy), we should avoid the questions, perhaps noting that (in terms of its linguistic role) ‘Aiehtela (-nu) accepts’ is an expressive device, one that affords generalizations that, for practical purposes, we couldn’t otherwise make.¹⁷⁹

The punchline, of course, comes with sentences like “‘Aiehtela does not accept this sentence.’” The rules governing Aiehtela-talk seem to dictate that Aiehtela does and does not accept that sentence, for reasons familiar from our discussion of the Strengthened Liar in Chapter Five. The children progressing through their lessons in Aiehtela-talk, in wry analogy to philosophers who find the notion of true contradictions difficult to comprehend, are puzzled by this. After all, they can’t picture it. Beall suggests that this may be a good time in their development for them to learn to make a more robust distinction between the language the pictures are designed to help them learn and the pictures themselves, but allows that it may be possible after all to provide a visual aid,

¹⁷⁹ Beall, pp. 200-204

since the paradoxical sentence “is a situation in which *both* Aihtela and Aihtelanu accept the same sentence... That’s all that need be pictured (if one insists on picturing at all).”¹⁸⁰

Of course, in a language that globally substituted Aihtela-talk for truth-talk, the meaningless solution would work just as well as it does for the Liars of our language. In fact, I think there would be far less intuitive resistance to it from some quarters, because Aihtela-talk is much more obviously a device for expressing agreement with the content of the sentence to which Aihtela’s acceptance is attributed than is truth-talk a device for expressing agreement with the content of the sentence to which truth is attributed. Or, at least, it would be far more obvious to adults, who understand that there is no such person as Aihtela, and that any attempt to claim that Aihtela sometimes raises his/her hand in response to questions, commands, meaningless jumbles of words, or for that matter, bits of melting candle wax, while it could be *pictured*, is nonsensical. Those simply aren’t the sorts of things with reference to which the question of Aihtela raising or not raising his/her hand arises.

The important point is this. Beall, as a dialetheist, thinks that, once we know that Aihtela raises his/her hand to something, there is a *further question* about whether or not Aihtelanu raises his/her hand to it as well. Further, as a gap theorist, he takes it that, if Aihtela does not raise his/her hand to something, there’s a further question about whether or not Aihtelanu will raise her hand to something. The Aihtelanu questions are not *settled* by the answers to Aihtela.

To tie things back to Priest’s game metaphor, we can imagine the children being shown videos as well as pictures, and seeing these strange deities boxing. Sometimes Aihtela knocks out Aihtelanu, sometimes Aihtelanu knocks out Aihtela, sometimes they

¹⁸⁰ Beall, p. 203

manage to somehow each knock the other out and sometimes they circle each other for round after round without either knocking the other out.

This, then, is the point of all the argumentation of the last nine chapters. I deny that this anything about this imagery is accurate or helpful. When it comes the still pictures, no extra information is given to the children by showing them pictures of Aihtelanu raising his/her hand or failing to do so in addition to pictures of Aihtela raising his/her hand or failing doing so, and much possibly confusion is introduced by the idea that these are two separate issues.

Coming to the video, I context that Aihtela has no one to box with. S/he isn't boxing. S/he is, rather, playing solitaire. In solitaire, you win, or you lose (by failing to win), but not both. There's no one to tie with, and once one realizes this, the question of whether ties *would* be best conceptualized as a matter of both players winning or both players losing is sublimely irrelevant. Solitaire is a one-player game.

So is truth.

References

- Barwise, J. and Etchemendy 1987, J. *The Liar: An Essay on Truth and Circularity* (Oxford).
- Batens, Diderik 1990, "Against Global Paraconsistency," *Studies in Soviet Thought* 39: 209-229
- Beall, JC, "Transparent Disquotationalism," in Beall and Armour-Garb 2006, p. 7-22
- ___ (ed.) (2003), *Liars and Heaps: New Essays on Paradox* (Oxford).
- ___ (ed.) (2007), *Revenge of the Liar* (Oxford).
- ___ and Brad Armour-Garb and Graham Priest (eds.) (2004), *The Law of Non-Contradiction: New Philosophical Essays* (Oxford).
- ___ and Brad Armour-Garb (eds.) (2005), *Deflationism and Paradox* (Oxford).
- ___ and Greg Restall (2006), *Logical Pluralism* (Oxford).
- Blackburn, Simon 1999, *Think: A Compelling Introduction To Philosophy* (Oxford).
- Boolos, G. 1998, *Logic, Logic and Logic* (Harvard).
- Bueno, Otavio and Mark Colyvan, "Logical Non-Apriorism and the 'Law' of Non-Contradiction" in Priest, Beall and Armour-Garb 2004, pp. 156-175
- ___ and Newton da Costa and Stephen French 1998, "Is There a Zande Logic?" *History and Philosophy of Logic* 19: 41-54
- ___ and Jean-Yves Béziau and Newton da Costa (2005), "On The Usefulness Of Paraconsistent Logic," in Vanderveken, 465-478
- ___ and Oystein Linnebo (eds.) (2009), *New Waves In Philosophy Of Mathematics* (Macmillan).
- ___ and Scott Shalkowski 2009, "Modalism and Logical Pluralism," *Mind* 118: 295-321
- Carnap, Rudolf 2001, "Empiricism, Semantics and Ontology" in Sosa nad Martinich 2001, pp. 425-435
- Chihara, Charles 1979, "The Semantic Paradoxes: A Diagnostic Investigation," *The Philosophical Review* 88: 590-618

- Colyvan, Mark 2008, "Who's Afraid Of Inconsistent Mathematics?" *Protosociology* 25: 24-35
- ___ 2009, "Applying Inconsistent Mathematics," in Bueno and Linnebo, 2009
- Denyer, Nicholas, "Priest's Paraconsistent Arithmetic," *Mind* 104: 567-575
- Frege, Gottlob 1979, *Posthumous Writings*, Edited by Hans Hermes, Friedrich Kambartel and Friedrich Kaulbach, Translated by Peter Long and Roger White (Chicago).
- ___ 2007, *The Foundations of Arithmetic*, Translated by Dale Jacquette (Longman.)
- Goldstein, Laurence 2004, "The Barber, Russell's Paradox, Catch-22, God and More: A Defense of a Wittgensteinian Conception of Contradiction," in Priest, Beall and Armour-Garb 2004, pp. 295-313
- ___ 2006, "Fibonacci, Yablo, and the Cassationist Approach to Paradox," *Mind* 115: 867-889
- Hallett, Michael (1984) *Cantorian Set Theory and Limitation of Size* (Oxford).
- Harman, Gilbert 1986, *Change in View: Principles Of Reasoning* (MIT).
- Hofweber, Thomas, "Validity, Paradox and the Ideal of Deductive Logic" in Beall 2007, pp. 145-158
- ___ and Anthony Everett (eds.) (2000), *Empty Names, Fiction and the Puzzles of Non- Existence* (Stanford)
- Kripke, Saul 1975, "Outline of a Theory of Truth," *Journal of Philosophy* 72: 690-716
- Lewis, David 1991, *Parts of Classes* (Oxford).
- ___ 2004, "Letters to Beall and Priest," in Priest, Beall and Armour-Garb 2004, pp. 176-177
- MacFarlane, John 2004, *In What Sense (If Any) Is Logic Normative For Thought?* (Unpublished Manuscript).
- Maddy, Penelope 1990, *Realism in Mathematics* (Oxford).
- ___ 2007, *Second Philosophy: A Naturalistic Method* (Oxford).

- Mares, Edwin 2004, "Semantic Dialetheism," in Priest, Beall and Armour-Garb, pp. 264-275
- McGee, Van 1990, *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth* (Hackett).
- McGinn, Colin 2000, *Logical Properties: Identity, Existence, Predication, Necessity, Truth*, (Oxford).
- Nolan, Daniel 1997, "Impossible Worlds: A Modest Approach," *Notre Dame Journal of Formal Logic* 38: 535-572
- Patterson, Douglass 2007, "Inconsistency Theories: The Significance of Semantic Ascent," *Inquiry* 50: 575-589
- Pelletier, Jeffrey 2006, "Review of *The Law of Non-Contradiction: New Philosophical Essays* by Graham Priest; JC Beall; Bradley Armour-Garb," *Bulletin of Symbolic Logic* 12: 131-135
- Priest, Graham 1979, "Two Dogmas of Quineanism," *The Philosophical Quarterly* 29: 298-301
- ____ 1995, *Beyond the Limits of Thought* (Cambridge).
- ____ 1996, "On Inconsistent Arithmetics: A Reply To Denyer," *Mind* 105: 649-659
- ____ 1998a, "What's So Bad About Contradictions?" *Journal Of Philosophy* 98 410-26
- ____ 1998b, "The Import Of Inclosure: Some Comments On Grattan-Guinness," *Mind* 107: 835-840
- ____ 2000, "On The Principle Of Uniform Solution: A Reply To Smith," *Mind* 109: 123-126
- ____ 2005, *Towards Non-Being: The Logic and Metaphysics of Intentionality* (Oxford).
- ____ 2006a, *In Contradiction* (Oxford).
- ____ 2006b, *Doubt Truth To Be A Liar* (Oxford).
- Prior, Arthur 1967, *Past, Present and Future* (Oxford).

- Quine, Willard Van Orman 1970, *Philosophy of Logic* (Prentice Hall).
- ____ 1992, *Pursuit of Truth* (Harvard).
- Ripley, David 2009, *Contradictions at the borders* (Unpublished Manuscript).
- Routley, Richard 1982, *Exploring Meinong's Jungle and Beyond: An Investigation of Noneism and the Theory of Items* (ANU).
- Smith, Quentin 1993, *Language and Time* (Oxford).
- ____ 2003, "Reference to the Past and Future," in Smith and Jokic, pp. 357-390
- ____ and L. Nathan Oaklander 1995, *Time, Change and Freedom: An Introduction to Metaphysics* (Routledge).
- ____ and Aleksander Jokic (eds.) (2003), *Time, Tense and Reference* (MIT).
- Soames, Scott 1999, *Understanding Truth* (Oxford).
- Sosa, David and A.P. Martinich 2001, *Analytic Philosophy: An Anthology* (Blackwell).
- Streumer, Bart 2007, "Reasons and Entailment," *Erkenntnis* 66: 353-374
- Swinburne, Richard 1993, *The Coherence Of Theism* (Oxford).
- Tappenden, Jamie 1999, "Negation, Denial, and Language Change in Philosophical Logic" in Wansing and Gabbay, pp. 261-298
- Takho, Tuomas 2009, "The Law of Non-Contradiction as a Metaphysical Principle," *Australasian Journal Of Logic* 7: 32-47
- Tanaka, Koji 2003, "Three Schools Of Paraconsistency," *Australasian Journal of Philosophy* 1: 28-42
- Vanderveken, Daniel 2008, *Logic, Thought and Action (Logic, Epistemology and the Unity of Science)* (Springer).
- Wansing, Heinrich and Dov M. Gabbay (eds.) (1999), *What Is Negation?* (Kluwer).
- Weir, Alan 1998, "Naïve Set Theory Is Innocent!" *Mind* 107: 763-798
- Yablo, Stephen 2000, "A Paradox of Existence," in Everett and Hofweber 2000, pp. 275-312