PARACONSISTENCY AND DIALETHEISM

Graham Priest

1 INTRODUCTION

1.1 Delineating the Topic of this Article

This article is about paraconsistent logic, logic in which contradictions do not entail everything. Though the roots of paraconsistency lie deep in the history of logic, its modern developments date to just before the middle of the 20th century. Since then, paraconsistent logic—or better, logics, since there are many of them—have been proposed and constructed for many, and very different, reasons. The most philosophically challenging of these reasons is dialetheism, the view that some contradictions are true. Though this article will also discuss other aspects of paraconsistency, it will concentrate specifically on its dialetheic aspects. Other aspects of the subject can be found in the article 'Paraconsistency: Preservational Variations' in this volume of the *Handbook*. The subject also has close connections with relevant logic. Many related details can therefore be found in the article 'Relevant and Substructural Logics', in Volume 4 of the *Handbook*.

In the following two parts of this article, we will look at the history of the subject before about 1950. We will look at the history of paraconsistency; then we will look at the history of dialetheism. In the next two parts, we will turn to the modern developments, those since about 1950; first paraconsistency, then dialetheism. In the final three parts of the article will look at some important issues that bear on paraconsistency, or on which paraconsistency bears: the foundations of mathematics, the notion of negation, and rationality.

1.2 Defining the Key Notions: Paraconsistency

Let us start, however, with definitions of the two central notions of the article. Perhaps the major motivation behind paraconsistency in the modern period has been the thought that there are many situations where we wish to handle inconsistent information in a sensible way—and specifically, where we have to infer from it. (We *may* also wish to revise the information; but that is another matter. And a knowledge of what does or does not follow sensibly from the information may be necessary for an intelligent revision.)

Let $\vdash$ be any relation of logical consequence.[1] Let $\neg$ denote negation. (What, exactly, this is, we will come back to later in this essay.) Then the relation is called *explosive* if it satisfies the principle of *Explosion*:

$$\alpha, \neg\alpha \vdash \beta$$

or, as it is sometimes called, *ex contradictione quodlibet*. Explosion is, on the face of it, a most implausible looking inference. It is one, however, that is valid in "classical logic", that is, the orthodox logic of our day.

Clearly, an explosive notion of logical consequence is not a suitable vehicle for drawing controlled inferences from inconsistent information. A necessary condition for a suitable vehicle is therefore that Explosion fail. This motivates the now standard definition: a consequence relation is *paraconsistent* if it is not explosive. The term was coined by Miró Quesada at the Third Latin American Symposium on Mathematical Logic in 1976.[2]

Given a language in which to express premises and conclusions, a set of sentences in this language is called *trivial* if it contains all sentences. Let Σ be a set of sentences, and suppose that it is inconsistent, that is: for some α, Σ contains both α and $\neg\alpha$. If $\vdash$ is explosive, the deductive closure of Σ under $\vdash$ (that is, the set of consequences of Σ) is trivial. Conversely, if $\vdash$ is paraconsistent it may be possible for the deductive closure of Σ to be non-trivial.[3] Hence, a paraconsistent logic allows for the possibility of inconsistent sets of sentences whose deductive closures are non-trivial.

Paraconsistency, in the sense just defined, is not a sufficient condition for a consequence relation to be a sensible one with which to handle inconsistent information. Consider, for example, so-called minimal logic, that is, essentially, intuitionist logic minus Explosion. This is paraconsistent, but in it $\alpha, \neg\alpha \vdash \neg\beta$, for all α and β.[4] Hence, one can infer the *negation* of anything from an inconsistency. This is not triviality, but it is clearly antithetical to the *spirit* of paraconsistency, if not the letter. It is possible to try to tighten up the definition of 'paraconsistent' in various ways.[5] But it seems unlikely that there is any purely formal necessary and sufficient condition for the spirit of paraconsistency: inconsistent information may make a nonsense of a consequence relation in so many, and quite different,

[1] In this article, I will think of such a relation as one between a set of premises and a single conclusion. However, as should be clear, multiple-conclusion paraconsistent logics are also quite feasible. In listing the premises of an inference, I will often omit set braces. I will use lower case Greek letters for individual premises/conclusions, and upper case Greek letters for sets thereof. Lower case Latin letters, p, q, r, will indicate distinct propositional parameters.

[2] The prefix 'para' has a number of different significances. Newton da Costa informed me that the sense that Quesada had in mind was 'quasi', as in 'paramedic' or 'paramilitary'. 'Paraconsistent' is therefore 'consistent-like'. Until then, I had always assumed that the 'para' in 'paraconsistent' meant 'beyond', as in 'paranormal' and 'paradox' (beyond belief). Thus, 'paraconsistent' would be 'beyond the consistent'. I still prefer this reading.

[3] Though, of course, for *certain* inconsistent Σ and paraconsistent $\vdash$, the set of consequences may be trivial.

[4] Since $\alpha \vdash \beta \to \alpha$ and $\beta \to \alpha \vdash \neg\alpha \to \neg\beta$.

[5] See, for example, Urbas (1990).

ways.[6] Better, then, to go for a clean, simple, definition of paraconsistency, and leave worrying about the spirit to individual applications.

1.3 Defining the Key Notions: Dialetheism

No similar problems surround the definition of 'dialetheism'. The fact that we are faced with, or even forced into operating with, information that is inconsistent, does not, of course, mean that that information is true. The view that it may be is *dialetheism*. Specifically, a *dialetheia* is a true contradiction, a pair, α and $\neg\alpha$, which are both true (or equivalently, supposing a normal notion of conjunction, a truth of the form $\alpha \wedge \neg\alpha$). A *dialetheist* is therefore a person who holds that some contradictions are true. The word 'dialetheism' and its cognates were coined by Priest and Routley in 1981, when writing the introduction to Priest, Routley, and Norman (1989).[7] Before that, the epithet 'paraconsistency' had often been used, quite confusingly, for both dialetheism and the failure of explosion.[8]

A *trivialist* is a person who believes that *all* contradictions are true (or equivalently, and more simply, who believes that everything is true). Clearly, a dialetheist need not be a trivialist (any more than a person who thinks that some statements are true must think that all statements are true). As just observed, a person may well think it appropriate to employ a paraconsistent logic in some context, or even think that there is a uniquely correct notion of deductive logical consequence which is paraconsistent, without being a dialetheist. Conversely, though, it is clear that a dialetheist must subscribe to a paraconsistent logic—at least when reasoning about those domains that give rise to dialetheias—unless they are a trivialist.

A final word about truth. In talking of true contradictions, no particular notion of truth is presupposed. Interpreters of the term 'dialetheia' may interpret the notion of truth concerned in their own preferred fashion. Perhaps surprisingly, debates over the nature of truth make relatively little difference to debates about dialetheism.[9]

---

[6] For example, as we will see later, the T-schema (plus self-reference) gives triviality in any logic that contains Contraction ($\alpha \to (\alpha \to \beta) \vdash \alpha \to \beta$). Yet Contraction is valid in many logics that standardly get called paraconsistent.

[7] Chapters 1 and 2 of that volume cover the same ground as the next two parts of this essay, and can be consulted for a slightly different account.

[8] The term *dialetheia* was motivated by a remark of Wittgenstein (1978), p. 256, where he compares the liar sentence to a Janus-headed object facing both truth and falsity. A di/aletheia is, thus, a two-way truth. Routley, with an uncharacteristic purism, always preferred 'dialethic', 'dialethism', etc., to 'dialetheic', 'dialetheism', etc. Forms with and without the 'e' can now both be found in the literature.

[9] See Priest (2000a).

2 PARACONSISTENT LOGIC IN HISTORY

2.1 *Explosion in Ancient Logic*

Having clarified the central notions of this essay, let us now turn to its first main theme. What are the histories of these notions? Paraconsistency first. It is sometimes thought that Explosion is a principle of inference coeval with logic. Calling the received theory of inference 'classical' may indeed give this impression. Nothing could be further from the truth, however. The oldest system of formal logic is Aristotle's syllogistic;[10] and syllogistic is, in the only way in which it makes sense to interpret the term, paraconsistent. Consider, for example, the inference:

$$\frac{\begin{array}{c}\text{Some men are animals.}\\ \text{No animals are men.}\end{array}}{\text{All men are men.}}$$

This is not a (valid) syllogism. Yet the premises are contradictories. Hence contradictions do not entail everything. Aristotle is well aware of this, and points it out explicitly:

> In the first figure [of syllogisms] no deduction whether affirmative or negative can be made out of opposed propositions: no affirmative deduction is possible because both propositions must be affirmative, but opposites are the one affirmative, the other negative... In the middle figure a deduction can be made both of opposites and of contraries. Let A stand for good, let B and C stand for science. If then one assumes that every science is good, and no science is good, A belongs to every B and to no C, so that B belongs to no C; no science, then is science. Similarly if after assuming that every science is good one assumed that the science of medicine is not good; for A belongs to every B but to no C, so that a particular science will not be a science... Consequently it is possible that opposites may lead to a conclusion, though not always or in every mood.[11]

Syllogistic is not a propositional logic. The first logicians to produce a propositional logic were the Stoics. But there is no record of any Stoic logician having endorsed Explosion either. Nor do any of the critics of Stoic logic, like Sextus Empiricus, mention it. (And this surely would have been grist for his mill!) Stoic logicians did not, therefore, endorse Explosion.

---

[10]The investigation of logic in the East, and especially in India, starts at around the same time as it does in Greece. But for some reason, Indian logic never developed into a formal logic in anything like the Western sense. There is, at any rate as far as I am aware, no Indian logician who endorsed Explosion or anything like it. There are good reasons for this, as we will see in due course.

[11]*Prior Analytics* 63^b31-64^a16. The translation is from Barnes (1984). Note also that there is nothing suspicious about taking some of the terms of the syllogism to be the same. As this quote shows, Aristotle explicitly allows for this.

It might be thought that Stoic logic was, none the less, explosive, on the following grounds. Consider the principle of inference called the Disjunctive Syllogism:

$$\alpha, \neg\alpha \vee \beta \vdash \beta$$

Given this, Explosion is not far away, as can be seen by the following argument, which we will call *William's argument* (for reasons that will become clear in a moment):

$$\dfrac{\alpha \quad \dfrac{\neg\alpha}{\neg\alpha \vee \beta}}{\beta}$$

(Premises are above lines; corresponding conclusions are below.) Now, Stoic logicians did explicitly endorse the Disjunctive Syllogism. It was one of their five "axioms" (indemonstrables).[12] So perhaps their logic was explosive, though they did not notice it? No. It is too much to ask one to believe that such good logicians missed a two-line argument of this kind.

The most likely explanation is that Stoic logicians did not endorse William's argument since they did not endorse the other principle it employs, Addition:

$$\alpha \vdash \alpha \vee \beta$$

Though the precise details of the Stoic account of disjunction are somewhat moot, there are reasons to suppose that the Stoics would not account a disjunction of an arbitrary α and β even as grammatical: disjunctions were legitimate when the disjuncts were exclusive, and enumerated an exhaustive partition of some situation or other (as in: It's either Monday, or it's Tuesday, or ... or it's Sunday).[13]

2.2 Explosion in Medieval Logic

The understanding of disjunction—and conjunction for that matter—in anything like its contemporary truth-functional sense seems not to emerge in logic until about the 12th century.[14] It is therefore not surprising that the first occurrence of William's argument seems to appear at about the same time. Though the evidence is circumstantial, it can be plausibly attributed to the 12th century Paris logician, William of Soissons, who was one of the parvipontinians, logicians who made a name for themselves advocating Explosion.[15] William's argument was well known within about 100 years. It can be found quite clearly in Alexander Neckham at the end of the 12th century,[16] and it is clearly stated in the writings of the mid-14th century logician now known only as Pseudo-Scotus.[17]

---

[12] See, e.g., Bocheński (1963), p. 98.
[13] For a discussion of Stoic disjunction, see Jennings (1994), ch.10.
[14] See Sylvan (2000), section 5.3.
[15] See Martin (1986).
[16] See Read (1988), p. 31.
[17] See Kneale and Kneale (1962), p. 281f.

The history of the principle of Explosion in medieval logic after this time is a tangled one, and surely much of it still remains to be discovered. What one can say for sure is that logical consequence, and with it Explosion, was one of the topics that was hotly debated in medieval logic. (One thing that muddies the waters is the fact that logicians tended to run together logical consequence and the conditional, calling both *consequentiae*.)

Most of the major logicians distinguished between different notions of logical consequence.[18] The various notions go by different names for different logicians. But it was not uncommon to distinguish between a "material" notion of validity, according to which Explosion held, and a "formal" notion of validity, requiring some sort of connection between premises and conclusion. Unsurprisingly, Explosion did not hold in the latter.[19]

One factor that drove towards accepting Explosion, at least for material consequences, was a definition of validity that started to become popular around the 13th century, and which may be stated roughly as follows:[20]

> A valid inference is one in which it is impossible for the premises to be true *and* the conclusion to be false.

The account was by no means accepted by all. But given the common assumption that it is impossible for contradictions to be true, $\neg\Diamond(\alpha \wedge \neg\alpha)$, and a few plausible principles concerning truth functional conjunction and modality, it follows that, for arbitrary β, $\neg\Diamond((\alpha \wedge \neg\alpha) \wedge \beta)$. Assuming that the 'and' italicized in the above definition is truth functional, this is just Explosion.[21]

Of particular note in the present context are the Cologne logicians of the late 15th century. These rejected Explosion as a formally valid principle, and with it the Disjunctive Syllogism (thereby prefiguring modern paraconsistent and relevant logic), specifically on ground that both fail if we are reasoning about situations in which, maybe *per impossibile* both α and $\neg\alpha$ hold.[22]

As is well known, the study of logic went into decline after this period. The subtle debates of the great medieval logicians were forgotten. Formal logic came to be identified largely with syllogistic. A few propositional inferences, such as *modus ponens* ($\alpha, \alpha \rightarrow \beta \vdash \beta$) and the Disjunctive Syllogism, are sometimes stated in logic texts, but Explosion is not one of them (and neither is Addition). Even the greatest logician between the middle ages and the end of the 19th century, Leibniz,

---

[18] Perhaps with some indication of which notion of consequence was appropriate in which sort of case. See Stump (1989), p. 262f.

[19] See, e.g., Sylvan (2000), 5.4.

[20] See, e.g., Boh (1982), Ashworth (1974), pp. 120ff.

[21] Many definitions of validity are to be found in medieval logic. The one in question goes back well beyond the 13th century. Indeed, arguably it goes back to Megarian logicians. But in earlier versions, the conjunction was not necessarily interpreted truth functionally. For a full discussion, see Sylvan (2000).

[22] See Ashworth (1974), p. 135. A similar line was run by de Soto in the early 16th century. See Read (1993), pp. 251-5.

does not mention Explosion in his writings.[23] It seems fair to say, therefore, that oblivion ensured that paraconsistency became the received position in logic once more.

2.3 Explosion in Modern Logic

Things changed dramatically with the rise of modern logic at the end of the 19th century. For the logical theory invented by Frege, and subsequently taken up by Russell—classical logic—is explosive. (This needs no documentation for contemporary readers.) But Frege and Russell were introducing (or reintroducing) into logic something very counter-intuitive.[24] Since neither of them was much of a student of medieval logic (nor could they have been, given the poor scholarship of the period at the time), what needs discussion is where the drive for Explosion came from. The motors are at least two.[25]

Frege and Russell realised the power of a truth-functional analysis of connectives, and exploited it relentlessly. But they were over-impressed by it, believing, incorrectly, that all interesting logical connectives could be given a truth functional analysis. The point was later to be given central dogmatic status by Russell's student, Wittgenstein, in the *Tractatus*. Now, if one gives a truth functional analysis of the conditional (*if...then...*), the only plausible candidate is the material conditional, $\neg\alpha \vee \beta$ ($\alpha \supset \beta$). Given this, the most natural principle for the conditional, *modus ponens,* collapses into the Disjunctive Syllogism, to which the logic is therefore committed. Given that the truth functional understanding of disjunction immediately vouchsafes Addition, Explosion is an immediate corollary.

The second source of Explosion is, in many ways, more fundamental. It is a fusion of two things. The first is an account of negation. How, exactly, to understand negation is an important issue in the history of logic, though one that often lurks beneath the surface of other disputes (especially concerning the conditional). (More of this later.) In the middle of the 19th century an account of propositional negation was given by George Boole. According to Boole, negation acts like set-theoretic complementation. Specifically, for any α, α and $\neg\alpha$ partition the set of all situations: the situations in which $\neg\alpha$ is true are exactly those where α fails to be true. (Note that this is not entailed by a truth functional account of negation. Some paraconsistent logics have a different, but still truth functional, theory of negation.) Boole's way of looking at negation, and more generally, the

[23] At least according to the account of Leibniz' logic provided by Kneale and Kneale (1962), pp. 336ff.

[24] And Russell, at least, was aware of this. There is a folklore story concerning Russell—Nick Griffin tells me that a version of it can be found in Joad (1927)—which goes as follows. Russell was dining at high table at Trinity, when he mentioned to one of his fellow dons that in his logic a contradiction implies everything. According to one version, the don, righly incredulous, challenged him to deduce the fact that he was the Pope from the claim that $2 = 3$. After some thought, Russell replied: 'Well, if $2 = 3$ then, subtracting 1 from both sides, it follows that $1 = 2$. Now the Pope and I are two. Hence, the Pope and I are one. That is, I am the Pope'.

[25] There are certainly others. For example, Explosion is endorsed by by Peano (1967), p.88, but his reasons for it are not stated.

analysis of propositional operators in set-theoretic terms, was highly influential on the founders of modern logic. Russell, for example, took Boole's work to be the beginning of 'the immense and surprising development of deductive logic' of which his own work formed a part.[26]

The second element entering into the fusion is an account of validity, to the effect that an inference is valid if there are no situations, or models, as they were to come to be called, in which the premises are true and the conclusion is false. The account is not stated by either Frege or Russell, as far as I am aware. It is implicit, however, at least for propositional logic, in the truth-tabular account of validity, and was developed and articulated, by Tarski and other logicians, into the modern model-theoretic account of validity.

Neither the Boolean theory of negation nor the model-theoretic account of validity, on its own, delivers explosion.[27] But together they do. For a consequence of Boole's account is that exactly one of α and $\neg\alpha$ holds in every model. It follows that there is no model in which α and $\neg\alpha$ hold and which β fails. The model-theoretic account does the rest. (The argument is clearly a relative of the medieval argument for Explosion based on the modal definition of validity.)

It is interesting to note that William's argument for Explosion does not seem to figure in discussions during this period. It was left to C.I.Lewis to rediscover it. (It is stated in Lewis and Langford (1932), p. 250.) There is a certain irony in this, since Lewis was one of the major early critics of Russell on the matter of the conditional. Lewis, whilst rejecting a material account of the conditional, was driven by William's argument to accepting an account according to which contradictions do imply everything ("strict implication"). It is perhaps also worth noting that both Russell and Lewis perpetuate the medieval confusion of validity and the conditional, by calling both 'implication'. Pointing out this confusion was to allow Quine to defend the material conditional as an account of conditionality.[28] The problems with the material conditional go much deeper than this, though.[29]

Lewis was not the only critic of "classical logic" in the first half of the century. The most notable critics were the intuitionists. But though the intuitionists rejected central parts of Frege/Russell logic, they accepted enough of it to deliver Explosion.[30] First, they, accepted both the Disjunctive Syllogism and Addition. They also accepted a model-theoretic account of validity (albeit with models of a somewhat different kind). They did not accept the Boolean account of negation. But according to their account, though α and $\neg\alpha$ may both fail in some situations,

[26] Russell (1997), p. 497. I also have a memory of him calling Boole the 'father of modern logic', but I am unable to locate the source. Boole himself was not a modern logician. Though he may have stretched this to its limits, syllogistic was squarely the basis of his work. He might plausibly, therefore, be thought of as the last of the great traditional logicians.

[27] We will see later that the model-theoretic account of validity is quite compatible with paraconsistent logic. As for negation, it may follow from the Boolean account that contradictions are true in no situation; but this says nothing about consequence.

[28] Quine (1966), p. 163f.

[29] See Priest (2001a), ch. 1.

[30] Though they were criticized on this ground, for example by Kolmogorov. Dropping Explosion from Intuitionist logic gives Johannson's "minimal logic". See Haack (1974), p. 101f.

they cannot, at least, both hold. This is sufficient to give Explosion.

So this is how things stood half way through the 20th century. Classical logic had become entrenched as the orthodox logical theory. Various other logical theories were known, and endorsed by some "deviant" logicians—especially modal and intuitionist logic; but all these accounts preserved enough features of classical logic to deliver Explosion. Explosion, therefore, had no serious challenge.

We will take up the story concerning paraconsistency again in a later section. But now let us back-track, and look at the history of dialetheism.

3 DIALETHEISM IN HISTORY

3.1 *Contradiction in Ancient Philosophy*

Can contradictions be true? At the beginning of Western philosophy it would seem that opinions were divided on this issue. On the face of it, certain of the Presocratics took the answer to be 'yes'. Uncontroversially, Heraclitus held that everything was in a state of flux. Any state of affairs described by α changes into one described by $\neg\alpha$. More controversially, the flux state was one in which both α and $\neg\alpha$ hold.[31] Hence, we find Heraclitus asserting contradictions such as:[32]

> We step and do not step into the same rivers; we are and we are not.

On the other hand, Parmenides held that *what is* has certain amazing properties. It is one, changeless, partless, etc. A major part of the argument for this is that one cannot say of what is that it is not, or vice versa:[33]

> For never shall this be forcibly maintained, that things that are not are, but you must hold back your thought from this way of inquiry, nor let habit, born of much experience, force you down this way, by making you use an aimless eye or an ear and a tongue full of meaningless sounds: judge by reason the strife-encompassed refutation spoken by me.

This certainly sounds like a proto-statement of the Law of Non-Contradiction. And Zeno, according to tradition Parmenides' student, made a name for himself arguing that those who wished to deny Parmenides' metaphysics ended up in contradiction—which he, at least, took to be unacceptable.

The dialogues of Plato are somewhat ambivalent on the matter of contradiction. For a start, in the *Republic* we find Socrates enunciating a version of the Law of Non-Contradiction, and then arguing from it:[34]

---

[31] And Heraclitus held, it would seem, that the flux state is *sui generis*. That is, $\alpha\wedge\neg\alpha$ entails neither α nor $\neg\alpha$.

[32] Fragment 49a; translation from Robinson (1987).

[33] Fragment 7; translation from Kirk and Raven (1957), p. 248.

[34] 436*b*. Hamilton and Cairns (1961).

> It is obvious that the same thing will never do or suffer opposites in the same respect in relation to the same thing and at the same time.

In the later dialogue, the *Parmenides*, the same Socrates expresses less confidence:

> Even if all things come to partake of both [the form of like and the form of unlike], and by having a share of both are both like and unlike one another, what is there surprising in that? ... when things have a share in both or are shown to have both characteristics, I see nothing strange in that, Zeno, nor yet in a proof that all things are one by having a share in unity and at the same time many by sharing in plurality. But if anyone can prove that what is simple unity itself is many or that plurality itself is one, then shall I begin to be surprised.[35]

Thus, it may be possible for things in the familiar world to have inconsistent properties, though not the forms.[36] What to make of the later part of this puzzling dialogue is notoriously hard. But taking the text at face value, Parmenides does succeed in showing that oneness itself does have inconsistent properties of just the kind to surprise Socrates.

Interpreting texts such as these, especially the Presocratics, is fraught with difficulty, and it may well be thought that those I have cited as countenancing violations of Law of Non-Contradiction did not really do so, but were getting at something else. It should be noted, then, that a commentator no less than Aristotle interpreted a number of the Presocratics as endorsing contradictions.[37] In Book 4 of the *Metaphysics*, he takes them in his sights, and mounts a sustained defence of the Law of Non-Contradiction, which he enunciates as follows (5^b18-22):[38]

> For the same thing to hold good and not hold good simultaneously of the same thing and in the same respect is impossible (given any further specifications which might be added against dialectical difficulties).

The rest of the text is something of an exegetical nightmare.[39] The Law is, Aristotle says, so certain and fundamental that one cannot give a proof of it (5^b22-27). He then goes on straight away to give about seven or eight arguments for it (depending on how one counts). He calls these *elenchic* demonstrations, rather than

[35] 129*b*, *c*. Hamilton and Cairns (1961).

[36] It is tempting to read Socrates as saying that things may be inconsistent in relational ways. That is, an object may be like something in some ways and unlike it in others. This would not be a real contradiction. But this cannot be what Socrates means. For exactly the same can be true of the forms. The form of the good might be like the form of unity in that both are forms, but unlike it in that it is the highest form.

[37] Heraclitus and Protagoras are singled out for special mention. Protagoras claimed that if someone believes something, it is true (for them). Hence α may be true (for some person), and $\neg\alpha$ may be true (for someone else). This does not sound quite like a contradiction. But of course, if someone believes $\alpha \wedge \neg\alpha$, then that is true (for them).

[38] Kirwan (1993).

[39] For a full analysis of the text, see Priest (1998).

proofs. Exactly what this means is not clear; what is clear is that the opponent's preparedness to utter something meaningful is essential to the enterprise. But then, just to confuse matters, only the first of the arguments depends on this preparedness. So the latter arguments do not seem to be elenchic either.

Leaving this aside, the arguments themselves are varied bunch. The first argument (6^a28-7^b18) is the longest. It is tangled and contorted, and it is not at all clear how it is supposed to work. (Some commentators claim to find two distinct arguments in it.) However one analyses it, though, it must be reckoned a failure. The *most generous* estimate of what it establishes is that for any predicate, P, it is impossible that something should be P and not be P ($\neg\Diamond(Pa \wedge \neg Pa)$) which sounds all well and good at first. But one who really countenances violations of the Law of Non-Contradiction may simply agree with this! For they may still hold that for some P and a, $Pa \wedge \neg Pa$ as well. It will follow, presumably, that $\Diamond(Pa \wedge \neg Pa)$, and hence that $\Diamond(Pa \wedge \neg Pa) \wedge \neg\Diamond(Pa \wedge \neg Pa)$. This is a contradiction (we might call it a "secondary contradiction"), but contradiction is clearly not a problem in this context.[40]

When we turn to the other arguments (7^b18-9^a6), matters are even worse. For the majority of these arguments, if they establish anything at all—and they certainly have steps at which one might cavil—establish not the Law of Non-Contradiction, but what we might call the Law of Non-Triviality: it is not possible that *all* contradictions be true. Dialetheists may of course agree with this. Aristotle, in fact, seems to slide between the two Laws with gay abandon, possibly because he took his main targets to be not just dialetheists, but trivialists.[41] A couple of the arguments do not even attempt to establish the Law of Non-Triviality. What they conclude is that it is impossible for anyone to *believe* that all contradictions are true. It is, of course, compatible with this that all contradictions *are* true, nonetheless.

Aristotle's defence of the Law of Non-Contradiction must therefore be reckoned a failure. It's historical importance has been completely out of proportion to its intellectual weight, however. Since the entrenchment of Aristotelian philosophy in the medieval European universities, the Law of Non-Contradiction has been high orthodoxy in Western philosophy. It is taken so much for granted that there has, improbably enough, been no sustained defence of the Law since Aristotle's. (Of which other of Aristotle's philosophical views can one say this?)

It is worth noting, finally, that the Law of Non-Contradiction—and its mate the Law of Excluded Middle, also defended in Book 4 of *Metaphysics*—are not *logical* principles for Aristotle, but *metaphysical* principles, governing the nature of beings *qua* beings. By the time one gets to Leibniz, however, the Laws have been absorbed into the logical canon.

[40] In fact, in many paraconsistent logics, such as LP, $\neg(\alpha \wedge \neg\alpha)$ is a *logical truth*, and in their modalised versions, so is $\neg\Diamond(\alpha \wedge \neg\alpha)$. Every contradiction therefore generates secondary contradictions.

[41] The slide between 'some' and 'all' is also not uncommon in others who have tried to defend the law.

3.2 A Minority Voice: Neoplatonism and its Successors

There is just one tradition that stands out against the orthodox acceptance of the Law of Non-Contradiction. This is the metaphysical tradition that starts with the Neoplatonists, and goes through the great Christian mystics, Eruigina and Eckhart, and their Renaissance successors, such as Cusanus. What holds this tradition together is the belief that there is an ultimate reality, the One, or in its Christian form, the Godhead. This reality is, in some sense, responsible for the existence of everything else, including humankind. Humankind, being alienated from the reality, finds its ultimate fulfillment in union with it.

This tradition draws on, amongst other things, some of the later Platonic dialogues, and especially the *Parmenides*. As we noted, in the second half of this dialogue Parmenides shows the One to have contradictory properties. It is perhaps not surprising, then, to find writers in this tradition having a tendency to say contradictory things, especially about the ultimate reality. For example, referring explicitly to *Parmenides* 160^b2-3, Plotinus says:[42]

> The One is all things and no one of them; the source of all things is not all things and yet it is all things...[43]

Eckhart says, sometimes, that the Godhead is being; and, at other times, that it is beyond being—and thus not being.[44] And Cusanus says that:[45]

> in no way do they [distinctions] exist in the absolute maximum. The absolute maximum... is all things, and whilst being all, is none of them; in other words, it is at once the maximum and minimum of being.

Cusanus also attacked contemporary Aristotelians for their attachment to the Law of Non-Contradiction.[46]

The contradictory claims about the One are no mere aberration on the part of these writers, but are driven by the view of the One as the ground of all things that are. If it were itself anything, it would not be this: it would be just another one of those things. Consequently, one cannot say truly anything to the effect that the One is such and such, or even that it is (*simpliciter*); for this would simply make it one of the many. The One is therefore ineffable. As Plotinus puts it (*Ennead* V.5.6):

> The First must be without form, and, if without form, then it is no Being; Being must have some definition and therefore be limited; but the First cannot be thought of as having definition and limit, for thus

[42] *Ennead* V.2.1. Translation from MacKenna (1991).

[43] MacKenna inserts the words 'in a transcendental sense' here; but they are not in the text. I think that this is a misplaced application of the principle of charity.

[44] See Smart (1967), p. 450.

[45] *Of Learned Ignorance* I.3. Translation from Heron (1954).

[46] See Maurer (1967).

> it would not be the Source, but the particular item indicated by the definition assigned to it. If all things belong to the produced, which of them can be thought of as the supreme? Not included among them, this can be described only as transcending them: but they are Being and the Beings; it therefore transcends Being.

But even though the One is ineffable, Plotinus still describes it as 'the source of all things', 'perfect' (*Ennead* V.2.1), a 'Unity', 'precedent to all Being' (*Ennead* VI.9.3). Clearly, describing the ineffable is going to force one into contradiction.[47]

3.3 Contradiction in Eastern Philosophy

We have not finished with the Neoplatonist tradition yet, but before we continue with it, let us look at Eastern Philosophy, starting in India. Since very early times, the Law of Non-Contradiction has been orthodox in the West. This is not at all the case in India. The standard view, going back to before the Buddha (a rough contemporary of Aristotle) was that on any claim of substance there are four possibilities: that the view is true (and true only), that it is false (and false only), that it is neither true nor false, and that it is both true and false. This is called the *catuskoti* (four corners), or tetralemma.[48] Hence, the possibility of a contradiction was explicitly acknowledged. The difference between this view and the orthodox Western view is the same as that between the semantics of classical logic and the four-valued semantics for the relevant logic of First Degree Entailment (as we shall see). In classical logic, sentences have exactly one of the truth values T (true) and F (false). In First Degree Entailment they may have any combination of these values, including both and neither. Just to add complexity to the picture, some Buddhist philosopers argued that, for some issues, all or none of these four possibilities might hold. Thus, the major 2nd century Mahayana Buddhist philosopher Nāgārjuna is sometimes interpreted in one or other of these ways. Arguments of this kind, just to confuse matters, are also sometimes called *catuskoti*. Interpreting Nāgārjuna is a very difficult task, but it is possible to interpret him, as some commentators did, as claiming that these matters are simply ineffable.[49]

The Law of Non-Contradiction has certainly had its defenders in the East, though. It was endorsed, for example, by logicians in the Nyāyā tradition. This influenced Buddhist philosophers, such as Darmakīrti, and, via him, some Buddhist schools, such as the Tibetan Gelug-pa. Even in Tibet, though, many Buddhist schools, such as the Nyngma-pa, rejected the law, at least for ultimate truths.

Turning to Chinese philosophy, and specifically Taoism, one certainly finds utterances that look as though they violate the Law of Non-Contradiction. For ex-

[47]Nor can one escape the contradiction by saying that the One is not positively characterisable, but may be characterised only negatively (the *via negativa*). For the above characterisations are positive.

[48]See Raju (1953-4).

[49]This is particularly true of the Zen tradition. See Kasulis (1981), ch.2.

ample, in the *Chuang Tzu* (the second most important part of the Taoist canon), we find:[50]

> That which makes things has no boundaries with things, but for things to have boundaries is what we mean by saying 'the boundaries between things'. The boundaryless boundary is the boundary without a boundary.

A cause of these contradictions is not unlike that in Neoplatonism. In Taoism, there is an ultimate reality, Tao, which is the source and generator of everything else. As the *Tao Te Ching* puts it:[51]

The Tao gives birth to the One.
The One gives birth to the two.
The Two give birth to the three—
The Three give birth to every living thing.

It follows, as in the Western tradition, that there is nothing that can be said about it. As the *Tao Te Ching* puts it (ch. 1):

The Tao that can be talked about is not the true Tao.
The name that can be named is not the eternal name.
Everything in the universe comes out of Nothing.
Nothing—the nameless—is the beginning...

Yet in explaining this situation, we are forced to say things about it, as the above quotations demonstrate.

Chan (Zen) is a fusion of Mahayana Buddhism and Taoism. As might therefore be expected, the dialetheic aspects of the two metaphysics reinforce each other. Above all, then, Zen is a metaphysics where we find the writings of its exponents full of apparent contradictions. Thus, for example, the great Zen master Dōgen says:[52]

> This having been confirmed as the Great Teacher's saying, we should study immobile sitting and transmit it correctly: herein lies a thorough investigation of immobile sitting handed down in the Buddha-way. Although thoughts on the immobile state of sitting are not limited to a single person, Yüeh-shan's saying is the very best. Namely: 'thinking is not thinking'.

or:[53]

[50] 22.6. Translation from Mair (1994).

[51] Ch. 42. Translation from Kwok, Palmer and Ramsey (1993). What the one, two and three are is a moot point. But in one interpretation, the one is the T'ai-Chi (great harmony); the two are Yin and Yang.

[52] Kim (1985), p. 157.

[53] Tanahashi (1985), p. 107.

> An ancient buddha said, 'Mountains are mountains, waters are waters.' These words do not mean that mountains are mountains; they mean that mountains are mountains. Therefore investigate mountains thoroughly...

Now interpreting all this, especially the Chinese and Japanese writings, is a hard and contentious matter. The writings are often epigrammatic and poetical. Certainly, the writings contain assertions of contradictions, but are we meant to take them literally? It might be thought not. One suggestion is that the contradictions are uttered for their perlocutionary effect: to shock the hearer into some reaction. Certainly, this sort of thing plays a role in Zen, but not in Mahayana Buddhism or Taoism. And even in Zen, contradictions occur in even the theoretical writings.

More plausibly, it may be suggested that the contradictions in question have to be interpreted in some non-literal way. For example, though ultimate reality is literally indescribable, what is said about it gives some metaphorical description of its nature. This won't really work either, though. For the very reason that ultimate reality is indescribable is precisely because it is that which brings all beings into being; it can therefore be no being (and so to say anything about it is contradictory). At least this much of what is said about the Tao must be taken literally, or the whole picture falls apart.[54]

3.4 Hegel

Let us now return to Western philosophy, and specifically to Hegel. With the philosophers we have met in the last two sections, because their utterances are often so cryptic, it is always possible to suggest that their words should not be taken at face value. By contrast, Hegel's dialetheism is ungainsayable. He says, for example:[55]

> ...common experience... says that *there is* a *host* of contradictory things, contradictory arrangements, whose contradiction exists not merely in external reflection, but in themselves... External sensuous motion is contradiction's immediate existence. Something moves, not because at one moment it is here and at another there, but because at one and the same moment it is here and not here, because in this "here", it at once is and is not.

Why does he take this view?

For a start, Hegel is an inheritor of the Neoplatonic tradition.[56] Hegel's One is Spirit (*Geist*). This creates Nature. In Nature there are individual consciousnesses

[54] It is true that in Chinese philosophy, unlike in Neoplatonism, the arguments that tie the parts of the picture together are not made explicit; but they are there implicitly: readers are left to think things through for themselves.

[55] Miller (1969), p. 440.

[56] The genealogy is well tracked in Kolakowski (1978), ch. 1.

(Spirit made conscious), who, by a process of conceptual development come to form a certain concept, the Absolute Idea, which allows them to understand the whole system. In this way Spirit achieves self-understanding, in which form it is the Absolute.[57]

There is much more to the story than this, of course, and to understand some of it, we need to backtrack to Kant. In the Transcendental Dialectic of the *Critique of Pure Reason*, Kant argues that Reason itself has a tendency to produce contradiction. Specifically, in the Antinomies of Pure Reason, Kant gives four pairs of arguments which are, he claims, inherent in thought. Each pair gives a pair of contradictory conclusions (that the world is limited in space and time, that it is not; that matter is infinitely divisible, that it is not; etc.) The only resolution of these contradictions, he argues, lies in the distinction between phenomena and noumena, and the insistence that our categories apply only to phenomena. The antinomies arise precisely because, in these arguments, Reason over-stretches itself, and applies the categories to noumena. There is a lot more to things than this, but that will suffice for here.[58]

Hegel criticised Kant's distinction between phenomena and noumena. In particular, he rejected the claim that the two behave any differently with respect to the categories. The conclusions of Kant's Antinomies therefore have to be accepted—the world is inconsistent:[59]

> to offer the idea that the contradictions introduced into the world of Reason by the categories of the Understanding is inevitable and essential was to make one of the most important steps in the progress of Modern Philosophy. But the more important the issue raised the more trivial was the solution. Its only motive was an excessive tenderness for the things in the world. The blemish of contradiction, it seems, could not be allowed to mar the essence of the world; but there could be no objection to attaching it to the thinking Reason, to the essence of the mind. Probably, nobody will feel disposed to deny that the phenomenal world presents contradictions to the observing mind; meaning by 'phenomenal' the world as it presents itself to the senses and understanding, to the subjective mind. But if a comparison is instituted between the essence of the world, and the essence of the mind, it does seem strange to hear how calmly and confidently the modest dogma has been advanced by one and repeated by others, that thought or Reason, and not the World, is the seat of contradiction.

Moreover, the Kantian contradictions are just the tip of an ice-berg. All our categories (or at least, all the important ones), give rise to contradiction in the same way. Thus, the contradictions concerning motion with which we started this section arise from one of Zeno's paradoxes of motion. And it is reflection on these

[57] For fuller discussion, see Priest (1989-90).

[58] Details and discussion can be found in Priest (1995), chs. 5, 6.

[59] *Lesser Logic*, Section 48. Translation from Wallace (1975).

contradictions which drives the conceptual development that forces the emergence of the concept of the Absolute Idea.

Famously, many aspects of Hegel's thought were taken up by Marx (and Engels). In particular, Marx materialised the dialectic. In the process, much of the dialetheic story was simply taken over. This adds little of novelty that is important here, though, and so we do not need to go into it.[60]

3.5 Precursors of Modern Dialetheism

So much for the Neoplatonic tradition. Outside this, dialetheists and fellow travellers are very hard to find in Western philosophy. Around the turn of the 20th century, intimations of the failure of the Law of Non-Contradiction did start to arise in other areas, however. Let us look at these, starting with Meinong.

Meinong's theory of objects had two major postulates. The first is that every term of language refers to an object, though many of these objects may not exist. The second is that all objects may have properties, whether or not they exist. In particular, with reservations that we will come back to in a moment, all objects which are characterised in certain ways have those properties which their characterisations attribute to them (the *Characterisation Principle*). Thus, for example, the fabled Golden Mountain is both golden and a mountain; and, notoriously, the round square is both round and square.

As the last example shows, some objects would appear to violate the Law of Non-Contradiction by being, for example, both round and square. Meinong was criticised on just these grounds by Russell (1905). He replied that one should expect the Law to hold only for those things that exist, or at least for those things that are possible. Impossible objects have—what else?—impossible properties. That is how one knows that they cannot exist. As he puts it:[61]

> B.Russell lays the real emphasis on the fact that by recognising such objects the principle of contradiction would lose its unlimited validity. Naturally I can in no way avoid this consequence... Indeed the principle of contradiction is directed by no one at anything other than the real and the possible.

Things are not quite as straightforward as may appear, however.[62] It is not entirely clear that Meinong does countenance violations of the Law of Non-Contradiction in the most full-blooded sense of the term. The round square is round and square, but is it round and not round? One would naturally think so, since being square entails not being round; but Meinong may well have thought that this entailment held only for existent, or at least possible, objects. Hence he may not have held there to be things with literally contradictory properties.

[60] Details can be found in Priest (1989-90).
[61] Meinong (1907), p.16.
[62] The following is discussed further in Routley (1980).

But what about the thing such that it is both round and it is not the case that it is round? This would seem to be such that it is round and it is not the case that it is round. Not necessarily. For Meinong did not hold that *every* object has the properties it is characterised as having. One cannot characterise an object into existence, for example. (Think of the existent round square).[63] The Characterisation Principle holds only for certain properties, those that are *assumptible*, or characterising. It is clear that Meinong thought that existence and like properties are not characterising, but Meinong never came clean and gave a general characterisation of characterising properties themselves. So we just do not know whether negation could occur in a characterising property. Hence, though there are certainly versions of Meinong's theory in which some objects have contradictory properties, it is not clear whether these are Meinong's.

The next significant figure in the story is the Polish logician Łukasiewicz. In 1910, Łukasiewicz published a book-length critique of Aristotle on the Law of Non-Contradiction. This has still to be translated into English, but in the same year he published an abbreviated version of it, which has.[64] Łukasiewicz gives a damning critique of Aristotle's arguments, making it clear that they have no substance. Following Meinong's lead, he also states that the Law of Non-Contradiction is not valid for impossible objects.[65] However, he does claim that the Law is a valid "practical-ethical" principle. For example, without it one would not be able to establish that one was absent from the scene of a crime by demonstrating that one was somewhere else, and so not there.[66] Given the logical acumen of the rest of the article, Łukasiewicz's position here is disappointing. One does not need a *universally* valid law to do what Łukasiewicz requires. It is sufficient that the situation in question is such as to enable one to rule out inconsistency in that particular case. (Compare: even a logical intuitionist can appeal to the Law of Excluded Middle in finite situations.) For the same reason, an inductive generalisation from this sort of situation to the *universal* validity of the Law—or even to a law covering existent objects—is quite groundless.

Another philosopher who was prepared to brook certain violations of the Law of Non-Contradiction, at around the same time, was the Russian Vasiliev.[67] Like Łukasiewicz, Vasiliev held the Law to be valid for the actual world, but he held that it might fail in certain "imaginary worlds". These are worlds where logic is different; there can be such things, just as there can be worlds where geometry is non-Euclidean. (Recall that he was writing before the General Theory of Relativity.) He did not think that all of logic could change from world to world, however. Essentially, positive logic, logic that does not concern negation (what he called

---

[63] In fact, using an unbridled form of this principle, one can establish triviality. Merely consider the thing such that it is self-identical and α, for arbitrary α.

[64] The book is Łukasiewicz (1910); the English translation of the abbreviated version is Łukasiewicz (1970).

[65] (1970), section 19.

[66] (1970), section 20.

[67] Only one of his papers has been translated into English, Vasiliev (1912-13). For further discussion of Vasiliev, see Priest (2000b).

'metalogic') is invariant across all worlds. Only negation could behave differently in different worlds.

Vasiliev also constructed a formal logic which was supposed to be the logic of these imaginary worlds, *imaginary logic*. This was not a modern logic, but a version of traditional logic. In particular, Vasiliev added to the two traditional syntactic forms 'S is P (and not also not P)', and 'S is not P (and not also P)', a third form, 'S is P and not P'. He then constructed a theory of syllogisms based on these three forms. (For example, the following is a valid syllogism: all S is M; all M is P and not P; hence, all S is P and not P.)

Though Vasiliev's logic is paraconsistent, it is not a modern paraconsistent logic: it is paraconsistent for exactly the same reason that standard syllogistic is. Nor, in a sense, is Vasiliev a dialetheist, since he held that no contradictions are true. His work clearly marks a departure from the traditional attitude towards the Law of Non-Contradiction, though.

The final figure to be mentioned in this section is Wittgenstein. Though Wittgenstein's views evolved throughout his life, they were mostly inhospitable to dialetheism. For most of his life, he held that contradictions, and especially the contradictions involved in the logical paradoxes, were senseless (in the *Tractatus*), or failed to make statements (in the transitional writings). However, towards the end of his life, and specifically in the *Remarks on the Foundations of Mathematics*, he came to reject this view:[68]

> There is *one* mistake to avoid: one thinks that a contradiction must be senseless: that is to say, if e.g. we use the signs 'p', '$\sim$', '.' *consistently*, then '$p. \sim p$' cannot say anything.

The crucial view here seems to have been that concerning language games. People play a variety of these, and if people play games in which contradictions are accepted, then contradictions are indeed valid in those games (shades of Protagoras here). The logical paradoxes might just be such. As he says:[69]

> But you can't allow a contradiction to stand: Why not? We do sometimes use this form of talk, of course, not often—but one could imagine a technique of language in which it was a regular instrument. It might, for example be said of an object in motion that it existed and did not exist in this place; change might be expressed by means of contradiction.

Unsurprisingly, Wittgenstein also had a sympathy towards paraconsistency. In 1930, he even predicted the modern development of the subject in the most striking fashion:[70]

---

[68] Wittgenstein (1978), pp. 377f.
[69] Wittgenstein (1978), p. 370.
[70] Wittgenstein (1979), p. 139.

> I am prepared to predict that there will be mathematical investigations of calculi containing contradictions and that people will pride themselves in having emancipated themselves from consistency too.

But his own efforts in this direction were not very inspired, and never came to much more than the directive 'infer nothing from a contradiction'.[71] Hence, Wittgenstein exerted no influence on future developments. Indeed, of all the people mentioned in this section it was only Łukasiewicz who was to exert an (indirect) influence on the development of paraconsistency, to which we now return.

4 MODERN PARACONSISTENCY

4.1 Background

The revolution that produced modern logic around the start of the 20th century depended upon the application of novel mathematical techniques in proof-theory, model theory, and so on. For a while, these techniques were synonymous with classical logic. But logicians came to realise that the techniques are not specific to classical logic, but could be applied to produce quite different sorts of logical systems. By the middle of the century, the basics of many-valued logic, modal logic, and intuitionist logic had been developed. Many other sorts of logic have been developed since then; one of these is paraconsistent logic.

The commencement of the modern development of paraconsistent logics occurred just after the end of the Second World War. At that juncture, it was an idea whose time had come—in the sense that it seems to have occurred to many different people, in very different places, and quite independently of each other. The result was a whole host of quite different paraconsistent logics. In this section we will look at these.[72] I will be concerned here only with propositional logics. Though the addition of quantifiers certainly raises novel technical problems sometimes, it is normally conceptually routine. I shall assume familiarity with the basics of classical, modal, many-valued and intuitionist logic. I will use $\models_X$ as the consequence relation of the logic X. C is classical logic; I is intuionist logic.

4.2 Jaśkowski and Subsequent Developments

The first influential developments in the area are constituted by the work of the Polish logician Jaśkowski, who had been a student of Łukasiewicz. Jaśkowski published a system of paraconsistent logic in 1948,[73] which he called *discussive* (or *discursive*) logic. Jaśkowski cites a number of reasons why there might be situations in which one has to deal with inconsistent information, but the main

[71] See Goldstein (1989) for discussion. According to Goldstein, the view that a contradiction entails nothing is present even in Wittgenstein's earlier writings, including the *Tractatus*.

[72] For technical information concerning the systems, see Priest (2003), where details explained in this section are discussed further. Proofs not given or referenced here can be found there.

[73] Translated into English as Jaśkowski (1969).

idea that drives his construction is indicated in the name he gives his logic. He envisages a number of people engaged in a discussion or discourse (such as, for example, the witnesses at a trial). Each participant vouchsafes certain information, which is consistent(!); but the information of one participant may contradict that of another.

Technically, the idea is implemented as follows.[74] An interpretation, I, is a Kripke-interpretation for $S5$. It helps (but is not necessary) to think of I as coming with a distinguished base-world, g. What is true at any one world is thought of as the information provided by a participant of the discourse, and what holds in the discourse is what is true at any one of its worlds. This motivates the following definitions (where $\Diamond$ is the usual possibility operator of modal logic):

α holds in I iff $\Diamond\alpha$ is true at g

$\Sigma \models_d \alpha$ iff for all I, if β holds in I, for all $\beta \in \Sigma$, then α holds in I

It is a simple matter to show that $\models_d$ is paraconsistent. A two-world model where p is true at w_1 $(= g)$ and false at w_2, but q is true at neither w_1 nor w_2 will demonstrate that $p, \neg p \not\models_d q$. It should be noted, though, that $\alpha \wedge \neg\alpha \models_d \beta$, since, whatever α is, $\alpha \wedge \neg\alpha$ holds in no I. It follows that the rule of Adjunction, $\alpha, \beta \models_d \alpha \wedge \beta$, fails. This approach may therefore be classified under the rubric of *non-adjunctive* paraconsistent logic. As is clear, different discussive logics can be obtained by choosing underlying modal logics different from $S5$.[75]

A notable feature of discussive logic is the failure of *modus ponens* for the material conditional, $\supset$: $p, p \supset q \not\models_d q$. (The two-world interpretation above demonstrates this.) In fact, it can be shown that for sentences containing only extensional connectives there is no such thing as multi-premise validity, in the sense that if $\Sigma \models_d \alpha$, then for some $\beta \in \Sigma$, $\beta \models_d \alpha$. Moreover, single-premise inference is classical. That is, $\alpha \models_d \beta$ iff $\alpha \models_C \beta$.

In virtue of this, Jaśkowski defined a new sort of conditional, the discursive conditional, $\supset_d$, defined as follows: $\alpha \supset_d \beta$ is $\Diamond\alpha \supset \beta$. It is easy to check that $\alpha, \alpha \supset_d \beta \models_d \beta$ (that is, that $\Diamond\alpha, \Diamond(\Diamond\alpha \supset \beta) \models \Diamond\beta$), provided that the accessibility relation in the underlying modal logic is at least Euclidean (that is, if wRx and wRy then xRy). This holds in $S5$, but may fail in weaker logics, such as $S4$.

The weakness produced by the failure of Adjunction, and multi-premise inferences in general, may be addressed with a quite different approach to constructing a non-adjunctive paraconsistent logic. The idea is to allow a certain amount of

[74]What follows is somewhat anachronistic, since it appeals to possible-world semantics, which were developed only some 10-15 years later, but it is quite faithful to the spirit of Jaśkowski's paper.

[75]A somewhat different approach is given in Rescher and Brandom (1980). They define validity in terms of truth preservation at all worlds, but they allow for inconsistent and incomplete worlds. What holds in an inconsistent world is what holds in any one of some bunch of ordinary worlds; and what holds in an incomplete world is what holds in all of some bunch of ordinary worlds. It can be shown that this results in the same consequence relation as discussive logic.

conjoining before applying classical consequence. Since arbitrary conjoining cannot be allowed on pain of triviality, the question is how to regulate the conjoining. One solution to this, due first, as far as I know, to Rescher and Manor (1970-71), is as follows. Given any set of sentences, Σ, a *maximally consistent* (*mc*) subset of Σ is a set $\Pi \subseteq \Sigma$, such that Π is classically consistent, but if $\alpha \in \Sigma - \Pi$, $\Pi \cup \{\alpha\}$ is classically inconsistent. Then define: $\Sigma \models_{rm} \alpha$ iff there is some mc subset of Σ such that $\Sigma \models_C \alpha$.

$\models_{rm}$ is non-adjunctive, since $p, \neg p \not\models_{rm} p \wedge \neg p$. ($\{p, \neg p\}$ has two mc subsets, $\{p\}$ and $\{\neg p\}$.) It does allow multi-premise inference, however. For example, $p, p \supset q \models_{rm} q$. ($\{p, p \supset q\}$ has only one mc subset, namely itself.) $\models_{rm}$ has an unusual property for a notion of deductive consequence, however: it is not closed under uniform substitution. For, as is easy to check, $p, q \models_{rm} p \wedge q$, but $p, \neg p \not\models_{rm} p \wedge \neg p$.[76]

A different way of proceeding, due to Schotch and Jennings (1980), is as follows. Define a *covering* of a set, Σ, to be a finite collection of disjoint sets, Σ_1, ..., Σ_n, such that for all $1 \leq i \leq n$, $\Sigma_i \subseteq \Sigma$ and is classically consistent, and for all $\alpha \in \Sigma$, at least one of the sets classically entails α. Define the *level of incoherence* of Σ, $l(\Sigma)$, to be the smallest n such that Σ has a covering of size n; if it has no such covering, then set $l(\Sigma)$ (conventionally) as ∞. If Σ is classically consistent, then $l(\Sigma) = 1$. A set such as $\{p, \neg p, q\}$ has level 2, since it has two coverings of size 2: $\{p, q\}$ and $\{\neg p\}$; $\{\neg p, q\}$ and $\{p\}$. And if Σ contains a member that is itself classically inconsistent, then $l(\Sigma) = \infty$.

Now define: $\Sigma \models_{sj} \alpha$ iff $l(\Sigma) = \infty$, or $l(\Sigma) = n$ and for every covering of Σ of size n, there is some member of it that classically entails α. The intuition to which this answers is this. We may suppose that Σ comes to us muddled up from different sources. The level of a set tells us the simplest way we can unscramble the data into consistent chunks; and however we unscramble the data in this way, we know that some source vouchsafes the conclusion.

Like $\models_{rm}$, $\models_{sj}$ is non-adjunctive, since $p, \neg p \not\models_{rm} p \wedge \neg p$. ($\{p, \neg p\}$ has level 2, with one covering: $\{p\}$, $\{\neg p\}$.) But it does allow multi-premise inference. For example, $p, p \supset q \models_{sj} q$. ($\{p, p \supset q\}$ is of level 1.) And $\models_{sj}$ is not closed under uniform substitution. For $p, q \models_{sj} p \wedge q$, but $p, \neg p \not\models_{sj} p \wedge \neg p$.

But $\models_{rm}$ and $\models_{sj}$ are not the same. For a start, since $p \wedge \neg p$ is classically inconsistent, $\{p \wedge \neg p\}$ has level ∞ and so $\{p \wedge \neg p\} \models_{sj} q$. But $\{p \wedge \neg p\}$ has one mc subset, namely the empty set, ϕ; and $\phi \not\models_C q$; hence, $p \wedge \neg p \not\models_{rm} q$. Moreover, let $\Sigma = \{p, \neg p, q, r\}$. Then Σ has two mc subsets $\{p, q, r\}$, and $\{\neg p, q, r\}$. Hence $\Sigma \models_{rm} q \wedge r$. But Σ has level 2, and one covering has the the members: $\{p, q\}$, $\{\neg p, r\}$. Hence, $\Sigma \not\models_{sj} q \wedge r$. Finally, $\models_{rm}$ is monotonic: if Σ has an mc subset that classically delivers α, so does $\Sigma \cup \Pi$. But $\models_{rm}$ is not: $p, q \models_{sj} p \wedge q$, whilst $p, \neg p, q \not\models_{sj} p \wedge q$, since $\{p, \neg p, q\}$ has level 2, and one covering is: $\{\neg p, q\}$, $\{p\}$.

We can look at the Schotch/Jennings account in a somewhat different, but illuminating, fashion. A standard definition of classical consequence is the familiar:

[76] We could define another consequence relation as the closure of $\models_{rm}$ under uniform substitution. This would still be paraconsistent.

$\Sigma \models_C \alpha$ iff for every evaluation, ν, if every member of Σ is true in ν, so is α.

Equivalently, we can put it as follows. If Σ is consistent:

$\Sigma \models_C \alpha$ iff for every $\Pi \supseteq \Sigma$, if Π is consistent, so is $\Pi \cup \{\alpha\}$

(If Σ is not consistent, then the biconditional holds vacuously.) In other words, a valid inference preserves consistency of supersets. Or, to put it another way, it preserves coherence of level 1.

Now, if Π is inconsistent, there is not much consistency to be preserved, but we may still consider it worth preserving higher levels of coherence. This is exactly what $\models_{sj}$ does. For, as is noted in Brown and Schotch (1999), if for some n, $l(\Sigma) = n$:

$\Sigma \models_{sj} \alpha$ iff for every $\Pi \supseteq \Sigma$, if $l(\Pi) = n$ then $l(\Pi \cup \{\alpha\}) = n$

(If $l(\Sigma) = \infty$, the biconditional holds vacuously.)[77] Thus, the Schotch/Jennings construction gives rives to a family of paraconsistent logics in which validity may be defined in terms of the preservation of something other than truth. Such preservational logics are the subject of another article in this *Handbook*, and so I will say no more about them here.

4.3 Dualising Intuitionism

The next sort of system of paraconsistent logic was the result of the work of the Brazilian logician da Costa starting with a thesis in 1963.[78] Da Costa, and his students and co-workers, produced many systems of paraconsistent logic, including more discussive logics. But the original and best known da Costa systems arose as follows.

In intuitionist logic, and because of the intuitionist account of negation, it is possible for neither α nor $\neg\alpha$ to hold. Thus, in a logic with a dual account of negation, it ought be possible for both α and $\neg\alpha$ to hold. The question, then, is how to dualise.

Da Costa dualised as follows. We start with an axiomatisation of positive intuitionist logic (that is, intuitionist logic without negation). The following[79] will do. The only rule of inference is *modus ponens*.

$\alpha \supset (\beta \supset \alpha)$

---

[77]¿From left to right, suppose that $l(\Sigma) = n$, $\Sigma \models_{sj} \alpha$, $\Pi \supseteq \Sigma$, and $l(\Pi) = n$. Let $\Pi_1, ..., \Pi_n$ be a covering of Π. Let $\Sigma_i = \Sigma \cap \Pi_i$. Then $\Sigma_1, ..., \Sigma_n$ is a coveringof Σ. Thus, for some i, $\Sigma_i \models_C \alpha$. Hence, $\Pi_i \models_C \alpha$, and $\Pi_1, ..., \Pi_n$ is a covering of $\Pi \cup \{\alpha\}$. Conversely, suppose that $l(\Sigma) = n$, and for every $\Pi \supseteq \Sigma$, if $l(\Pi) = n$ then $l(\Pi \cup \{\alpha\}) = n$; but $\Sigma \not\models_{sj} \alpha$. Then there is some partition of Σ, $\Sigma_1, ..., \Sigma_n$, such that for no i, $\Sigma_i \models_C \alpha$. Hence, for each i, $\Sigma_i \cup \{\neg\alpha\}$ is consistent. Thus, if $\Pi = \Sigma \cup \{\neg\alpha \wedge \sigma;\ \sigma \in \Sigma\}$, $l(\Pi) = n$. Hence, $l(\Pi \cup \{\alpha\}) = n$. But this is impossible, since α cannot be consistently added to any member of a covering of Π of size n.

[78]The most accessible place to read the results of da Costa's early work is his (1974).

[79]Taken from Kleene (1952).

$(\alpha \supset \beta) \supset ((\alpha \supset (\beta \supset \gamma)) \supset (\alpha \supset \gamma))$
$(\alpha \wedge \beta) \supset \alpha \qquad (\alpha \wedge \beta) \supset \beta$
$\alpha \supset (\beta \supset (\alpha \wedge \beta))$
$\alpha \supset (\alpha \vee \beta)$
$\beta \supset (\alpha \vee \beta)$
$(\alpha \supset \gamma) \supset ((\beta \supset \gamma) \supset ((\alpha \vee \beta) \supset \gamma))$

One obtains an axiomatization for full intuitionist logic if one adds:

$(\alpha \supset \beta) \supset ((\alpha \supset \neg\beta) \supset \neg\alpha))$
$\alpha \supset (\neg\alpha \supset \beta)$

It is clear that one certainly does not want the second of these in a paraconsistent logic; the first, being a version of *reductio ad absurdum*, is also suspect.[80]

The two most notable consequences of these principles for negation are:

$\alpha \supset \neg\neg\alpha$
$\neg(\alpha \wedge \neg\alpha)$

(though not, of course, the converse of the first). Both of these, in their own ways, can be thought of as saying that if something is true, it is not false, whilst leaving open the possibility that something might be neither. To obtain a paraconsistent logic, it is therefore natural to take as axioms the claims which are, in some sense, the duals of these:

$\mathbf{1}^{\neg}$ $\neg\neg\alpha \supset \alpha$

$\mathbf{2}^{\neg}$ $\alpha \vee \neg\alpha$

Both of these, in their ways, can be thought of as saying that if something is not false, it is true, whilst leaving open the possibility that something may be both. Adding these two axioms to those of positive intuitionist logic gives da Costa's system C_ω.

Next, da Costa reasoned, there ought to be a way of expressing the fact that α behaves consistently (that is, is not both true and false). The natural way of doing this is by the sentence $\neg(\alpha \wedge \neg\alpha)$. Write this as α^o, and consider the principles:

$\mathbf{1}^o$ $\beta^o \supset ((\alpha \supset \beta) \supset ((\alpha \supset \neg\beta) \supset \neg\alpha)))$

$\mathbf{2}^o$ $(\alpha^o \wedge \beta^o) \supset ((\alpha \wedge \beta)^o \wedge (\alpha \vee \beta)^o \wedge (\alpha \supset \beta)^o \wedge (\neg\alpha)^o)$

The first says that the version of the *reductio* principle we have just met works provided that the contradiction deduced behaves consistently. The second (in which the last conjunct of the consequent is, in fact, redundant) expresses the plausible thought that if any sentences behave consistently, so do their compounds. Adding these two axioms to C_ω gives the da Costa system C_1.

[80] Though, note, even a paraconsistent logician can accept the principle that if something entails a contradiction, this fact establishes its negation: versions of this inference are valid in many relevant logics.

The addition of this machinery in C_1 allows us to define the strong negation of α, $\neg^*\alpha$, as: $\neg\alpha \wedge \alpha^o$. $\neg^*\alpha$ says that α is consistently false. It is possible to show that $\neg^*$ has all the properties of classical negation.[81] But as is well known, the addition of classical negation to intuitionist logic turns the positive part into classical logic. (Using the properties of classical negation, it is possible, reasoning by cases in a standard fashion, to establish Peirce's Law: $((\alpha \supset \beta) \supset \alpha) \supset \alpha$, which is the difference between positive intuitionist and classical logics.) Hence, the positive logic of C_1 is classical logic.

It might be thought that one needs more than α^o to guarantee that α behaves consistently. After all, in contexts where contradictions may be acceptable, why might we not have $\alpha^o \wedge \alpha \wedge \neg\alpha$? In virtue of this, it might be thought that what is required in condition 1^o is not α^o, but $\alpha^o \wedge \alpha^{oo}$. Of course, there is no *a priori* guarantee that this behaves consistently either. So it might be thought that what is required is $\alpha^o \wedge \alpha^{oo} \wedge \alpha^{ooo}$; and so on. Let us write α^n as $\alpha^o \wedge ... \wedge \alpha^{o...o}$ (where the last conjunct has n 'o's). Then replacing 'o' by 'n' in 1^o and 2^o gives the da Costa system C_n $(1 \leq n < \omega)$. Just as in C_1, in each C_n, a strong negation $\neg^*\alpha$ can be defined as $\neg\alpha \wedge \alpha^n$, and the collapse of the positive part into classical logic occurs as before.

Semantics for the C systems were discovered by da Costa and Alves (1977). Take the standard truth-functional semantics for positive classical logic. Thus, if ν is an evaluation, $\nu(\alpha \vee \beta) = 1$ iff $\nu(\alpha) = 1$ or $\nu(\beta) = 1$; $\nu(\alpha \supset \beta) = 1$ iff $\nu(\alpha) = 0$ or $\nu(\beta) = 1$, etc. Now allow ν to behave non-deterministically on negation. That is, for any α, $\nu(\neg\alpha)$ may take any value. Validity is defined in the usual way, in terms of truth preservation over all evaluations. It is clear that the resulting system is paraconsistent, since one can take an evaluation that assigns both p and $\neg p$ the value 1, and q the value 0.

The system just described is, in fact, none of the da Costa systems. In a certain sense, it is the most basic of a whole family of logics which extend positive classical logic with a non-truth-functional negation. The C_n systems can be obtained by adding further constraints on evaluations concerning negation. Thus, if we add the conditions:

(i) If $\nu(\neg\neg\alpha) = 1$ then $\nu(\alpha) = 1$

(ii) If $\nu(\alpha) = 0$ then $\nu(\neg\alpha) = 1$

we validate $1^\neg$ and $2^\neg$. Adding the conditions:

If $\nu(\beta^n) = \nu(\alpha \supset \beta) = \nu(\alpha \supset \neg\beta) = 1$ then $\nu(\alpha) = 0$

If $\nu(\alpha^n) = \nu(\beta^n) = 1$ then $\nu((\alpha \wedge \beta)^n) = \nu((\alpha \vee \beta)^n) = \nu((\alpha \supset \beta)^n) = (\neg\alpha)^o = 1$

[81] Specifically, $\neg^*$ satisfies the conditions:

$(\alpha \supset \beta) \supset ((\alpha \supset \neg^*\beta) \supset \neg^*\alpha))$

$\neg^*\neg^*\alpha \supset \alpha$

which give all the properties of classical negation. See da Costa and Guillaume (1965).

then gives the system C_n $(1 \leq n < \omega)$.

The semantics for C_ω are not quite so simple, since positive intuitionist logic is not truth-functional. However, non-deterministic semantics can be given as follows.[82] A *semi-evaluation* is any evaluation that satisfies the standard conditions for conjunction and disjunction, plus (i), (ii), and:

If $\nu(\alpha \supset \beta) = 1$ then $\nu(\alpha) = 0$ or $\nu(\beta) = 1$

If $\nu(\alpha \supset \beta) = 0$ then $\nu(\beta) = 0$

A valuation is now any semi-evaluation, ν, satisfying the further condition: if α is anything of the form $\alpha_1 \supset (\alpha_2 \supset (...\alpha_n)...)$, where α_n is not itself a conditional, then if $\nu(\alpha) = 0$, there is a semi-valuation, ν', such that for all $1 \leq i < n$, $\nu'(\alpha_i) = 1$ and $\nu'(\alpha_n) = 0$. Validity is defined in terms of truth preservation over all evaluations in the usual way.

As we have seen, all the C systems can be thought of as extending a positive logic (either intuitionistic or classical) with a non-truth-functional negation. They are therefore often classed under the rubric of *positive plus* logics. A singular fact about all the positive plus logics is that the substitution of provable equivalents breaks down. For example, α and $\alpha \wedge \alpha$ are logically equivalent, but because negation is not truth functional, there is nothing in the semantics to guarantee that $\neg\alpha$ and $\neg(\alpha \wedge \alpha)$ take the same value in an evaluation. Hence, these are not logically equivalent.

Da Costa's systems are the result of one way of producing something which may naturally be thought of as the dual of intuitionist logic. There are also other ways. Another is to dualise the Kripke semantics for intuitionist logic. A Kripke semantics for intuitionist logic is a structure $\langle W, R, \nu \rangle$, where W is a set (of worlds), R is a binary relation on W that is reflexive and transitive, and ν maps each propositional parameter to a truth value at every world, subject to the heredity condition: if xRy and $\nu_x(p) = 1$, $\nu_y(p) = 1$. The truth conditions for the operators are:

$\nu_w(\alpha \wedge \beta) = 1$ iff $\nu_w(\alpha) = 1$ and $\nu_w(\beta) = 1$

$\nu_w(\alpha \vee \beta) = 1$ iff $\nu_w(\alpha) = 1$ or $\nu_w(\beta) = 1$

$\nu_w(\alpha \supset \beta) = 1$ iff for all w' such that wRw', if $\nu_{w'}(\alpha) = 1$ then $\nu_{w'}(\beta) = 1$

$\nu_w(\neg\alpha) = 1$ iff for all w' such that wRw', $\nu_{w'}(\alpha) = 0$

(Alternatively, $\neg\alpha$ may be defined as $\alpha \supset \perp$ where $\perp$ is a logical constant that takes the value 0 at all worlds.) It is not difficult to show that the heredity condition follows for all formulas, not just parameters. An inference is valid if it is truth-preserving at all worlds of all interpretations.

Dualising: everything is exactly the same, except that we dualise the truth conditions for negation, thus:

[82] Folowing Loparić (1986).

$\nu_w(\neg\alpha) = 1$ iff there is some w', such that $w'Rw$ and $\nu_{w'}(\alpha) = 0$

It is easy to check that the general heredity condition still holds with these truth conditions. Since nothing has changed for the positive connectives, the positive part of this logic is intuitionist, but whereas in intuitionist logic we have $\alpha \wedge \neg\alpha \models_I \beta$ and $\alpha \models_I \neg\neg\alpha$, but not $\beta \models_I \alpha \vee \neg\alpha$ or $\neg\neg\alpha \models_I \alpha$, it is now the other way around. Details are left as an exercise. Here, though, is a counter-model for Explosion. Let $W = \{w_0, w_1\}$; $w_0 R w_1$; at w_0, p and q are both false; at w_1, p is true and q is false. It follows that $\neg p$ is true at w_1, and hence w_1 gives the counter-model.[83]

Despite similarities, the logic obtained in this way is distinct from any of the C systems. It is easy to check, for example, that $\neg\alpha$ is logically equivalent to $\neg(\alpha \wedge \alpha)$, and more generally, that provable equivalents are inter-substitutable.

Yet a third way to dualise intuitionist logic, is to dualise its algebraic semantics.[84] A Heyting algebra is a distributive lattice with a bottom element, $\perp$, and an operator, $\supset$, satisfying the condition:[85]

$a \wedge b \leq c$ iff $a \leq b \supset c$

which makes $\perp \supset \perp$ the top element. We may define $\neg\alpha$ as $\alpha \supset \perp$. A standard example of a Heyting algebra is provided by any topological space, T. The members of the algebra are the open subsets of T; $\wedge$ and $\vee$ are union and intersection; $\perp$ is the empty set, and $a \supset b$ is $(\overline{a} \vee b)^o$, where overlining indicates complementation, and ^{o} here is the interior operator of the topology. It is easy to check that $\neg a = \overline{a}^o$. It is well known that for finite sets of premises, intuitionist logic is sound and complete with respect to the class of all Heyting algebras—indeed with respect to the class of Heyting algebras defined by topological spaces. That is, $\alpha_1, ..., \alpha_n \models_I \beta$ iff for every evaluation into every such algebra $\nu(\alpha_1 \wedge ... \wedge \alpha_n) \leq \nu(\beta)$.

We now dualise. A *dual Heyting algebra* is a distributive lattice with a top element, $\top$, and an operator, $\subset$, satisfying the condition:

$a \leq b \vee c$ iff $a \subset b \leq c$

which makes $\top \subset \top$ the bottom element. We may define $\neg a$ as $\top \subset a$. It is not difficult to check that if T is any topological space, then it produces a dual Heyting algebra whose elements are the *closed* sets of the space; $\wedge$ and $\vee$ are union and intersection; $\top$ is the whole space; and $a \subset b$ is $(a \wedge \overline{b})^c$, where ^{c} is the closure operator of the space. $\neg b$ is clearly $\overline{b}^c$. Validity is defined as before.

We may call the logic that this construction gives *closed set* logic. Again, we have $\beta \models_I \alpha \vee \neg\alpha$ and $\neg\neg\alpha \models_I \alpha$, but not their duals. Verification is left as an

[83] A version of these semantics can be found, in effect, in Rauszer (1977). In this, Rauszer gives a Kripke semantics for a logic he calls 'Heyting-Brower Logic'. This is intuionist logic *plus* the duals of intuionist $\neg$ and $\supset$.

[84] As discovered by Goodman (1981).

[85] I use the same symbols for logical connectives and the corresponding algebraic operators, context sufficing to disambiguate.

exercise, but a counter-model to Explosion is provided by the real numbers with their usual topology. Consider an evaluation, ν, such that $\nu(p) = [-1, +1]$, and $\nu(q) = \phi$. Then $\nu(p \wedge \neg p) = \{-1, +1\}$, which is not a subset of ϕ. (This example illustrates how the points in the set represented by $p \wedge \neg p$ may be thought of as the points on the topological boundary between the sets represented by p and $\neg p$.) It is to be noted that closed set logic is distinct from all the C systems. For example, it is easy to check that $\neg\alpha$ and $\neg(\alpha \wedge \alpha)$ are logically equivalent—and more generally, that provable equivalents are inter-substitutable. Finally, as one would expect, *modus ponens* fails for $\subset$. (It is, after all, the dual of $\supset$.) It is a simple matter to construct a topological space where $a \cap (a \cap \bar{b})^c$ is not a subset of b. (Hint: take a to be the whole space.) Indeed, it may be shown that there is no operator definable in terms of $\wedge$, $\vee$, $\subset$ and $\perp$ that satisfies *modus ponens*. Hence closed set logic is distinct from the logic obtained by dualising Kripke semantics as well.[86]

4.4 Many-Valued Logics

It is not only intuitionism that allows for truth value gaps. In many-valued logics it is not uncommon to think of one of the values as *neither true nor false*. Hence another way of constructing a paraconsistent logic is to dualise this idea, with a many-valued logic that employs the value *both true and false* or something similar. The idea that paradoxical sentences might take a non-classical truth value goes back to at least Bochvar (1939). But the idea that this might be used to construct a many-valued logic that was paraconsistent first seems to have been envisaged by the Argentinian logician Asenjo in 1954, though the ideas were not published until (1966). As well as having the standard truth values, t and f, there is a third value i, which is the semantic value of paradoxical or antinomic sentences. The truth tables for conjunction, disjunction and negation are:

$\neg$	
t	f
i	i
f	t

$\wedge$	t	i	f
t	t	i	f
i	i	i	f
f	f	f	f

$\vee$	t	i	f
t	t	t	t
i	t	i	i
f	t	i	f

and defining $\alpha \supset \beta$ as $\neg\alpha \vee \beta$ gives it the table:

$\supset$	t	i	f
t	t	i	f
i	t	i	i
f	t	t	t

[86] I suspect that they have the same conditional-free fragment, though I have never checked the details. According to Goodman (1981), p. 124, closed set logic does have a Kripke semantics. The central feature of this is that it is not truth that is hereditary, but falsity. That is, if xRy and $\nu_x(\alpha) = 0$ then $\nu_y(\alpha) = 0$.

The designated values are t and i. That is, a valid inference is one such that there is no evaluation where all the premises take the value t or i, and the conclusion does not.[87]

The logic is a very simple and natural one, and has been rediscovered a number of times since. For example, it and its properties were spelled out in more detail in Priest (1979), where it is termed LP (the Logic of Paradox), a name by which it is now standardly known. It is not difficult to see that LP is a paraconsistent logic: take the evaluation that sets p to the value i, and q to the value f to see that $p, \neg p \not\models_{LP} q$. Despite this, it is not difficult to show that the logical truths of LP are exactly the same as those of classical logic.

The same evaluation that invalidates Explosion shows that *modus ponens* for $\supset$ is not valid: $p, \neg p \vee q \not\models_{LP} q$. The logic may be extended in many ways with a many-valued conditional connective that does satisfy *modus ponens*. Perhaps the simplest such connective has the following truth table:

$\rightarrow$	t	i	f
t	t	f	f
i	t	i	f
f	t	t	t

Adding this conditional gives the logic RM_3.[88]

It is clear that many-valued paraconsistent logics may be produced in many different ways. Any many-valued logic will be paraconsistent if it has a designated value, i, such that if $\nu(p) = i$, $\nu(\neg p) = i$. Thus, Łukasiewicz continuum-valued logic (better known as a fuzzy logic) will be paraconsistent provided that the designated values include 0.5; but we will not go into this here.[89]

The semantics of LP may be reformulated in an illuminating fashion. Let 1 and 0 be the standard truth values *true* and *false*. And let us suppose that instead of taking an evaluation to be a *function* that relates each parameter to one or other of these, we take it to be a *relation* that relates each parameter to one or other, or maybe both. Let us write such an evaluation as ρ. We may think of $\alpha\rho 1$ as 'α is true (under ρ)' and $\alpha\rho 0$ as 'α is false (under ρ)'. Given an evaluation of the propositional parameters, this can be extended to an evaluation of all formulas by the standard truth-table conditions:

$\neg\alpha\rho 1$ iff $\alpha\rho 0$
$\neg\alpha\rho 0$ iff $\alpha\rho 1$

[87] Designation is crucial here. The truth tables are the same as those of Kleene's strong three valued logic. But there, the value i is thought of as *neither true nor false*, and hence not designated. This logic is not a paraconsistent logic. The designated values are not actually specified in Asenjo (1966), but designating i does seem to be faithful to his intentions.

[88] The logic is so called because it is one of a family of n-valued logics, RM_n, whose intersection is the semi-relevant logic RM (R-Mingle). I am not sure who first formulated RM_3. The earliest reference to it in print that I know is in Anderson and Belnap (1975).

[89] An argument for paraconsistency, based on a semantics with degrees of truth, was mounted by Peña in a doctoral thesis of 1979, and subsequently, e.g., in (1989). His semantics is more complex than standard Łukaziewicz continuum-valued logic, though.

$\alpha \wedge \beta\rho 1$ iff $\alpha\rho 1$ and $\beta\rho 1$
$\alpha \wedge \beta\rho 0$ iff $\alpha\rho 0$ or $\beta\rho 0$

$\alpha \vee \beta\rho 1$ iff $\alpha\rho 1$ or $\beta\rho 1$
$\alpha \vee \beta\rho 0$ iff $\alpha\rho 0$ and $\beta\rho 0$

It is an easy matter to check that ρ relates every formula to 1 or 0 (or both). Moreover, if we write:

t for: α is true and not false
f for: α is false and not true
i for: α is true and false

then one can check that the conditions produce exactly the truth tables for LP. Further, under this translation, α takes a designated value (t or i) iff it relates to 1. So the definition of validity reduces to the classical one in terms of truth-preservation:

$$\Sigma \models_{LP} \alpha \text{ iff for every } \rho, \text{ if } \beta\rho 1 \text{ for all } \beta \in \Sigma, \text{then } \alpha\rho 1$$

Hence, LP is exactly classical logic with the assumption that each sentence is either true or false, but not both, replaced with the assumption that each sentence is either true or false or both.

Given these semantics, it is natural to drop the constraint that ρ must relate every parameter to at least one truth value, and so allow for the possibility that sentences may be neither true nor false, as well as both true and false. Thus, if we repeat the above exercise, but this time allow ρ to be an arbitrary relation between parameters and $\{0, 1\}$, we obtain a semantics for the logic of First Degree Entailment (FDE). These semantics were discovered by Dunn in his doctoral dissertation of 1966, though they were not published until (1976), by which time they also had been rediscovered by others.[90] Since the semantic values of FDE extend those of LP it, too, is paraconsistent. But unlike LP it has no logical truths. (The empty value takes all these out.) It is also not difficult to show that FDE has a further important property: if $\alpha \models_{FDE} \beta$ then α and β share a propositional parameter. FDE is, in fact, intimately related with the family of relevant logics that we will come to in the next subsection.

Dunn's semantics can be reformulated again. Instead of taking evaluations to be relations, we can take them, in a classically equivalent way, to be functions whose values are *subsets* of $\{1, 0\}$. It is not difficult to check that the truth conditions of the connectives can then be represented by the following *diamond lattice*:

$$\begin{tikzcd}
& \{1\} & \\
\phi \arrow[ur] & & \{1,0\} \arrow[ul] \\
& \{0\} \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

[90]It is interesting to note that when Dunn was a student at the University of Pittsburgh he took some classes in the mathematics department where he was taught by Asenjo. Apparently, neither realised the connection between their work at this time.

If ν is any evaluation of formulas into this lattice, $\nu(\alpha \wedge \beta) = \nu(\alpha) \wedge \nu(\beta)$;[91] $\nu(\alpha \vee \beta) = \nu(\alpha) \vee \nu(\beta)$; and $\nu(\neg\alpha) = \neg\nu(\alpha)$, where $\neg$ maps top to bottom, vice versa, and maps each of the other values to itself.

Suppose that we now define validity in the standard algebraic fashion:

$\alpha_1, ..., \alpha_n \models \beta$ iff for every ν, $\nu(\alpha_1) \wedge ... \wedge \nu(\alpha_n) \leq \nu(\beta)$

Then the consequence relation is again FDE.[92] The proof of this is relatively straightforward, though not entirely obvious.

These semantics may be generalised as follows. A *De Morgan lattice* is a structure $\langle L, \wedge, \vee, \neg\rangle$, where $\langle L, \wedge, \vee\rangle$ is a distributive lattice, and $\neg$ is an involution of period two; that is, for all a, b in L:

$\neg\neg a = a$

If $a \leq b$ then $\neg b \leq \neg a$

It is easy to check that the diamond lattice is a De Morgan lattice. One may show that FDE is sound and complete not just with respect to the diamond lattice, but with respect to the class of De Morgan lattices. (Thus, the class of De Morgan lattices relates to the diamond lattice as the class of Boolean algebras relates to the two-valued Boolean algebra in classical logic.) All these results are also due to Dunn.

De Morgan lattices have a very natural philosophical interpretation. The members may be thought of as propositions (that is, as the Fregean senses of sentences). The ordering $\leq$ may then be thought of as a containment relation. Thus, $\alpha \models \beta$ iff however the senses of the parameters are determined, the sense of α contains that of β.

4.5 Relevant Logic

The final approach to paraconsistent logic that we will consider is relevant logic. What drove the development of this was a dissatisfaction with accounts of the conditional that validate "paradoxes" such as the paradoxes of material implication:

$\alpha \models (\beta \supset \alpha)$

$\neg\alpha \models (\alpha \supset \beta)$

As soon as the material account of the conditional was endorsed by the founders of classical logic, it came in for criticism. As early as a few years after *Principia Mathematica*, C.I.Lewis started to produce theories of the strict conditional, $\alpha \dashv \beta$ ($\Box(\alpha \supset \beta)$), which is not subject to these paradoxes. This conditional was, however, subject to other "paradoxes", such as:

[91] Again, I write the logical connectives and the corresponding algebraic operators using the same symbol.

[92] If we omit ϕ from the picture then, as is to be expected, we obtain LP.

$\Box\beta \models \alpha \dashv\!\exists \beta$

$\Box\neg\alpha \models \alpha \dashv\!\exists \beta$

Lewis eventually came to accept these. It is clear, though, that such inferences are just as counter-intuitive. In particular, intuition rebels because there may be no connection at all between α and β.

This motivates the definition of a *relevant logic*. If L is some propositional logic with a conditional connective, $\rightarrow$, then L is said to be *relevant* iff whenever $\models_L \alpha \rightarrow \beta$, α and β share a propositional parameter.[93] Commonality of the parameter provides the required connection of content. Though closely connected with paraconsistency, relevant logics are quite distinct. None of the paraconsistent logics that we have met so far is relevant.[94] Moreover, a relevant logic may not be paraconsistent. One of the first relevant logics, Π' of Ackermann (1956), contained the Disjunctive Syllogism as a basic rule. If this is interpreted as a rule of inference (i.e., as applying to arbitrary assumptions, not just to theorems), then Explosion is forthcoming in the usual way.

The history of relevant logic goes back, in fact, to 1928, when the Russian logician Orlov published an axiomatisation of the fragment of the relevant logic R whose language contains just $\rightarrow$ and $\neg$. This seems to have gone unnoticed, however.[95] Axiomatizations of the fragment of R whose language contains just $\rightarrow$ were given by Moh (1950) and Church (1951). The subject took off properly, though, with the collaboration of the two US logicians Anderson and Belnap, starting at the end of the 1950s. In particular, in (1958) they dropped the Disjunctive Syllogism from Ackermann's Π' to produce their favourite relevance logic E. Both E and R are paraconsistent. The results of some 20 years of collaboration between Anderson, Belnap, and their students (especially Dunn, Meyer, and Urquhart) is published as Anderson and Belnap (1975), and Anderson, Belnap, and Dunn (1992).

Initially, relevance logic was given a purely axiomatic form. For reasons that will become clear later, let us start with an axiom system for a relevant logic that Anderson and Belnap did not consider, B.

A1. $\alpha \rightarrow \alpha$
A2. $\alpha \rightarrow (\alpha \vee \beta)$ (and $\beta \rightarrow (\alpha \vee \beta)$)
A3. $(\alpha \wedge \beta) \rightarrow \alpha$ (and $(\alpha \wedge \beta) \rightarrow \beta$)
A4. $\alpha \wedge (\beta \vee \gamma) \rightarrow ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma))$
A5. $((\alpha \rightarrow \beta) \wedge (\alpha \rightarrow \gamma)) \rightarrow (\alpha \rightarrow (\beta \wedge \gamma))$
A6. $((\alpha \rightarrow \gamma) \wedge (\beta \rightarrow \gamma)) \rightarrow ((\alpha \vee \beta) \rightarrow \gamma)$
A7. $\neg\neg\alpha \rightarrow \alpha$

[93] According to this definition FDE is not a relevant logic, since it has no conditional connective. However, if we add a conditional connective, subject to the constraint that $\models \alpha \rightarrow \beta$ iff $\alpha \models_{FDE} \beta$, it is. This is how the system first arose.
[94] With the exception of FDE as understood in the previous footnote.
[95] It was rediscovered by Došen (1992).

R1. $\alpha, \alpha \to \beta \vdash \beta$
R2. $\alpha, \beta \vdash \alpha \wedge \beta$
R3. $\alpha \to \beta \vdash (\gamma \to \alpha) \to (\gamma \to \beta)$
R4. $\alpha \to \beta \vdash (\beta \to \gamma) \to (\alpha \to \gamma)$
R5. $\alpha \to \neg\beta \vdash \beta \to \neg\alpha$

The logic R can be obtained by adding the axioms:

A8. $(\alpha \to \beta) \to ((\beta \to \gamma) \to (\alpha \to \gamma))$
A9. $\alpha \to ((\alpha \to \beta) \to \beta)$
A10. $(\alpha \to (\alpha \to \beta)) \to (\alpha \to \beta)$
A11. $(\alpha \to \neg\beta) \to (\beta \to \neg\alpha)$

(and dropping R3-R5, which are now redundant).[96] FDE is, it turns out, the core of all the relevant systems, in that if α and β contain no occurrences of $\to$ then $\alpha \models_{FDE} \beta$ iff $\alpha \to \beta$ is provable (in no matter which of the above-mentioned systems). Like FDE, B has no logical truths expressible in terms of only $\wedge$, $\vee$, and $\neg$. In R, however, $\alpha \vee \neg\alpha$ is a logical truth, as, in fact, are all classical tautologies.

The axiom systems, by themselves, are not terribly illuminating. An important problem then became to find appropriate semantics. The first semantics, produced by Dunn, was an algebraic one. Define a *De Morgan monoid* to be a structure $\langle L, \wedge, \vee, \neg, \to, \circ, e\rangle$. Where $\langle L, \wedge, \vee, \neg\rangle$ is a de Morgan lattice and $\to$ is a binary operator (representing the conditional). It is convenient to extract the properties of the conditional from a corresponding residuation operator (a sort of intensional conjunction); this is what $\circ$ is. e is a distinguished member of L; it's presence is necessary since we need to define logical truth, and this cannot be done in terms of the top member of the lattice (as in the algebraic semantics for classical and intuitionist logics), since there may be none. The logical truths are those which are always at least as great as e. In a De Morgan monoid, the additional algebraic machinery must satisfy the conditions:

$e \circ a = a$
$a \circ b \leq c$ iff $a \leq b \to c$
If $a \leq b$ then $a \circ c \leq b \circ c$ and $c \circ a \leq c \circ b$
$a \circ (b \vee c) = (a \circ b) \vee (a \circ c)$ and $(b \vee c) \circ a = (b \circ a) \vee (c \circ a)$

Note that $e \leq a \to b$ iff $e \circ a \leq b$ iff $a \leq b$, so conditionals may be thought to express containment of propositional content.

[96] E is obtained from R by deleting A9 and adding:

$(\alpha \to \neg\alpha) \to \neg\alpha$
$(\alpha \to \gamma) \to (((\alpha \to \gamma) \to \beta) \to \beta)$
$(N(\alpha) \wedge N(\beta)) \to N(\alpha \wedge \beta)$

where $N(\gamma)$ is $(\gamma \to \gamma) \to \gamma$. E is a much clumsier system than R. Initially, Anderson and Belnap thought that the $\to$ of E was exactly the modalised $\to$ of R. That is, they believed that if one adds an appropriate modal operator, $\Box$, to R, then $\Box(\alpha \to \beta)$ behaves in R, just like $\alpha \to \beta$ behaves in E. They even stated that should this not turn out to be the case, they would prefer the modalised version of R. It turned out not to be the case.

Finally, define:

$\Sigma \models \alpha$ iff for all evaluations into all De Morgan monoids, if $e \leq \nu(\beta)$ for all $\beta \in \Sigma$, $e \leq \nu(\alpha)$

This consequence relation is exactly one for B.

Stronger relevant logics can be obtained by putting further constraints on $\circ$. In particular, the logic R is produced by adding the following constraints:

$\circ$8 $a \circ (b \circ c) = (a \circ b) \circ c$

$\circ$9 $a \circ b = b \circ a$

$\circ$10 $a \leq a \circ a$

$\circ$11 $a \circ b \leq c$ iff $a \circ \neg c \leq \neg b$

$\circ$8-$\circ$11 correspond to A8-A11, respectively, in the sense that the structures obtained by adding any one of them are sound and complete with respect to the axiom system obtained by adding the corresponding axiom to B.[97]

Perhaps the most robust semantics for relevant logics are world semantics. These were produced by the Australian logician R.Routley (later Sylvan), in conjunction with Meyer, who moved to Australia, in the early 1970s.[98] The results of some 20 years of collaboration between Routley, Meyer, and their students, especially Brady, are published in Routley, Plumwood, Meyer and Brady (1984) and Brady (2003).

Historically, the world semantics piggy-backed upon yet another semantics for FDE produced by Sylvan and V.Routley (later Plumwood).[99] An interpretation for the language of FDE is a structure $\langle W, *, \nu \rangle$, where W is a set of worlds, and ν is a function that assigns every propositional parameter a truth value (0 or 1) at every world. Thus, for all $w \in W$, $\nu_w(p) = 0$ or $\nu_w(p) = 1$. The novel element here is $*$. This is a function from worlds to worlds, satisfying the condition: $w^{**} = w$. w^* is often glossed as the "mirror image" world of w; but its philosophical understanding is still a matter of some debate.[100] The truth conditions for the connectives are:

$\nu_w(\alpha \wedge \beta) = 1$ iff $\nu_w(\alpha) = 1$ and $\nu_w(\beta) = 1$

$\nu_w(\alpha \vee \beta) = 1$ iff $\nu_w(\alpha) = 1$ or $\nu_w(\beta) = 1$

$\nu_w(\neg\alpha) = 1$ iff $\nu_{w^*}(\alpha) = 0$

[97] Dunn worked out the details for R. It was Meyer who worked out the details for B and the logics between B and R. See Meyer and Routley (1972).

[98] Related ideas were published by Urquhart (1972) and by Fine (1974).

[99] See Routley and Routley (1972).

[100] For what is, I think, the most coherent story, see Restall (1999).

Thus, in the case where $w^* = w$, the truth conditions for $\neg$ collapse into the standard ones of modal logic. Validity is defined in terms of truth-preservation at all worlds of all interpretations. Again, it is not entirely obvious that these semantics deliver FDE, but it is not difficult to establish this. Essentially, it is because a relational evaluation, ρ, and a pair of worlds, w, w^*, are equivalent if they are related by the conditions:

$$\nu_w(\alpha) = 1 \text{ iff } \alpha\rho 1$$

$$\nu_{w^*}(\alpha) = 0 \text{ iff } \alpha\rho 0$$

Thus, a counter-model to Explosion is provided by the interpretation with two worlds, w, w^*, such that p is true at w and false at w^* (so that $\neg p$ is true at w); but q is false at w.

We can build an account of the conditional on top of this machinery as one would in a standard modal logic. Thus, $\alpha \multimap \beta$ is true at world w iff at every (accessible) world either α is false or β is true. The behavior of $*$ suffices to ensure that neither $\alpha \multimap (\beta \vee \neg\beta)$ nor $(\alpha \wedge \neg\alpha) \multimap \beta$ is valid. But the logic is not a relevant logic. The trouble is, for example, that $q \multimap q$ is true at all worlds. Hence $p \multimap (q \multimap q)$ is also true at all worlds, and so logically valid. To finish the job of producing the semantics for a relevant logic, we therefore need further machinery.

In Routley/Meyer semantics, a new class of worlds is introduced.[101] The worlds we have employed so far may be called *normal* worlds. The new worlds are *non-normal* worlds. Non-normal worlds are logically impossible worlds, in the sense that in these worlds the laws of logic may be different from what they are at possible (normal) worlds—just as the laws of physics may be different at physically impossible worlds. In particular, if one thinks of conditionals as expressing the laws of logic—so that, for example $\alpha \to \alpha$ expresses the fact that α follows from α—then non-normal worlds are worlds where logically valid conditionals (like $\alpha \to \alpha$) may fail. Thus $p \to (q \to q)$ will not be logically valid, since there are worlds where p is true, but $q \to q$ is false.

Specifically, an interpretation is a structure $\langle W, N, *, R, \nu\rangle$. W, $*$, and ν are as before. N is a subset of W, and is the class of normal worlds, so $W - N$ is the class of non-normal worlds. The truth conditions for $\wedge$, $\vee$, and $\neg$ are as before.[102] At normal worlds, w:

$$\nu_w(\alpha \to \beta) = 1 \text{ iff for all } w' \in W, \text{ either } \nu_{w'}(\alpha) = 0 \text{ or } \nu_{w'}(\beta) = 1$$

These are the simple $S5$ truth conditions for $\multimap$. To state the truth conditions for $\alpha \to \beta$ at non-normal worlds we require the relation R. This is an arbitrary relation on worlds; but unlike the binary accessibility relation of standard modal logic, it is a *ternary* relation. Thus, for all $w \in W - N$:

[101] The following are not quite the original Routley/Meyer semantics, but are a simplified form due to Priest and Sylvan (1992) and Restall (1993).

[102] It is possible to perform exactly the same construction concerning conditionals, but imposed not on $*$ semantics for negation, but on the Dunn four-valued semantics. The result is a family of perfectly good relevant logics, but not the Anderson Belnap family under consideration here.

$\nu_w(\alpha \to \beta) = 1$ iff for all $x, y \in W$ such that $Rwxy$, either $\nu_x(\alpha) = 0$ or $\nu_y(\beta) = 1$

Given these truth conditions, it is clear that a conditional such as $q \to q$ may fail at a non-normal world, w, since we may have $Rwxy$, with q true at x, but false at y. In this way, relevance is obtained. Note that if $x = y$ the truth conditions for $\to$ at non-normal worlds collapse into the $S5$ truth conditions. Hence, we may state the truth conditions for $\to$ at all worlds uniformly in terms of the ternary relation, provided that at normal worlds we *define* R in terms of identity. That is, for normal worlds, w:

$Rwxy$ iff $x = y$

Validity is defined in terms of truth preservation at *normal* worlds. Thus:

$\Sigma \models \alpha$ iff for every interpretation and every $w \in N$, if $\nu_w(\beta) = 1$ for all $\beta \in \Sigma$, $\nu_w(\alpha) = 1$

These semantics are a semantics for the relevant logic B. Stronger relevant logics may be produced by adding constraints on the ternary relation R. For example, the relevant logic R is produced by adding the following constraints. For all $x, y, z, u, v \in W$:

R8. If $\exists w(Rxyw$ and $Rwuv)$ then $\exists w(Rxuw$ and $Rywv)$
R9. If $Rxyz$ then $Ryxz$
R10. If $Rxyz$ then $\exists w(Rxyw$ and $Rwyz)$
R11. If $Rxyz$ then Rxy^*z^*

Each of these constraints corresponds to one of A8-A11, in the sense that the axiom is sound and complete with respect to the class of interpretations in which the corresponding constraint is in force.

An important issue to be faced is what, exactly, the ternary relation means, and why it should be employed in stating the truth conditions of conditionals. Whether there are sensible answers to these questions, and, if so, what they are, is still a matter for debate. Some, for example, have tried to explicate the notion in terms of the flow of information.[103] It is worth noting that the ternary relation can be avoided if one simply assigns conditionals arbitrary truth values at non-normal worlds—which makes perfectly good sense, since at logically impossible worlds, logical principles could, presumably, do anything. This construction gives a relevant logic weaker than B.[104]

At any rate, the relevant logic B is the analogue of the modal logic K, in the following sense. K is the basic (normal) modal logic. In its semantics, the binary accessibility relation is arbitrary. Stronger logics are obtained by adding constraints on the relation. Similarly, B is the basic relevant logic (of this family). In its semantics, the ternary accessibility relation is arbitrary. Stronger logics are obtained by adding constraints on the relation. It was this fact that became clear

[103] For further details, see Priest (2001a), 10.6.
[104] See Priest (2001a), ch.9.

with the invention of the world-semantics for relevant logics by the Australian logicians. Moreover, just as the early work on modal logic had concentrated on systems at the strong end of the modal family, so Anderson and Belnap's work had concentrated on systems at the strong end of the relevant family.[105] Further details concerning relevant logic can be found in the chapter on the subject in this *Handbook*, so we will pursue the issue no further here.

We have now looked at the development of paraconsistent logics in the modern period, based on four distinct ideas. This survey is certainly not exhaustive: there are other approaches.[106] But we have tracked the major developments, and it is now time to return to dialetheism.

5 MODERN DIALETHEISM

5.1 Inconsistent Information

As we noted in 1.2, the major motive for modern paraconsistency is the idea that there are situations in which we need to reason in a non-trivial way from inconsistent information. The early proponents of paraconsistent logics mentioned various such situations, but the first sustained discussion of the issue (that I am aware of) is Priest and Routley (1989).[107] A list of the situations involving inconsistent

[105] A word on terminology. The Americans called the subject *relevance* logic, since they took the logic to be spelling out what relevance was. This was rejected by Sylvan, who argued that the logics did not provide an analysis of relevance as such. The logics were relevant, but this fact fell out of something more fundamental, namely, truth preservation over a suitably rich class of (and especially impossible) worlds. Following Sylvan, Australian logicians have called the logics *relevant* logics.

[106] A quite different approach goes back to research starting in the late 1950s. This also has relevance connections. It is a natural idea that classical logical consequence lets in too much, and specifically, that it lets in inferences where the premises and conclusion have no connection with each other. The thought then is to filter out the irrelevant inferences by imposing an extra condition. Specifically, define the inference from α to β to be *prevalid* if $\alpha \models_C \beta$ and $F(\alpha, \beta)$. Prevalid inferences may not be closed under substitution. So define an inference to be valid if it is obtained from a prevalid inference by uniform substitution. The condition F is a filter that removes the Bad Guys. A suitable choice of F gives a paraconsistent logic. The first filter logic was given by Smiley (1959). His filter was the condition that α not be a classical contradiction and β not be a classical tautology. It is clear that this makes the inference $p \wedge \neg p \vdash q$ invalid. It is also easy to check that the following inferences are valid under the filter: $p \wedge \neg p \vdash p \wedge (\neg p \vee q)$, $p \wedge (\neg p \vee q) \vdash q$. (The first is a substitution instance of $p \wedge r \vdash p \wedge (r \vee q)$.) This shows two things: first, that the disjunctive syllogism holds, unlike in most other paraconsistent—and particularly relevant—logics; second, that the transitivity of deducibility breaks down. The failure of transitivity is, in fact, typical of filter logics (though not invariably so). Perhaps the most interesting filter logic was developed by Tennant (1984), a student of Smiley. It is given most naturally in multiple-conclusion terms. (Thus, $\Sigma \models_C \Pi$ iff every classical evaluation that makes every member of Σ true makes some member of Π true.) Accordingly, $\Sigma \models \Pi$ iff $\Sigma \models_C \Pi$ and there are no proper subsets $\Sigma' \subset \Sigma$, $\Pi' \subset \Pi$, such that $\Sigma' \models_C \Pi'$. The filter takes out redundant "noise". Suitably developed, this approach can be used to construct a family of relevant but non-transitive logics. See Tennant (1992).

[107] The essay, which can be consulted for further discussion of the material that follows, is one of the introductory chapters of Priest, Routley, and Norman (1989). This was the first collection

information that have been mooted include:

1. Information collected from different sources, at different times, etc., especially in computational information processing.

2. Various theories in science and mathematics.

3. Various theories in philosophy.

4. Various bodies of law and other legal documents.

5. Descriptions of borderline cases concerning vague predicates.

6. Descriptions of certain states of change.

7. Information concerning over-determination and multi-criterial terms.

8. Information generated by paradoxes of self-reference.

Of these, the most straightforward is 1.[108] Information collected in this way is clearly liable to be inconsistent. The situation is particularly crucial in modern information processing, where the amount of information is humanly unsurveyable. Whilst, no doubt, one would normally wish to revise inconsistent information when it occurs in this context, we might be in a situation in which we do not know how to revise consistently. Worse, as is well known, there is no algorithm for inconsistency, so we may not even know that the information is inconsistent.

For 2, it is a fact that various theories in the history of science have been inconsistent, and known to be so. Perhaps the most striking example of this is the Bohr theory of the atom, whose inconsistency was well recognised—even by Bohr. To explain the frequency of radiation emitted in quantum transitions, classical electromagnetic theory had to be employed. But the same electromagnetic theory contradicts the existence of stationary states for an electron in orbit; it entails that such electrons, since they are accelerating, will radiate (and so lose) energy.[109]

An example of an inconsistent theory in the history of mathematics is the original calculus of Newton and Leibniz. Again, the inconsistency of this was well

---

of essays on paraconsistency, and contains essays by most of the founders of the subject. It may be noted that the completed manuscript of the book was sent to the publisher in 1982, which is a more accurate dating of its contents. The book contains a useful bibliography of paraconsistecy to that date.

[108] A supposed example of this that is sometimes cited is the information provided by witnesses at a a trial, who frequently contradict one another—and themselves. This example, though, is not very persuasive. For, plausibly, the pertinent information in this sort of case is not of the form 'the car was red', 'the car was not red', but of the form 'witness x says that the car was red', 'witness y says that the car was not red'. (The judge and jury may or may not conclude something about the colour of the car.) Information of this kind is consistent.

[109] The Bohr theory has long since been displaced by modern quantum theory. But this, too, sails close to the paraconsistent wind in a number of places. To mention just one: the Dirac δ-function has mathematically impossible properties. The integral of the function is non-zero; yet its value at all but one point is zero.

known at the time. It was pointed out forcibly by Berkeley, for example. In computing derivatives one needed to divide by infinitesimals, at one stage, and so suppose them to be non-zero. In the final stage of the computation, however, one had to ignore infinitesimal summands, hence assuming, in effect, that they are zero.[110] We will return to the issue of inconsistent mathematical theories later.

Turning to 3, the examples of inconsistent theories in the history of philosophy are legion. Indeed, most philosophers who have constructed theories of any degree of complexity have endorsed principles that turned out to be contradictory. No doubt, many of these philosophers contradicted themselves unwittingly. However, in Section 2 above, we noted various philosophers for whom this was not the case: Heraclitus, Hegel, and Meinong (at least, as many people interpreted him). Again, we will return to inconsistent philosophical theories later.

We will also come to the other cases on the list above in a minute. But given even just these cases, it is clear that inferences must be, or were, drawn from inconsistent information. What inference mechanism was employed in each of the historical cases is a matter for detailed historical investigation. There is no *a priori* reason to suppose that it was one of the formal paraconsistent logics we looked at in the last section—though there is no *a priori* reason to suppose that it was not, either. What is ungainsayable is that in all these cases, where inference goes ahead in contexts whose inconsistency—or the possibility thereof—is explicitly acknowledged, some inference procedure that is *de facto* paraconsistent must (have) be(en) employed.

5.2 The Rise of Modern Dialetheism

In none of the cases so far discussed is there much temptation to suppose that the inconsistent information in question is true, that is, that we have an example of dialetheism—unless one endorses one of the philosophical theories mentioned, such as Meinongianism. Even in the cases of inconsistent theories in science and mathematics, we may suppose that the theories were important, not because they were taken to be true, but because they were useful instrumentally, or perhaps they were taken to be good *approximations* to the (consistent) truth.

In fact, none of the paraconsistent logicians mentioned in the previous section who wrote before the 1970s, with the exception of Asenjo, comes close to endorsing dialetheism.[111] Indeed, it is clear that some of the formal paraconsistent logics of the last section do not even lend themselves to dialetheism. Non-adjunctive logics, in particular, though they concern the aggregation of information that is, collectively, inconsistent, have no truck with the idea that the information from any one source is inconsistent. To bring this home, note that for each of the non-adjunctive constructions, one can formulate explicitly dialetheic versions.

[110] For an analysis of this, and many other inconsistent mathematical theories, see Mortensen (1995).

[111] This is true even of da Costa, who was much concerned with inconsistent set-theories. He tended to regard these simply as interesting and possibly important mathematical theories.

For example, consider discussive logic. Repeat the construction, but based not on a classical modal logic, but on a paraconsistent modal logic that allows for inconsistent worlds (for example, of the kind in the world-semantics of relevant logic). Or in the Rescher/Manor construction, instead of considering maximal consistent sets, consider maximal non-trivial sets, and then apply a paraconsistent consequence relation to these. How to handle pieces of information from multiple sources, which do not fit together happily, is a problem for everyone, dialetheist or otherwise.

The rise of the modern dialetheist movement can most naturally be seen as starting in the 1970s with the collaboration between Priest and Routley in Australia.[112] Priest argued for dialetheism in (1973) in an argument based on paradoxes of self-reference and Gödel's Theorem. The case was mounted in detail in a paper, later published as (1979), given at a meeting of the Australasian Association for Logic in Canberra in 1976, where Priest and Routley first met. Priest (1987) is a sustained defence of dialetheism. Routley became sympathetic to dialetheism because of his work on the semantics of relevant logics, and the possibility of applying relevant logic to logical paradoxes and to Meinong. He endorsed the position in (1977) and (1979).[113]

It is worth noting that it was the development of the world-semantics for relevant logic which brought the dialetheic potential of relevant logic to the fore. If there are inconsistent worlds, a person of a naturally curious disposition will ask how one knows that the actual world is not one of them. The American relevant logicians never showed any tendency towards dialetheism. Even Dunn, who was responsible for the four-valued semantics, preferred to read 1 and 0 as 'told true' and 'told false', rather than as 'true' and 'false': inconsistent information could be given, but not the truth. Endorsing the world-semantics for relevant logic does not require dialetheism, however. It is quite possible to suppose that all the inconsistent worlds are non-normal, that is, that for all $w \in W - N$, $w = w^*$. The logic will still be relevant, but will validate Explosion, and so not be paraconsistent. Alternatively, one may suppose that some normal worlds are inconsistent, so that the logic is paraconsistent, but that the actual world has special properties; in particular, consistency.

5.3 Arguments for Dialetheism

Let us now return to the list of examples in 5.1. The rest of the examples on the list have been mooted as dialetheias. Let us start with 4. It is not uncommon for legal documents to have unforeseen consequences; sometimes, these can be

[112] Readers must remember, especially at this point, that this essay is not being written by an impartial historian, and make due allowances for this.

[113] In this paper Routley describes his position as 'dialectical', taking the view to be identical with aspects of dialectical logic in the Hegel/Marx tradition. Whilst there certainly are connections here, the simple identification is, at the very least, somewhat misleading, and Routley dropped the description after the term 'dialetheism' was coined.

contradictory. Suppose, for example, that the constitution of a certain country contains the clauses:

All university graduates are required to perform jury service.

No woman shall be a member of a jury.

We may suppose that when the constitution was written, university admission was restricted to male clergy, as it had been for hundreds of years. Some time later, however, universities open their doors to women. Women graduates are then both required to perform and forbidden from performing jury service.[114] Of course, once the contradiction came to light, the constitution would presumably be changed, or a judge would rule one way or the other (which is tantamount to the same thing). But until and unless this is done, we have a legal contradiction.

The law has a number of mechanisms for defusing *prima facie* contradictions. For example it is a general principle that constitutional law outranks statute law, and that a later law overrides an earlier law. Clearly, such principles may well resolve an explicit contradiction in legislation. However, equally clearly, the situation may be such that none of the principles applies. (The situation just described might be one of these.) And where this is the case, the contradictions are not just *prima facie*.

Turning to 5, the idea is this. Given a vague predicate, there is a grey area between cases in which it clearly applies and cases where it clearly does not. Thus, there is no point at which a tadpole ceases to be a tadpole and becomes a frog. Suppose that Fred is a creature in this grey area. Intuition says that Fred is as much tadpole as not tadpole, and as little tadpole as not tadpole. In other words, the semantic value of 'Fred is a tadpole' is symmetrically poised between truth and falsity. It is commonplace to suppose that a sentence such as this is neither true nor false. But as far as the story so far goes, *both true and false* is just as good. Moreover, for any consideration that drives one towards truth value gaps, there would seem to be dual considerations that drive one towards truth value gluts.[115]

To be honest, any simple three-valued solution to the problem of vagueness is going to be problematic for very simple reasons. Just as the boundary between being true and being false is grey in such cases, so the boundary between being true and being neither true nor false, or being both true and false, is also grey. Little therefore seems to have been gained by moving to three semantic values. Considerations of this kind have led some logicians to endorse a continuum-valued semantics to deal with vagueness. Assuming, as is standard, that such values are numbers in the range $[0,1]$, and that if the value of α is x, the value of $\neg\alpha$ is $1-x$, then a contradiction $\alpha\wedge\neg\alpha$ may certainly be half-true—and 0.5 may be a designated value in the context.

[114] In a similar way, the rules of a game, such as chess, may well have untoward consequences, such as a contradiction in certain recondite situations that come to light.

[115] See Hyde (1997).

In some ways, issues are similar when we move to 6. Consider a state of affairs described by α, which changes, perhaps instantaneously, to one described by $\neg\alpha$. It may be that there is something about the point of transition that determines either α or $\neg\alpha$ as true at that transition. Thus, for example, if a car accelerates continuously from rest, there is a last point with zero velocity, but no first point with a non-zero velocity. But, again, it may be that the situation is completely symmetrical. Thus, if a subatomic particle makes an instantaneous transition from one quantum state to another, there are no continuity considerations to determine the situation at the point of transition one way or the other. In such situations, the transition state is symmetrically poised between α and $\neg\alpha$. Either, then, neither α nor $\neg\alpha$ is true, or both are. Moreover, in this case, there are some considerations, at least, which push towards the latter conclusion. The state, whatever it is, is a state of change. Such a state is naturally described as one in which $\alpha \wedge \neg\alpha$ holds. (Recall Heraclitus.) A state where neither α nor $\neg\alpha$ holds is less naturally thought of as a state of change. For if neither holds, then α has ceased to be true. That change is already *over*. It is true that if $\alpha \wedge \neg\alpha$ holds then α still holds, so its ceasing is *yet* to occur. But in this case, at least $\neg\alpha$ has already started: change is under way. Or to put it another way: an instant where neither α nor $\neg\alpha$ holds cannot be a transition state between one where α holds and one where $\neg\alpha$ holds. For it is quite possible that such a state might be followed by ones where $\neg\alpha$ does not hold: $\neg\alpha$ never starts at all!

The idea can be applied to one of Zeno's paradoxes of motion: the arrow. Recall that this goes as follows. Consider an arrow at an instant of its motion. During that instant it advances not at all on its journey. Yet somehow in the whole motion, composed of just such instants, it does advance. How can this be possible? Standard measure-theory tells us that an interval of non-zero measure is composed of points of zero measure. Fine. But how can a physical advance be constituted by a bunch of no advances? A bunch of nothings, even an infinite bunch, is nothing. A resolution is provided by the previous considerations concerning change. At an instant of the motion, the arrow is at point p. But it is in a state of change, so it is not there as well. Thus, it is also at other points; presumably those just before and just after p. In the instant, then, it does occupy more than one point; it does make some advance on its journey.

Finally in this section, let us consider 7. It is a commonplace to note that versions of verificationism may give rise to truth-value gaps since, for certain α, neither α nor $\neg\alpha$ may be verified—or even verifiable. It is less often noted that other versions may give rise to truth value gluts. Specifically, it is not uncommon for terms of our language to be multi-criterial—that is, for there to be different criteria which are semantically sufficient for the application of the term. For example, the appropriate reading from a correctly functioning alcohol thermometer is sufficient to determine the temperature of some water to be $4°c$. But the appropriate reading of a thermo-electric thermometer is equally sufficient for the same. Now, normally, if we test for both of these criteria, they will either both hold or both fail. But in circumstances of a novel kind, it might well happen that the criteria fall apart. The alcohol thermometer may tell us that the temperature is 4°; the thermo-electric thermometer may tell us that it is 3°, and so not 4°.

It might be argued that if such a situation occurs, what this shows is that the terms in question are ambiguous, so that '3°' is ambiguous between 3°-*by-an-alcohol-thermometer*, and 3°-*by-an-electro-chemical-thermometer*. And doubtless, should this situation arise, we probably would replace our old concept of temperature by two new concepts. In just this way, for example, the term 'mass', as employed before the Special Theory of Relativity, was replaced by two terms 'rest mass' and 'inertial mass', afterwards. But it can hardly be claimed that the old term was semantically *ambiguous* before, in the way that, say, 'cricket' is (the insect and the game). It had a single meaning; we just recognised that meaning as applicable in different, and logically independent, ways. Thus, the situation, as described in the old language, really was inconsistent.

5.4 Truth and the Paradoxes of Self-Reference

This brings us to the last item on the list: the paradoxes of self-reference. As a matter of documented fact, this is the consideration that has been historically most influential for dialetheism. It is also, I think it fair to say, the consideration to which it is hardest to object coherently. Paradoxes of this kind are apparently valid arguments, often very simple arguments, starting from things that seem obviously true, but ending in explicit contradictions. Unless one can fault them, they establish dialetheism. Though many arguments in the family are, historically, quite recent, paradoxes of the family have been known now for close to two and a half thousand years. It is a mark of their resilience that even now there is still no consensus amongst those who think that there is something wrong with them as to what this is. Better, then, to stop trying to find a fault where none exists, and accept the arguments at face value.

It is conventional wisdom to divide the paradoxes into semantic and set-theoretic. Though I think that this a profoundly misleading distinction,[116] it will be useful to employ it here. Let us start with the semantic paradoxes. These are paradoxes that concern notions such as truth, satisfaction, reference. Take everyone's favourite: the liar paradox.[117] At its simplest, this is the claim: this claim is false. If it is true then it is false; and if it is false then it is true. Contradiction in either case. To tighten up the argument, let us write T for 'is true'. Then the liar is a truth-bearer,[118] λ, of the form $\neg T\langle\lambda\rangle$. (The angle brackets here are some name-forming device.) Now, an almost irresistible principle concerning truth, stated first by Aristotle, is that something is true iff what it claims to be the case is in fact the case; as it is usually called now, the T-schema. For every α:

---

[116]See Priest (1995), Part 3.

[117]It should be noted that though the paradox is a paradigm of the family, it has features that other members of the family do not have, and vice versa. One can not simply *assume*, therefore, that a solution to it automatically generalises to all members of the family.

[118]One can choose whether these are sentences, propositions, beliefs or wot not, as one pleases.

$$T\langle\alpha\rangle \leftrightarrow \alpha$$

In particular, $T\langle\lambda\rangle \leftrightarrow \lambda$. And given what λ is:

$$T\langle\lambda\rangle \leftrightarrow \neg T\langle\lambda\rangle.$$

$T\langle\lambda\rangle \wedge \neg T\langle\lambda\rangle$ now follows, given various logical principles, such as the law of excluded middle, or *consequentia mirabilis* ($\alpha \rightarrow \neg\alpha \vdash \neg\alpha$).

The solutions to the liar and other semantic paradoxes that have been suggested—particularly in the last 100 years—are legion. This is not the place to attempt an exhaustive analysis of them. Further details can be found in the article on the paradoxes of self-reference in this *Handbook*. However, all attempts to provide a consistent analysis of the paradoxes run into fundamental problems. To see this, let us start by considering what are probably the two most influential such attempts in the last 100 years.

The first of these is based on the work of Tarski. According to this, a language may not contain its own truth predicate. That is, a predicate satisfying the T-schema for every sentence of a language L, must not occur in L itself, but must occur in a metalanguage. Of course, the move must be repeated, generating a whole hierarchy of languages, H, each of which contains a truth predicate for lower members of the hierarchy, but is semantically open: it does not contain its own truth predicate. In no sentence of the hierarchy may we therefore formulate a self-referential liar sentence.

Of the many objections that one may raise against this solution, note here only the following. Given the resources of H, one may formulate the sentence:

λ_H: λ_H is true in no member of H.

Now we have a choice: is λ_H a sentence of some language in H or not? Suppose it is. We may therefore reason about its truth in the next member of the hierarchy up. If it is true, then it is not true in any member of H. Contradiction. Hence it cannot be true in any member of the hierarchy. That is, we have established λ_H. Hence, λ_H is a true sentence of some language in H. And we have already seen that this leads to contradiction. Suppose, on the other hand, that λ_H is not a member of the hierarchy. Then H is not English, since λ_H clearly is a sentence of English. The construction does not, therefore, show that the rules governing the truth predicate in English are consistent.[119]

The other particularly influential theory is Kripke's. According to this, certain sentences may fail to take a truth value, and so be neither true nor false. Starting with a language which contains no truth predicate, we may augment the language with one, and construct a hierarchy. Not, this time, a hierarchy of languages, but a hierarchy of three-valued interpretations for the extended language. At the base level, every sentence containing T is neither true nor false. As we ascend the

[119]Here, and in what follows, I am assuming that English is the language of our vernacular discourse. Exacly the same considerations apply if it is some other natural language.

hierarchy, we acquire information to render sentences containing T determinately true or false. In particular, if we have shown that α is true at a certain level of the hierarchy, this suffices to render $T\langle\alpha\rangle$ true at the next. If we play our cards right, we reach a level, F (a *fixed point*), where everything stabilises; by then, every sentence has a fixed semantic status; in particular, for every α, α and $T\langle\alpha\rangle$ have the same status. It is this fixed-point interpretation that is supposed to provide an account of the behaviour of the truth predicate. Sentences that are determinately true or determinately false at the fixed point are called *grounded*. The liar sentence is, unsurprisingly, ungrounded. And being neither true nor false, it slips through the dilemma posed by the liar paradox argument.

Again, of the many objections that may be brought against the theory, we note just one. Consider the sentence:

λ_F: λ_F is not true at F

What status does λ_F have at F? If it has the status *true*, then it is not true at F. Contradiction. If it does not have the status *true* (in particular, if it is neither true nor false), then what it says to be the case *is* the case. Hence it is true. Contradiction again. One may object by noting that if λ_F is neither true nor false at F, then so are $T\langle\lambda_F\rangle$ and $\neg T\langle\lambda_F\rangle$. Hence the final step of the reasoning does not follow. But if one chooses to break the argument in this fashion, this just shows, again, that the behaviour of T at the fixed point is not that of the English truth predicate. For according to the theory, λ_F is *not* true at the fixed point; and the theorist is committed to the truth of this claim. At this point, the only option,[120] is to locate the discourse of the theorist *outside* the language L—in effect, taking the theorist's truth predicate to be in a metalanguage for L. But this just shows that the construction does not establish the truth predicate of English to behave consistently. For the theorist is speaking English, and the construction does not apply to that.

If we look at these two solutions, we can see a certain pattern. The machinery of the solution allows us to reformulate the liar paradox. Such reformulations are often call *extended paradoxes*. This is something of a misnomer, however. These paradoxes are not new paradoxes; they are just the same old paradox in a new theoretical context. What generates the paradox is a heuristic that allows us to construct a sentence that says of itself that it is not in the set of *bona fide* truths. Different solutions just characterise this set in different ways. At any rate, the only options in the face of these reformulated paradoxes are to accept contradiction or to deny that the machinery of the solution is expressible in the language in question. Since the machinery is part of the discourse of the theoretician, English, this shows that English discourse about truth has not been shown to be consistent.

The pattern we see here manifests itself, in fact, across all purported solutions to the liar paradox, showing them all to be deeply unsatisfactory for exactly the same reason.[121] Neither is this an accident. There are underlying reasons as to why it

[120]Which Kripke, in fact, exercised.
[121]For detailed arguments, See Priest (1987), ch.1, and Priest (1995), Part 3.

must happen. We can put the matter in the form of a series of dilemmas. The liar and its kind arise, in the first place, as arguments in English. One who would solve the paradoxes must show that the semantic concepts of English involved are not, despite appearances, inconsistent—and it is necessary to show this for all such concepts, for they are all embroiled in contradiction. Attempts to do this employing the resources of modern logic all show how, for a given language, L, in some class of languages, to construct a theory T_L, of the semantic notions of L, according to which they behave consistently.

The first dilemma is posed by asking the question of whether T_L is expressible in L. If the answer is 'yes', the liar heuristic always allows us to reformulate the paradox to generate inconsistency. Nor is this an accident. For since T_L is expressible in L, and since, according to T_L, things are consistent, we should be able to prove the consistency of T_L in T_L. And provided that T_L is strong enough in other ways (for example, provided that it contains the resources of arithmetic, which it must if L is to be a candidate for English), then we know that T_L is liable to be inconsistent by Gödel's second incompleteness Theorem. (Any theory of the appropriate kind that can prove its own consistency is inconsistent.)

If the answer to the original question in 'no', then we ask a second question: is English (or at least the relevant part of it), E, one of the languages in the family being considered? If the answer to this is 'yes', then it follows that T_E is not expressible in English, which is self-refuting, since the theorist has explained how to construct T_E in English. If, on the other hand, the answer to this question is 'no', then the original problem of showing that the semantic concepts of English are consistent has not been solved.

Hence, all attempts to solve the paradox swing uncomfortably between inconsistency and a self-refuting inexpressibility. The problem, at root, is that English is, in a certain sense, over-rich. The semantic rules that govern notions such as truth over-determine the truth values of some sentences, generating contradiction. The only way to avoid this is to dock this richness in some way. But doing this just produces incompleteness, making it the case that it is no longer English that we are talking about.[122]

What we have seen is that the liar paradox and its kind are more than just *prima facie* dialetheias. Attempts to show them to be only this, run into severe difficulties. At this point, a natural question is as follows: if consistent attempts to solve the paradoxes run into the problem of reformulated paradoxes, what about dialetheic solutions? In particular, if sentences may be both true and false, perhaps

[122] Another move is possible at this point: an explicitly revisionary one. This concedes that the rules that govern 'is true' in English generate contradictions, but insists that the concept should be replaced by one governed by rules which do not do this. This was, in fact, Tarski's own view, and was the spirit in which he offered the hierarchy of metalanguages. But why *must* we revise? If our notion of truth is inconsistent, does this just not show us that an inconsistent notion is perfectly serviceable? And if we must go in for some act of self-conscious conceptual revision, then a revision to a paraconsistent/dialetheic conceptual framework is clearly a possibility. The mere proposal of a consistent framework is not, therefore, enough: it must be shown to be superior. As we will see in the final part of this essay, this seems rather hard task.

the *bona fide* truths are the ones that are *just* true. So what about:

λ_D: λ_D is not (true only)

If it is true it is also false. If it is false, it is true only. Hence it is true. Hence, it would seem to be true and false. But if it is true, it is not false. Hence it is true, false, and not false. We have certainly run into contradiction. But unlike consistent accounts of the paradox, this is hardly fatal. For the very point of a dialetheic account of the paradoxes is not to show that self-referential discourse about truth is consistent—precisely the opposite. This is a confirmation, not a refutation!

There is an important issue here, however. Though some contradictions are acceptable to a dialetheist, not all are, unless the dialetheist is a trivialist. Now there is an argument which purports to show that the T -schema entails not just some contradictions; it entails everything. In particular, suppose that the conditional involved in the T-schema satisfies both *modus ponens* and Contraction: $\alpha \to (\alpha \to \beta) \vdash \alpha \to \beta$. Let α be any sentence, and consider the sentence:

λ_α: $T\langle\lambda_\alpha\rangle \to \alpha$

(if this sentence is true then α). The T-schema gives:

$$T\langle\lambda_\alpha\rangle \leftrightarrow (T\langle\lambda_\alpha\rangle \to \alpha)$$

whence Contraction from left to right gives:

$$T\langle\lambda_\alpha\rangle \to \alpha$$

whence *modus ponens* from right to left gives $T\langle\lambda_\alpha\rangle$. A final *modus ponens* delivers α. Arguments of this kind are usually called *Curry paradoxes*, after one of their inventors.

A dialetheic solution to the paradoxes therefore depends on endorsing a paraconsistent logic whose conditional does not satisfy Contraction.[123] Paraconsistent logics whose positive parts are classical or intuitionistic, such as the positive-plus logics of 4.3, contain Contraction, and so are unsuitable. Even the stronger relevant logics in the vicinity of R endorse Contraction. But weaker relevant logics, in the vicinity of B, do not. It can be shown that a theory containing the T-schema and self-reference (even all of arithmetic), and based on a weaker relevant logic, though inconsistent, is non-trivial. It can be shown, in fact, that all the sentences that are grounded in Kripke's sense (and so contain only extensional connectives) behave consistently.[124]

We have yet to deal with the set-theoretic paradoxes, but before we turn to these, let us return to the issue of inconsistencies in philosophical theories.

[123] Or *modus ponens*, though this is a less easy position to defend. It has been defended by Goodship (1996).

[124] The result was first proved for a version of set theory by Brady (1989). Its adaptation to truth is spelled out in Priest (2003), Section 8.

5.5 The Limits of Thought

A few philosophers have endorsed explicitly contradictory theories. Many have endorsed theories that turned out to be accidentally inconsistent—accidental in the sense that the inconsistencies could be trimmed without fundamental change. But there is a third group of philosophers. These are philosophers who, though they could hardly be said to be dialetheists, yet endorsed theories that were essentially inconsistent: inconsistency lay at the very heart of their theories; it could not be removed without entirely gutting them.

Such inconsistencies seem to occur, in particular, in the works of those philosophers who argue that there are limits to what can be thought, conceived, described. In the very act of theorising, they think, conceive, or describe things that lie beyond the limit. Thus, many philosophers have argued that God is so different from anything that people can conceive, that God is literally beyond conception or description. This has not prevented them from saying things about God, though; for example, in explaining why God is beyond conception.

A famous example of the same situation is provided by Kant in the first *Critique*. Kant espoused the distinction between phenomena (things that can be experienced) and noumena (things that cannot). Our categories of thought apply to the former (indeed, they are partly constitutive of them); but they cannot be applied to the latter (one reason for this: the criteria for applying each of the categories involves time, and noumena are not in time). In particular, then, one can say nothing about noumena, for to do so would be to apply categories to them. Yet Kant says much about noumena in the *Critique*; he explains, for example, why our categories cannot be applied to them.

Another famous example of the same situation is provided by Wittgenstein in the *Tractatus*. Propositions express the facts that constitute the world. They can do so because of a commonality of structure. But such structure is not the kind of thing that propositions can be about (for propositions are about objects, and structure is not an object). One can say nothing, therefore, about this structure. Yet the *Tractatus* is largely composed of propositions that describe this structure, and ground the conclusion that it cannot be described.

None of the philosophers referred to above was very happy about this contradictory situation; and all tried to suggest ways in which it might be avoided. In theology, it was not uncommon to draw a distinction between positive and negative attributions, and to claim that only negative assertions can be made of God (*via negativa*), not positive. But not only is the positive/negative distinction hard to sustain—to say, for example, that God is omnipotent is to say that God can do everything (positive); but it is equally to say that there is nothing that limits God's power (negative)—the very reasons for supposing that God is ineffable would clearly seem to be positive: ineffability arises because God's characteristics exceed any human ones by an infinite amount.

In the *Critique*, Kant tried to defuse the contradiction in a not dissimilar way, claiming that the notion of a noumenon had a merely negative, or limiting, function: it just serves to remind that there are bounds to the applicability of our categories. But this does not actually address the issue, which is how we can possibly say anything at all about noumena; indeed, it makes matters worse by saying *more* things about them. And again, Kant says lots of things about noumena which go well beyond a simple assertion of this limiting function; for example, he defends free will on the ground that the noumenal self is not subject to causation.

The issue was faced squarely in the *Tractatus*. Wittgenstein simply accepted that he could not *really* say anything about the structure of language or the world. The *Tractatus*, in particular, in mostly meaningless. But this is not at all satisfactory. Apart from the fact that we *do* understand what the propositions of the *Tractatus* say—and so they cannot be meaningless—if this were indeed so, we would have no ground for supposing that the propositions *are* meaningless, and so accepting Wittgenstein's conclusions. (You would not buy a second-hand ladder from such a person.)

None of the saving stratagems, then, is very successful. Nor is this surprising. For there is something inherently contradictory in the very project of theorising about limits of thought. In the very process, one is required to conceive or describe things that are on the other side—as Wittgenstein himself points out in the introduction to the *Tractatus*. The contradiction concerned is therefore at the very heart of the project. It is no mere accidental accretion to the theory, but is inherent in its very problematic. If there are limits to thought, they are contradictory—by their very nature.

Of course, one might reject the contradiction by rejecting the claim that there are things beyond the limit of thought. (This is exactly Berkeley's strategy in his argument that everything can be conceived.) There is no God; or if there is, God is perfectly effable. Hegel argued that our categories are just as applicable to noumena as they are to phenomena.[125] And in the introduction to the English version of the *Tractatus*, Russell argued that what could not be stated in the language of the *Tractatus* could be stated in a metalanguage for it.

How successful these *particular* moves are, is another matter. There certainly are *general* philosophical reasons for supposing there to be things beyond the limits of thought. The most definitive reasons for supposing this take us back to the semantical paradoxes of self-reference. There are so many objects that it impossible that all of them should have a name (or be referred to). There is, for example, an uncountable infinitude of ordinal numbers, but there is only a countable number of descriptions in English. Hence, there are many more ordinal numbers than can have names. In particular, to turn the screw, since the ordinal numbers are well-ordered, there is a least ordinal number that has no description. But we have just described it.

Perhaps, it may be thought, something fishy is going on here with infinity.

[125]This is ironical, to a certain extent, since Hegel was a philosopher who was prepared to accept contradictions. But in this respect, the move takes Hegel out of the frying pan, and into the fire. For the move undercuts Kant's solution to the Antinomies of Pure Reason, which contradictions must therefore be endorsed.

Historically, infinity has always, after all, been a notion with a question mark hanging over it. But similar paradoxes do not employ the notion of infinity. Given the syntactic resources of English, there is only a finite number of descriptions of some fixed length—say less than 100 words—and, *a fortiori*, only a finite number of (natural) numbers that are referred to by them. But the number of numbers exceeds any finite bound. Hence, there are numbers that cannot be referred to by a description with fewer than 100 worlds. And again, there must be a least. This cannot be referred to; but we have just referred to it.

These two paradoxes are well known. The first is König's paradox; the second is Berry's. They are semantic paradoxes of self-reference in the same family as the liar. We now see them in another light. They are paradoxes of the limits of thought; and contradiction is just what one should expect in such cases.[126]

6 THE FOUNDATIONS OF MATHEMATICS

6.1 Introduction: a Brief History

The development of modern logic has been intimately and inextricably connected with issues in the foundations of mathematics. Questions concerning consistency and inconsistency have been a central part of this. One might therefore expect paraconsistency to have an important bearing on these matters. Such expectations would not be disappointed. In this part we will see why. In the process, we will pick up the issue of the set-theoretic paradoxes left hanging in the previous section.

Let us start with a brief synopsis of the relevant history.[127] The nineteenth century was a time of great progress in the understanding of foundational matters in mathematics, matters that had been murky for a very long time. By the end of the century, the reduction of rational, irrational, and complex numbers to the natural numbers was well understood. The nature of the natural numbers still remained obscure. It was in this context that Frege and Russell proposed an analysis of the natural numbers (and thence of all numbers) in purely logical terms. A vehicle for this analysis needed to be built; the vehicle was classical logic. It was more than this, though; for what was also needed was a theory of extensions, or sets, which both Frege and Russell took to be part of logic. According to Frege's theory of extensions, the simplest and most obvious, every property has an extension. This is the unrestricted principle of set abstraction:

$$\forall y(y \in \{x; \alpha(x)\} \leftrightarrow \alpha(y))$$

The schema looks to be analytic, and very much like a part of logic.

The reduction was a very successful one... except that this theory of sets was found to be inconsistent. At first, the contradictions involved, discovered by Cantor, Burali-Forti and others, were complex, and it could be hoped that some error

[126]The issues of this section are discussed at much greater length in Priest (1995).

[127]Further details can be found in the articles on Frege, Russell, Hilbert, and Gödel in this *Handbook*.

of reasoning might be to blame. But when Russell simplified one of Cantor's arguments to produce his famous paradox, it became clear that contradiction lay at the heart of the theory of sets. Taking $x \notin x$ for $\alpha(x)$ gives:

$$\forall y(y \in \{x; x \notin x\} \leftrightarrow y \notin y)$$

Now writing $\{x; x \notin x\}$ as r, and instantiating the quantifier with this, produces $r \in r \leftrightarrow r \notin r$, and given some simple logical principles, such as the law of excluded middle or *consequentia mirabilis*, contradiction follows.

In response to this, mathematicians proposed ways of placing restrictions on the abstraction principle which were strong enough to avoid the contradictions, but not too strong to cripple standard set-theoretic reasoning, and particularly some version of the reduction of numbers to sets. How successful they were in this endeavour, we will return to in a moment. But the result for Frege and Russell's logicist programme was pretty devastating. It became clear that, though the reduction of numbers to sets could be performed, the theory of sets employed could hardly be taken as a part of logic. Whilst the unrestricted abstraction schema could plausibly be taken as an analytic principle, the things that replaced it could not be seen in this way.

Nor could this theory of sets claim any *a priori* obviousness or freedom from contradiction. This fact gave rise to another foundational programme, Hilbert's. Hilbert thought that there were certain mathematical statements whose meanings were evident, and whose truth (when true) was also evident, finitary statements—roughly, numerical equations or truth-functional compounds thereof. Other sorts of statements, and especially those containing numerical variables—which he termed *ideal*—had no concrete meaning. We can reason employing such statements, but we can do so only if the reasoning does not contradict the finitary base. And since Hilbert took the underlying logic to be classical logic, and so explosive, what this meant was that the reasoning had to be consistent. Hence, it was necessary to prove the consistency of our formalisation of mathematics. Of course, a proof could have significance only if it was secure. Hence, the proof had to be carried out finitistically, that is, by employing only finitary statements. This was Hilbert's programme.[128]

The programme was killed, historically, by Gödel's famous incompleteness theorems. Gödel showed that in any consistent theory of arithmetic there are sentences such that neither they nor their negations could be proved. Moreover, the consistency of the theory in question was one such statement. Hence, any consistent theory which includes at least finitary reasoning about numbers can not have its consistency shown in the theory itself. To confound matters further, Gödel demonstrated that, given a theory that was intuitively sound, some of the sentences that could not be proved in it could, none the less, be *shown* to be true.

Let us now turn to the issues of how paraconsistency bears on these matters and vice versa.

[128] See Hilbert (1925).

6.2 The Paradoxes of Set Theory

For a start, the set-theoretic paradoxes provide further arguments for dialetheism. The unrestricted abstraction schema is an almost irresistible principle concerning sets. Even those who deny it have trouble sticking to their official position. And if it is what it appears to be, an *a priori* truth concerning sets, then dialetheism is hard to resist.

As mentioned above, set theorists tried to avoid this conclusion by putting restrictions on the abstraction schema. And unlike the corresponding situation for the semantic paradoxes, there is now some sort of orthodoxy about this. Essentially, the orthodoxy concerns Zermelo Fraenkel set theory (ZF) and its intuitive model, the cumulative hierarchy. This model is the set-theoretic structure obtained by starting with the empty set, and applying the power-set iteratively. The construction is pursued all the way up the ordinals, collecting at limit ordinals. The instances of the abstraction schema that are true are the ones that hold in the hierarchy. That is, the sets postulated by the schema do not exist unless they are in the hierarchy.[129]

Notice that it is not contentious that the sets in the hierarchy exist. All may agree with that. The crucial claim is the one to the effect that there are no sets *outside* the hierarchy. Unfortunately, there seems to be no very convincing reason as to why this should be so. It is not the case, for example, that adding further instances of the abstraction schema must produce inconsistency. For example, one can postulate, quite consistently with ZF, the existence of non-well-founded sets (that is, sets, x_0, such that there is an infinitely descending membership sequence $x_0 \ni x_1 \ni x_2 \ni ...$; there are no such sets in the hierarchy).

Moreover, there are reasons as to why an insistence that there are no sets other than those in the hierarchy cannot be sustained. For a start, this is incompatible with mathematical practice. It is standard in category theory, in particular, to consider the category of all sets (or even all categories). Whatever else a category is, it is a collection of a certain kind. But the set of all sets in the hierarchy is not itself in the hierarchy. Indeed, if one supposes that there is such a set then, given the other resources of ZF, contradiction soon ensues.

More fundamentally, the insistence flies in the face of the *Domain Principle.* A version of this was first enunciated by Cantor. In a modern form, it is as follows: if statements quantifying over some totality are to have determinate sense, then there must be a determinate totality of quantification. The rationale for the Principle is simple: sentences that contain bound variables have no determinate sense unless the domain of quantification is determinate. Is it true, for example, that every quadratic equation has two roots? If we are talking about real roots, the answer is 'no'; if we are talking about complex roots, the answer is 'yes'. Now, statements of set theory have quantifiers that range over all sets, and, presumably, have a

[129] There are variations on the idea; for example, concerning whether or not to countenance proper classes (sub-collections of the whole hierarchy that cannot be members of anything); but these do not change the fundamental picture. In particular, all the arguments that follow can be reworked to apply to the collection of all classes (that is, sets or proper classes).

determinate sense. By the Domain Principle, the set of all sets must therefore be a determinate collection. But it is not a collection located in the hierarchy, as we have just noted.[130]

The orthodox solution to the paradoxes of set theory is therefore in just as much trouble as the plethora of solutions to the semantic paradoxes.

6.3 Paraconsistent Set Theory

In contrast with attempted consistent solutions to the set-theoretic paradoxes, a dialetheic approach simply endorses the unrestricted abstraction schema, and accepts the ensuing contradictions. But since it employs a paraconsistent consequence relation, these contradictions are quarantined. As with semantic paradoxes, not all paraconsistent logics will do what is required. For example, in a logic with *modus ponens* and Contraction, Curry paradoxes are quickly forthcoming. If α is any sentence, then the abstraction schema gives:

$$\forall y(y \in \{x; x \in x \to \alpha\} \leftrightarrow (y \in y \to \alpha))$$

Now write $\{x; x \in x \to \alpha\}$ as c, and instantiate the universal quantifier with it to obtain: $c \in c \leftrightarrow (c \in c \to \alpha)$; then argue as in the semantic case. It was shown by Brady (1989) that when based on a suitable relevant logic that does not endorse Contraction (but which contains the law of excluded middle), set theory based on the unrestricted abstraction schema, though inconsistent, is non-trivial.[131] Let us call this theory *naive relevant set theory.*

The next obvious question in this context concerns how much standard set theory can be derived in naive relevant set theory. In particular, can the reduction of number theory to set theory be obtained? If it can, then the logicist programme looks as though it can be made to fly again; Frege and Russell are vindicated.

In naive set theory, and with a qualification to which we will return in a moment, naive set theory is sufficient for most workaday set theory, concerning the basic set-theoretic operations (unions, pairs, functions, etc.).[132] As to whether it provides for the essential parts of the theory of the transfinite, or for the reduction of number theory to set theory, no definitive answer can presently be given. What can be said is that the standard versions of many of the proofs concerned fail, since they depend on properties of the conditional not present in the underlying logic. Whether there are other proofs is not known. But the best guess is that for most of these things there probably are not. If this is the case, a big question clearly hangs over the acceptability of the theory. If it cannot accommodate at least the elements of standard transfinite set theory in some way, it would seem to be inadequate.

---

[130] There are various (unsatisfactory) ways in which one may try to avoid this conclusion. These are discussed in Priest (1995), ch. 11.

[131] Brady (1983) also showed that without the law of excluded middle, the theory is consistent.

[132] Details can be found in Routley (1977).

Actually, the situation is more complex than I have so far indicated, due to considerations concerning extensionality. The natural identity condition for sets is coextensionality: two sets are the same if, as a matter of fact, they have the same members. That is:

$$\forall x(\alpha \equiv \beta) \rightarrow \{x;\alpha\} = \{x;\beta\}$$

where $\equiv$ is the material biconditional ($\alpha \equiv \beta$ is $(\alpha \wedge \beta) \vee (\neg\beta \wedge \neg\alpha)$). But if one formulates the identity conditions of sets in naive relevant set theory in this way, trouble ensues. Let r be $\{x; x \notin x\}$. We can show that $r \in r \wedge r \notin r$. Hence, for any α, we have $\alpha \equiv r \in r$, and so $\{x;\alpha\} = \{x; r \in r\}$.[133] Given standard properties of identity, it follows that all sets are identical.

One way around this problem is to replace the $\equiv$ in the identity conditions with an appropriate relevant biconditional $\leftrightarrow$.[134] But there is a cost. Let $\overline{x}$ be the complement of x, $\{y; y \notin x\}$. Then one can show that for any x and y, $\neg\exists z\ z \in x \cap \overline{x}$, and $\neg\exists z\ z \in y \cap \overline{y}$. Thus, $x \cap \overline{x}$ and $y \cap \overline{y}$ are both empty; but one cannot show that they are identical, since arbitrary contradictions are not equivalent: it is not the case that $(z \in x \wedge z \notin x) \leftrightarrow (z \in y \wedge z \notin y)$.

One might think this not too much of a problem. After all, many people find a unique empty set somewhat puzzling. However, the problem is quite pervasive. There are going to be many universal sets, for example, for exactly the same reason.[135] The structure of sets is not, therefore, a Boolean algebra. Unsurprisingly, it is a De Morgan algebra.[136] And assuming, as seems natural, that a universe of sets must have an underlying Boolean structure, this shows that using an intensional connective to state identity conditions is going to deliver a theory of some kind of entity other than sets.[137] Extensionality lies deep at the heart of set theory.

Can this fact be reconciled with a dialetheic account of sets? There is one way. Formulate the theory *entirely* in terms of material conditionals and biconditionals. Not only are these employed in the statement of identity conditions of sets, but they are also employed in the abstraction schema. After all, this is how it is done in ZF. Call this theory *simply naive set theory.* If one formulates set theory in this way, the argument that all sets are identical fails, since it requires a detachment for the material conditional:— in effect, the disjunctive syllogism. Indeed, it is now an easy matter to show that there are models of the theory with more than

[133] Or $\{x;\alpha\} = \{x; r \in r \wedge x = x\}$ if one does not like vacuous quantification.

[134] This is how extensionality is stated in Brady's formulation.

[135] And quite generally, every set is going to be duplicated many times; for if τ is any contingent truth, the same things satisfy $\alpha(x)$ and $\alpha(x) \wedge \tau$. But it is not the case that $\alpha(x) \leftrightarrow (\alpha(x) \wedge \tau)$.

[136] Indeed, as Dunn (1988) shows, if we add the assumption that there is a unique empty set and a unique universal set, the underlying logic collpses into classical logic.

[137] Possibly properties, which are more naturally thought of as intensional entities. If we read set abstracts as referring to properties and $\in$ as property instantiation then this problem does not arise, since there is no reason to expect a Boolean algebra. Note, also, that a naive theory of properties of this kind is not problematic if it is unable to deliver transfinite set theory. A dialetheic theory of properties is, in fact, quite unproblematic.

one member. Such a move radically exacerbates the problem concerning the proof-theoretic power of the theory, however. Since the material conditional does not detach, the theory is very weak indeed.

Fortunately, then, standard set theory may be interpreted in a different fashion. It can be shown that any model of ZF can be extended to a model of simply naive set theory.[138] The original model is, in fact, a consistent substructure of the new model. Hence, there are models of naive set theory in which the cumulative hierarchy is a consistent sub-structure. And we may take the standard model (or models) of naive set theory to be such (a) model(s). In this way, classical set theory, and therefore all of classical mathematics, can be interpreted as a description of a consistent substructure of the universe of sets. This fact does nothing much to help logicism, however. In particular, one cannot argue that the principles of arithmetic are analytic, since, even if the axioms of set theory are analytic, the former have not been deduced from the latter.

6.4 Gödel's Theorems

Let us now turn to Gödel's incompleteness theorems. These concern theories that contain arithmetic, phrased in a standard first order language (with only extensional connectives). Without loss of generality, we can consider just arithmetic itself. A simple statement of Gödel's first theorem says that any consistent theory of arithmetic is incomplete. This need not be disputed. Careless statements of the theorem often omit the consistency clause. What paraconsistency shows is that the clause is absolutely necessary. As we will see, there are complete but inconsistent theories of arithmetic.[139]

The existence of these follows from a general model-theoretic construction called the *Collapsing Lemma*. I will not go into all the formal details of this here, but the essential idea is as follows. Take any classical model, and consider any equivalence relation on its domain, that is also a congruence relation on the interpretations of the function symbols in the language. Now construct an LP interpretation by identifying all the elements in each equivalence class. Any predicate of the language is true of the elements identified if it is true of some one of them; and it is false if it is false of some one of them. The resulting interpretation is the *collapsed interpretation*; and the Collapsing Lemma states that anything true in the original interpretation is true in the collapsed interpretation. Hence, if the original interpretation is a model of some theory, so is the collapsed interpretation. Of course, it will be a model of other things as well. In particular, it will verify certain contradictions. Thus, for example, suppose that a and b are distinct members of an equivalence class. Then since $a \neq b$ was true before the collapse, it is true after the collapse. But since a and b have now been identified, $a = b$ is also true.

[138] See Restall (1992).

[139] This was first demonstrated, in effect, by Meyer (1978). The same paper shows that the non-triviality (though not the consistency) of a certain *consistent* arithmetic based on relevant logic may also be demonstrated within the theory itself. Further technical details of what follows can be found in Priest (2003), Section 9.

To apply this to the case at hand, take arithmetic to be formulated, as is usually done, in a first-order language containing the function symbols for successor, addition, and multiplication; and consider any model of the set of sentences in this language true in the standard model—maybe the standard model itself. It is easy to construct an appropriate equivalence relation, $\backsim$, and apply the Collapsing Lemma to give an interpretation that is a model of an inconsistent theory containing classical arithmetic. For example, the following will do: for a fixed n, $a \backsim b$ iff $(a, b \geq n)$ or $(a, b < n$ and $a = b)$. (This leaves all the numbers less than n alone, and identifies all the others.)

To bring this to bear on Gödel's theorem, choose an equivalence relation which makes the collapsed model finite. The one just mentioned will do nicely. Let T be the theory of the collapsed model (that is, the set of sentences true in it). Since what holds in a finite model is decidable (essentially by LP truth tables; quantifiers are equivalent to finite conjunctions and disjunctions), T is decidable. *A fortiori*, it is axiomatic. Hence, T is an axiomatic theory of arithmetic. It is inconsistent but complete.

Let us turn now to the second incompleteness theorem. According to this, if a theory of arithmetic is consistent, the consistency of the theory cannot be proved in the theory itself. Inconsistent theories hardly bear on this fact. Classically, consistency and non-triviality are equivalent. Indeed, the canonical statement of consistency in these matters is a statement of non-triviality. In a paraconsistent logic the two are not equivalent, of course. T, for example, is inconsistent; but it is not trivial, provided that the equivalence relation is not the extreme one which identifies all elements of the domain (in the example of $\backsim$ just given, provided that $n > 0$). The question of whether the non-triviality of an inconsistent but non-trivial theory can be proved in the theory itself is therefore a real one. And it can.

Consider T. Since it is decidable, its membership relation is expressible in the language of arithmetic. That is, there is a sentence of one free variable, $\pi(x)$, such that for any sentence, α, if $\alpha \in T$ then $\pi(\langle\alpha\rangle)$ is true, and if $\alpha \notin T$ then $\neg\pi(\langle\alpha\rangle)$ is true. (Here, $\langle\alpha\rangle$ is the numeral of the code number of α.) Hence, by the Collapsing Lemma:

π-in: if $\alpha \in T$, $\pi(\langle\alpha\rangle) \in T$

π-out: if $\alpha \notin T$ then $\neg\pi(\langle\alpha\rangle) \in T$

(Of course, for some αs, $\pi(\langle\alpha\rangle)$ and $\neg\pi(\langle\alpha\rangle)$ may both be in T.) Then provided that the equivalence relation does not identify 1 and 0, $1 = 0 \notin T$, and so $\neg\pi(\langle 1 = 0\rangle \in T$. Hence, T is non-trivial, and the statement expressing the non-triviality of T is provable in T. Gödel's second incompleteness Theorem does fail in this sense.

We have not finished with Gödel's Theorem yet, but let us ask how these matters bear on the issue of Hilbert's Programme. Hilbert's programme required that mathematics be formalised, and that the whole formalised theory be a conservative extension of the finitary part. Interestingly, Hilbert's motivating considerations did not require the formalisation to be consistent (though since he assumed that the underlying logic was classical, this was taken for granted). Like all instrumentalisms, it does not matter what happens *outside* the core (in this case, the finitary) area. The point is that the extension be a conservative one over the core area. So the use of an inconsistent theory is quite compatible with Hilbert's programme, in this sense. Does the construction we have been looking at provide what is required, then?

Not exactly. First, as far as has been shown so far, it might be the case that both $\pi(\langle 1 = 0 \rangle)$ and $\neg\pi(\langle 1 = 0 \rangle)$ are in T. If this is the case, the significance of a non-triviality proof is somewhat moot. (It could be, though, that with careful juggling we can ensure that this is not the case.) More importantly, T is not a conservative extension of the true numerical equations. For since the model is finite, distinct numbers must have been identified. Hence, there are distinct m and n such that $m = n \in T$.[140] There are certainly collapsed models where this is not the case. Suppose, for example, that we collapse a classical non-standard model of arithmetic, identifying some of the non-standard numbers, but leaving the standard numbers alone. Then the equational part of the theory of the collapsed model is consistent. In this case, though, the collapsed model is not finite, so there is no guarantee that its theory is axiomatisable. Whether or not there are collapses of non-standard models of this kind where the theory of the collapsed model is axiomatisable, or there are other axiomatic inconsistent theories with consistent equational parts, is not known at present.

6.5 Gödel's Paradox

As we have noted, paraconsistency does not destroy Gödel's theorems provided that they are stated in the right way; and in particular, that the consistency clauses are spelled out properly. Otherwise, they fail. The theorems have been held to have many philosophical consequences. If consistency is simply taken for granted, paraconsistency entirely undercuts any such mooted consequence. But, it may be argued, we are interested only in true theories, and the inconsistent theories in question can hardly be true. This move is itself moot. Once dialetheism is taken on board, it cannot simply be assumed that any true mathematical theory is consistent—especially in areas where paradoxes play, such as set theory. But leave the flights of set theory out of this; what of arithmetic? Could it be seriously supposed that this is inconsistent?

This brings us back to the version of Gödel's theorem with which I ended the first section of this part. According to this, given any axiomatic and intuitively correct theory of arithmetic, there is a sentence that is not provable in the theory, but which we can yet establish as true by intuitively correct reasoning. The sentence

[140] There is a radical move that is possible here, though: to accept that the true equations are themselves inconsistent. See Priest (1994).

is the famous undecidable sentence that "says of itself that it is not provable"; that is, a sentence, γ, of the form $\neg\pi(\langle\gamma\rangle)$.[141]

Now consider the canons of mathematical proof, those procedures whereby we establish mathematical claims as true. These are certainly intuitively correct—or we would not use them. They are not normally presented axiomatically; they are learned by mathematics students by osmosis. Yet it is reasonable to suppose that they are axiomatic. We are finite creatures; yet we can recognise, in principle, an infinite number of mathematical proofs when we see them. Hence, they must be generated by some finite set of resources. That is, they are axiomatic. In the same way, we can recognise an infinite number of grammatical sentences. Hence, these, too, must be generatable by some finite rule system, or our ability to recognise them would be inexplicable. Now consider the undecidable sentence, γ, for this system of proof. By the theorem, if the system is consistent, we cannot prove γ in it. But—again by the theorem—we can prove γ in an intuitively correct way. Hence, it must be provable in the system, since this encodes our intuitively correct reasoning. By *modus tollens* it follows that the system is inconsistent. Since this system encoded precisely our means of establishing mathematical claims as true, we have a new argument for dialetheism.

What of the undecidable sentence? It is not difficult to see that it is provable. Let us use $\vdash$ as a sign for our intuitive notion of provability. It is certainly intuitively correct that what is provable is true (indeed, this is analytic), i.e., for all α, $\vdash \pi(\langle\alpha\rangle) \supset \alpha$. In particular, then, $\vdash \pi(\langle\gamma\rangle) \supset \neg\pi(\langle\gamma\rangle)$. It follows that $\vdash \neg\pi(\langle\gamma\rangle)$, i.e., $\vdash \gamma$. Of course, since we have a proof of γ, we have also demonstrated that $\vdash \pi(\langle\gamma\rangle)$, i.e., $\vdash \neg\gamma$. Thus, the "undecidable" sentence is one of the contradictions in question. It is worth noting that if T is the formal system introduced in the last section, both γ and $\neg\gamma$ are in T. For $\gamma \in T$ or $\gamma \notin T$. But in the latter case, $\neg\pi(\langle\gamma\rangle) \in T$ (by π-out), i.e., $\gamma \in T$ anyway. But then $\pi(\langle\gamma\rangle) \in T$ (by π-in), i.e., $\neg\gamma \in T$. Hence T captures these aspects of our intuitive proof procedures admirably.

At any rate, arithmetic is inconsistent, since we can prove certain contradictions to be true; and γ is one of them. In fact, dressed in the vernacular, γ is a very recognisable paradox, in the same family as the liar: this sentence is not provable. If it is provable, it is true, so not provable. Hence it is not provable. But then we have just proved it. We may call this *Gödel's paradox*; it returns us to the discussion of semantic paradoxes in the last part. We see that there is a very intimate connection between these paradoxes, Gödel's theorems, and dialetheism.

[141] The theorem is proved explicitly in this form in Priest (1987), ch.3, where the following argument is discussed at much greater length.

7 NEGATION

7.1 What is Negation?

We have now looked at the history of both paraconsistency and dialetheism. No account of these issues could be well-rounded, however, without a discussion of a couple of philosophical notions which are intimately related to both. One of these is rationality, which I will deal with in the next part. The other, which I will deal with in this part, is negation. This is a notion that we have been taking for granted since the start of the essay. Such a crucial notion clearly cannot be left in this state. So what is negation?[142]

A natural thought is that the negation of a sentence is simply one that is obtained by inserting the word 'not' at an appropriate point before the main verb (or by some similar syntactic construction in other languages). This, however, is not right. It may well be that the negation of:

1 Bessy is a cow

is:

1n Bessy is not a cow

But as Aristotle pointed out a long time ago[143] the negation of:

2 Some cows are black

is not:

2′ Some cows are not black

but rather:

2n No cows are black

Worse, inserting a 'not' in a sentence often has nothing to do with negation at all. Consider, for example, the person who says: 'I'm not British; I'm Scottish' or 'Australia was not established as a penal colony; it was established as a British territory using forced labour'. In both cases, the "notted" sentence is true, and the utterer would not suppose otherwise. What the 'not' is doing, as the second sentence in each pair makes clear, is rejecting certain (normal?) connotations of each first sentence. Linguists sometimes call this 'metalinguistic negation'.[144]

What these examples show is that we have a grasp of the notion of negation, independent of any particular use of the word 'not', which we can use to determine

[142] The material in this section is discussed further in Priest (1999a).

[143] *De Interpretatione*, ch. 7.

[144] See, e.g., Horn (1989), ch. 5, for an excellent discussion. In the context of logic, the terminology is clearly not a happy one.

when "notting" negates. We can see that this relationship holds between examples like 1 and 1n, and 2 and 2n, but not between 2 and $2'$. This is the relationship between contradictories; let us call it the *contradictory* relation. We can, and of course modern logicians usually do, use a symbol, $\neg$, with the understanding that for any α, α and $\neg\alpha$ bear the contradictory relation to each other, but $\neg$ is a term of art.[145] Perhaps it's closest analogue in English is a phrase like 'It is not true that' (or equivalently, 'It is not the case that'). But this is not exactly the same. For a start, it brings in explicitly the notion of truth. Moreover, these phrases can also be used as "metalinguistic" negations. Just consider: 'It's not true that he's bad; he's downright evil'.

Negation, then, is the contradictory relation. But what relation is that? Different accounts of negation, and the different formal logics in which these are embedded, are exactly different theories which attempt to provide answers to this question. One may call these different notions of negation simply different negations if one wishes, but one should recall that what they are, really, are different *conceptions* of how negation functions. In the same way, different theories of matter (Aristotelian, Newtonian, quantum) provided different conceptions of the way that matter functions.

7.2 Theories of Negation

There are, in fact, many different theories as to the nature of negation. Classical logic and intuitionist logics give quite different accounts, as do many other modern logics. Indeed, we have already looked at a number of paraconsistent accounts of negation in Part 4. The existence of different theories of negation is not merely a contemporary phenomenon, however. There are different theories of negation throughout the history of logic. Let me illustrate this fact by looking briefly at three, one from ancient logic, one from (early) medieval logic, and one from (early) modern logic.

The first account is Aristotle's. First, Aristotle has to say which sentences are the negations of which. This, and related information, is encapsulated in what later came to be known as the square of opposition:

All As are Bs.	No As are Bs
Some As are Bs.	Some As are not Bs.

The top two statements are contraries. The bottom two are sub-contraries. Formulas at the opposite corners of diagonals are contradictories, and each statement at the top entails the one immediately below it.

The central claims about the properties of contradictories are to be found in Book 4 of the *Metaphysics*. As we have seen, Aristotle there defends the claim that negation satisfies the laws of non-contradiction and excluded middle:

[145] The device goes back to Stoic logicians who simply prefixed the whole sentence with a 'not'—or at any rate its Greek equivalent. Medieveal logicians often did the same—in Latin.

LEM $\Box(\alpha \vee \neg\alpha)$

LNC $\neg\Diamond(\alpha \wedge \neg\alpha)$

Further discussion of the properties of contradictories is found in *De Interpretatione*. *Prima facie*, Aristotle appears there to take back some of the *Metaphysics* account, since he argues that if α is a contingent statement about the future, neither α nor $\neg\alpha$ is true, prefiguring theories that contain truth-value gaps. There is, however, a way of squaring the two texts, and this is to read Aristotle as endorsing supervaluation of some kind.[146] Even though α and $\neg\alpha$ may both be neither true nor false now, eventually, one will be true and the other will be false. Hence, if we look at things from an "eventual" point of view, where everything receives a truth value, $\alpha \vee \neg\alpha$ (and so its necessitation) is true. In this way, Aristotle can have his law of excluded middle and eat it too.

Whether the texts can reasonably be interpreted in this way, I leave Aristotle scholars to argue about. Whatever one says about the matter, this is still only a part of Aristotle's account of negation. It does not specify, for example, what inferential relations negations enter into.[147] What are these according to Aristotle? The major part of his answer to this question is to be found in the theory of syllogistic. This tells us, for example, that 'all As are Bs and no Bs are Cs' entails 'no As are Cs'.

Scattered through the Organon are other occasional remarks concerning negation and inference. For example, Aristotle claims (*Prior Analytics* 57^b3) that contradictories cannot both entail the same thing. His argument for this depends on the claim that nothing can entail its own negation. Aristotle never developed these remarks systematically, but they were to be influential on the next theory of negation that we will look at.

This was endorsed by medieval logicians including Boethius, Abelard, and Kilwardby. It can be called the *cancellation view* of negation, since it holds that $\neg\alpha$ is something that cancels out α.[148] As Abelard puts it:[149]

> No one doubts that [a statement entailing its negation] is improper since the truth of any one of two propositions that divide truth [i.e., contradictories] not only does not require the truth of the other but rather entirely expels and extinguishes it.

As Abelard observes, if negation does work like this then α cannot entail $\neg\alpha$. For if it did, α would contain as part of its content something that neutralises it, in which event, it would have no content, and so entail nothing (or at least, nothing with any content). This principle, and related principles such as that nothing can entail

[146] For further discussion of supervaluation, see Priest (2001), 7.10.

[147] To bring this point home, note that both $\Box(\alpha \vee \neg\alpha)$ and $\neg\Diamond(\alpha \wedge \neg\alpha)$ may well hold in a modal dialetheic logic.

[148] For details, see Martin (1987) and Sylvan (2000).

[149] De Rijk (1970), p. 290.

a sentence and its contradictory, are now usually called *connexivist* principles.[150] Such principles were commonly endorsed in early medieval logic.

Carried to its logical conclusion, the cancellation account would seem to imply something much stronger than any of the connexivist principles so far mentioned; namely, that a contradiction entails nothing (with any content). For since $\neg\alpha$ cancels out α, $\alpha \wedge \neg\alpha$ has no content, and so entails nothing. This, of course, is inconsistent with Aristotle's claim which we noted in 2.1, that contradictories sometimes entail conclusions and sometimes do not. So this is not Aristotle's view. But some philosophers certainly took the account to its logical conclusion. Thus, Berkeley, when criticising the infinitesimal calculus in the *Analyst*, says:[151]

> Nothing is plainer than that no just conclusion can be directly drawn from two inconsistent premises. You may indeed suppose anything possible: But afterwards you may not suppose anything that destroys what you first supposed: or if you do, you must begin *de novo*... [When] you ... destroy one supposition by another ... you may not retain the consequences, or any part of the consequences, of the first supposition destroyed.

Despite the fact that this quotation comes from Berkeley, allegiance to the cancellation view of negation, and to the connexivist principles that it delivers, waned in the later middle ages.[152]

The third account of negation we will look at is Boole's, as he explains it in the *Mathematical Analysis of Logic*.[153] Boole's starting point in his logical investigations was the theory of the syllogism. His aim was to express syllogistic premises as equations, and then to give algebraic rules for operating on these which draw out their consequences. To turn the syllogistic forms into equations, he invokes the extensions of the terms involved. Thus, if a is the set of things that are A, etc., appropriate translations are:

All As are Bs:	$a(1-b)=0$
No As are Bs:	$ab=0$
Some As are Bs:	$ab=\nu$
Some As are not Bs:	$a(1-b)=\nu$

Here, 1 is an appropriate universal class, so that $1-b$ is the complement of b, 0 is the empty class, juxtaposition is intersection, and ν is an arbitrary non-empty class (necessary since Boole wants equations, not inequations).

---

[150] For modern connexivism, see Priest (1999b).

[151] Luce and Jessop (1951), p. 73.

[152] The reason seems to be that a truth functional account of conjunction and disjunction gained ground at this time. This makes trouble for connexivist principles. For by truth functionality, $\alpha \wedge \neg\alpha \vdash \alpha$; so by contraposition $\neg\alpha \vdash \neg(\alpha \wedge \neg\alpha)$. But $\alpha \wedge \neg\alpha \vdash \neg\alpha$. Hence, by transitivity, $\alpha \wedge \neg\alpha \vdash \neg(\alpha \wedge \neg\alpha)$. See Martin (1987) and Sylvan (2000).

[153] The account given in the *Laws of Thought* is slightly different, but not in any essential ways.

Boole extends this machinery to a propositional logic. To do this, he thinks of propositions as the sorts of thing that may change their truth value from circumstance to circumstance.[154] He can then think of 'if X then Y' as 'all cases in which X is true are cases in which Y is true': $x(1-y)=0$. Moreover, we may translate the other standard connectives thus:

X and Y:	xy
X or Y:	$x+y$
It is not the case that X:	$1-x$

where $+$ is union, which Boole takes to make sense only when x and y are disjoint. Boole thus conceives negation as complementation: the negation of X is that statement which holds exactly where X fails to hold.

It should be observed that none of the historical theories of negation that we have just looked at are the same as each other. As observed, according to Aristotle, contradictions may imply some things; whilst according to the cancellation account, strictly applied, they entail nothing. According to both Aristotle and cancellation, 'if X then it is not the case that X' is false, but under the Boolean interpretation this becomes: $x(1-(1-x))=0$. But $x(1-(1-x))=xx=x$, and this is not equal to 0 in general.

7.3 Other Negations

But which account of negation is correct? This is a substantial question, and I will return to it in the next part. Before we get to that, there are some other issues concerning negation that are worth noting.[155] Let us suppose that some paraconsistent account of negation is correct. Other accounts are then incorrect, but it does not follow that they do not succeed in capturing other meaningful and important notions.

For example, in both classical and intuitionist logic there is an absurdity constant, $\perp$, such that for all β, $\perp \rightarrow \beta$ is a logical truth. Negation may then be defined as $\alpha \rightarrow \perp$, where $\rightarrow$ is the appropriate conditional. Let us write this as $-\alpha$. The constant $\perp$ makes perfectly good sense from a dialetheic point of view. If T is the truth predicate then $\perp$ may be defined as $\forall x Tx$; the T-schema then does the rest. Thus, $-\alpha$ makes perfectly good sense for a dialetheist too. But since its properties are inherited from those of $\rightarrow$, $-\alpha$ may behave in ways quite different from classical and intuitionist negation. For example, suppose that $\rightarrow$ is the conditional of some relevant logic.[156] Then we have Explosion for $-$, since $\alpha, -\alpha \vdash \perp$ (by *modus ponens*), and so $\alpha, -\alpha \vdash \beta$.[157] Moreover, in logics like R that contain $(\alpha \wedge (\alpha \rightarrow \beta)) \rightarrow \beta$, we will have $(\alpha \wedge (\alpha \rightarrow \perp)) \rightarrow \perp$, i.e., $-(\alpha \wedge -\alpha)$,

[154] In the *Laws of Thought*, this becomes from time to time.

[155] The following material is covered in more detail in Priest (1999a).

[156] In this context, $\perp$ would usually be written as F, not to be confused with the constant f. See Anderson and Belnap (1975), p. 342f.

[157] One may wonder, in virtue of this, what happens to the liar paradox, phrased in terms of $-$. The answer is that it transforms into a Curry paradox.

a version of the law of non-contradiction. But in weaker logics, such as B, this will not be the case. And in none of these logics will one have $\alpha \vee (\alpha \to \bot)$, i.e., a version of the law of excluded middle.

Despite this, $-\alpha$ may well have useful properties. For example, let Λ be the set of all instances of the law of excluded middle, $\alpha \vee \neg\alpha$. Then, as is well known, $\Lambda \cup \Sigma \vdash_I \alpha$ iff $\Sigma \vdash_C \alpha$. In other words, full classical logic may be used even by an intuitionist, in contexts in which the law of excluded middle may be assumed enthymematically. In a similar way, suppose that Ξ is the set of all instances of $-(\alpha \wedge \neg\alpha)$. Then it is not difficult to show that for many paraconsistent consequence relations, $\vdash$, $\Xi \cup \Sigma \vdash \alpha$ iff $\Sigma \vdash_C \alpha$.[158] Hence, full classical logic can be used even by a paraconsistent logician if this schema is enthymematically assumed. The schema is one way of expressing the fact that we are reasoning about a consistent situation.[159]

Another negation-like notion, $\dagger\alpha$, may be characterised by the classical truth conditions:

$\dagger\alpha$ is true at a world, w, iff α is **not** true at w

and, if truth and falsity are independent:

$\dagger\alpha$ is false at w iff α is true at w

It might be thought that these conditions will deliver a notion with the properties of classical logic, but whether this is so depends on the properties of the negation used in the truth conditions (printed in boldface). For example, suppose that we wish to establish Explosion for $\dagger$. Then we need to establish that, for any world, w, if α and $\alpha\dagger$ are true at w then β is true at w; i.e.:

if α is true at w and α is **not** true at w, β is true at w

Now, even given that **not**-(α is true at w and α is **not** true at w)—and this may be true even if **not** is a paraconsistent negation—to infer what we want we need to invoke the inference **not**-$\gamma \vdash \gamma \to \delta$. And we may well not be entitled to this.

---

[158] For in many such logics, adding the disjunctive syllogism is sufficient to recapture classical logic. Now suppose that we have $\neg\alpha$ and $\alpha \vee \beta$. Then it follows that $(\neg\alpha \wedge \alpha) \vee \beta$. But given that $(\alpha \wedge \neg\alpha) \to \bot$, and $\bot \to \beta$, β follows by disjunction elimination.

[159] A less heavy-handed way of recapturing classical logic is as follows. Suppose that one is employing the paraconsistent logic LP. (Similar constructions can be performed with some other paraconsistent logics.) Let an evaluation ν_1 be *more consistent* than an evaluation ν_2, $\nu_1 \prec \nu_2$, iff every propositional parameter which is both true and false according to ν_1 is both true and false according to ν_2, but not vice versa. As usual, ν is a model of α if it makes α true; and ν is a model of Σ if it is a model of every member. ν is a *minimally inconsistent model* of Σ iff ν is a model of Σ and if $\mu \prec \nu$, μ is not a model of Σ. α is a *minimally inconsistent consequence* of Σ iff every minimally inconsistent model of Σ is a model of α. The construction employed in this definition of consequence is a standard one in non-monotonic logic, and is a way of enforcing certain default assumptions. Specifically, in this case, it enforces the assumption of consistency. Things are assumed to be no more inconsistent than Σ requires them to be. Unsurprisingly, it is not difficult to show that if Σ is consistent then α is a minimally inconsistent consequence of Σ iff it is a classical consequence. Thus, assuming consistency as a default assumption, a paraconsistent logician can use classical logic when reasoning from consistent information. The original idea here is due to Batens (1989), who has generalised it into a much broader programme of adaptive logics. See Batens (1999), (2000).

One issue to which this is relevant is that of a dialetheic solution to the paradoxes of self-reference. For if there is a legitimate notion, say $*_1$, that behaves like classical negation (whether or not it really *is* negation) then the T-schema cannot be endorsed, as required by a dialetheic account. If it were, and given self-reference, we could simply apply the schema to a sentence, λ, of the form $*_1 T\langle\lambda\rangle$, to obtain $T\langle\lambda\rangle \wedge *_1 T\langle\lambda\rangle$. Explosion would then give triviality. What we have seen is that there is no way that † can be shown to satisfy Explosion without assuming that the notion of negation appropriate in stating truth conditions itself satisfies certain "paradoxical" conditions. A dialetheist may simply deny this. The properties of a connective depend not just on its truth conditions, but on what follows from these; and this depends, of course, on the underlying logic.

But can we not ensure that a connective, $*_1$, has all the properties of classical negation, including Explosion, by simply characterising it as a connective that satisfies the classical proof-theoretic rules of negation? No. As was shown by Prior (1960), there is no guarantee that characterising a connective by an arbitrary set of rules succeeds in giving it meaning. Prior's example was a supposed connective, $*_2$ (tonk), satisfying the rules $\alpha \vdash \alpha *_2 \beta$, $\alpha *_2 \beta \vdash \beta$. Clearly, given $*_2$, one could infer anything from anything. It is clear, then, that $*_2$ must lack sense, on pain of triviality. But a connective, $*_1$, possessing all the properties of classical negation equally gives rise to triviality, and so must lack sense. The triviality argument is essentially the liar argument concerning $*_1$ just given. It is true that this argument invokes the T-schema, and that that schema is not included in standard logical machinery. But if a dialetheic account of truth is correct, the instances of the schema are logical truths concerning the truth predicate, just as much as the instances of the substitutivity of identicals are logical truths concerning the identity predicate. The T-schema ought, then, to be considered part of logic.

7.4 Denial

The other issue connected with negation that needs discussion is denial. Let me start by explaining what I mean by the word here. Speech acts are of many different kinds (have different illocutory forces): questioning, commanding, exhorting, etc. Perhaps the most fundamental kind of act is asserting. When a person asserts that α their aim is to get the hearer to believe that α, or at least, to believe that the speaker believes that α.[160] Denial is another kind of speech act. When a person denies that α their aim is to get the hearer to reject (refuse to accept) α, or at least, to believe that the speaker rejects α.

There was a long-standing confusion in logic, going all the way back to Aristotle, concerning assertion. The word was used to mean both the act of uttering and the content of what was uttered. A similar confusion beset the notion of denial. These confusions were finally laid to rest by Frege. And, said Frege, once this confusion is remedied, we may dispense with a *sui generis* notion of acts of denial. To deny

[160] With such Gricean refinements as seem fit.

is simply to assert a sentence containing negative particles.[161] This conclusion is certainly not required by enforcing the distinction between act and content, however; and, in fact, is false.

For a start, one can deny without asserting a sentence with a negative particle: 'England win the world cup? Get real.' Perhaps less obviously, one can also assert a sentence containing a negative particle without denying. The existence of "metalinguistic" negation makes this patent, but the point stands even without that. For example, when a dialetheist asserts 'The liar sentence is true; the liar sentence is not true', the second utterance is not meant to convey to the hearer the fact that the dialetheist rejects the first sentence: after all, they do accept it. The second sentence conveys the fact that they accept its negation too. The issue does not depend in any essential way on dialetheism. Many people have inconsistent views (about religion, politics, or whatever). Sometimes they come to discover this fact by saying inconsistent things, perhaps under some probing questioning. Thus, for some α they may utter both α and $\neg\alpha$. The second utterance is not an indication that the speaker rejects α. They *do* accept α. They just accept $\neg\alpha$ as well, at least until they revise their views. (If they did not accept α, there would be no need to revise their views.)

Denial, then, is a linguistic act *sui generis*. This does not mean that uttering a sentence with a negative particle is never an act of denial; it certainly can be. You say to me 'Truth is a consistent notion'; I say 'It certainly is not'. What I am signalling here is exactly my rejection of what you say, and maybe trying to get you to revise your views in the process. Sometimes, then, an utterance containing a negative particle is an assertion; sometimes it is a denial. This is not an unusual situation. The very same words can often (if not always) be used in quite different speech acts. I say 'the door is open'. Depending on the context, this could be an assertion, a command (to close it), or even a question. Of course, this raises the question of how one determines the illocutory force of an utterance. The short answer is that the context provides the relevant information. The long answer is surely very complex. But it suffices here that we can do it, since we often do.[162]

It might be thought that the notion of denial provides a route back into a classical account of negation. If we write $\dashv \alpha$ to indicate a denial of α, then won't $\dashv$ behave in just this way? Not at all. For a start, $\dashv$ is a force operator: it applies only to whole sentences; it cannot be embedded. Thus, $\alpha \leftrightarrow \dashv \alpha$, for example, is a nonsense. But could there not be some operator on content, say Δ, such that asserting $\Delta\alpha$ is the same as denying α? Perhaps 'I deny that' is a suitable candidate here. If this is the case, Δ behaves in no way like classical negation. It is certainly not a logical truth, for example, that $\alpha \vee \Delta\alpha$: α may be untrue, and I may simply keep my mouth shut. $\alpha \wedge \Delta\alpha$ may also be true: I may deny a truth.

[161] See Frege (1919).

[162] Some, e.g., Parsons (1990), have objected to dialetheism on the ground that if it were true, it would be impossible for anyone to rule anything out, since when a person says $\neg\alpha$, it is perfectly possible for them to accept α anyway. If *ruling out* means *denying*, this is not true, as we have just seen. And that's a denial.

For just this reason, the inference from α and $\Delta\alpha$ to an arbitrary β is invalid. Is there not an operator on content, Δ, such that assertions of α and $\Delta\alpha$ commit the utterer to everything? Indeed there is. Take for Δ the negation-like operator, $-$, of the previous section. As we saw there, this will do the trick. But as we saw there, $-$ does not behave like classical negation either.[163]

8 RATIONALITY

8.1 Multiple Criteria

Let us now turn to the final issue intimately connected with paraconsistency and dialetheism: rationality. The ideology of consistency is so firmly entrenched in orthodox western philosophy that it has been taken to provide the cornerstone of some of its most central concepts: consistency has been assumed to be a necessary condition for truth, (inferential) validity, and rationality. Paraconsistency and dialetheism clearly challenge this claim in the case of validity and truth (respectively). What of rationality? How can this work if contradictions may be tolerated?

In articulating a reply to this question, the first thing to note is that consistency, if it is a constraint on rationality, is a relatively weak one. Even the most outrageous of views can be massaged into a consistent one if one is prepared to make adjustments elsewhere. Thus, consider the claim that the earth is flat. One can render this consistent with all other beliefs if one accepts that light does not travel in straight lines, that the earth moves in a toroid, that the moon landing was a fraud, etc.[164] It is irrational for all that. There must therefore be other criteria for the rationality of a corpus of belief. What these are, philosophers of science argue about. All can agree that adequacy to the data (whatever form that takes) is one criterion. Others are more contentious. Simplicity, economy, unity, are all standardly cited, as are many different notions.[165]

Sorting out the truth in all this is, of course, an important issue for epistemology; but we do not need to go into the details here. As long as there is a multiplicity of criteria, they can come into conflict. One theory can be simple, but not handle all the data well; another can be more complex, with various *ad hoc* postulations, but give a more accurate account of the data.[166] In such cases, which is the rationally acceptable theory? Possibly, in some cases, there may be no determinate answer to this question. Rationality may be a vague notion, and there may well be situations in which rational people can disagree. However, it seems reasonable to hold that if one theory is sufficiently better than all of its competitors on sufficiently many

[163] An assertion of $-\alpha$ would normally be a denial of α, but it need not be: a trivilist would assert $-\alpha$ without rejecting α.

[164] See the works of the Flat Earth Society. At the time of writing, these can be accessed at: http://www.flat-earth.org/platygaea/faq.mhtml.

[165] For various lists, see Quine and Ullian (1970), ch. 5, Kuhn (1977), Lycan (1988), ch. 7.

[166] For example, the relationship between late 19th Century thermodynamics and the early quantum theory of energy was like this.

of the criteria, then, rationally, one should believe this rather than the others.[167] That is the way that things seems to work in the history of science, anyway. In disputes in the history of science, it is rare that all the indicators point mercilessly in the same direction. Yet a new view will often be accepted by the scientific community even though it has some black marks.

8.2 Rationality and Inconsistency

The theory of rationality just sketched, nugatory though it be, is sufficient to show how rationality works in the presence of inconsistency. In particular, it suffices to show how inconsistent beliefs can be rational. If inconsistency is a negative criterion for rationality, it is only one of many, and in particular cases it may be trumped by performance on other criteria. This is precisely what seems to have happened with the various inconsistent theories in the history of science and mathematics that we noted in 5.1. In each case, the explanatory power of the inconsistent theory well outweighed its inconsistency. Of course, in each of these cases, the inconsistent theory was eventually replaced by a consistent theory.[168] But in science, pretty much *every* theory gets replaced sooner or later. So this is nothing special.

One may even question whether inconsistency is really a negative criterion at all. (That people have usually *taken it to be so* is not in dispute.) Consistency, or at least a certain amount of it, may well be required by other criteria. For example, if the theory is an empirical one, then adequacy to observational data is certainly an important criterion. Moreover, if α describes some observable situation, we rarely, if ever, see both α and $\neg\alpha$. Empirical adequacy will therefore standardly require a theory to be consistent about observable states of affairs.[169] The question is whether consistency is a criterion in its own right.

This raises the hard question of what makes something a legitimate criterion. Different epistemologies will answer this question in different ways. For example, for a pragmatist, the only positive criteria are those which promote usefulness (in some sense). The question is therefore whether a consistent theory is, *per se*, more useful than an inconsistent one (in that sense). For a realist, on the other hand, the positive criteria are those which tend to select theories that correctly describe the appropriate external reality. The question is therefore whether we have some (perhaps transcendental) reason to believe that reality has a low degree of inconsistency. These are important questions; but they are too complex, and too tangential to the present issues, to be pursued here.[170]

[167] This is vague, too, of course. One way of tightening it up can be found in Priest (2001b).

[168] Well, this can be challenged in the case of modern quantum theory, which dallies with inconsistent notions, such as the Dirac δ-function, and is generally agreed to be inconsistent with the Theory of Relativity.

[169] For further discussion, see Priest (1999c).

[170] It might be suggested that whatever the correct account, inconsistency must be a negative criterion. Why else would we find paradoxes, like the liar, intuitively unacceptable? The answer, of course, is that we mistakenly *took* consistency to be a desideratum (perhaps under the weight

The theory of rationality just sketched shows not only how and when it is rational to accept an inconsistent theory, but how and when it is rational to give it up: the theory is simply trumped by another theory, consistent or otherwise. A frequent objection to paraconsistency and dialetheism is that if they were correct, there could never be any reason for people to reject any of their views. For any objection to a view establishes something inconsistent with it; and the person could simply accept the original view *and* the objection.[171]

Now, it is not true that objections always work in this way. They may work, for example, by showing that the position is committed to something unacceptable to its holder. And many consistent consequences are more unacceptable than some inconsistent ones. That you are a poached egg, for example, is a much more damaging consequence than that the liar sentence is both true and false. But even waiving this point, in the light of the preceding discussion, the objection is clearly incorrect. To accept the theory plus the objection is to accept an inconsistent theory. And despite paraconsistency, this may not be the rational thing to do. For example, even if inconsistency is not, *per se*, a negative mark, accepting the objection may be entirely *ad hoc*, and thus make a mess of simplicity.

8.3 The Choice of Logic

Let us now return to the question raised but deferred in the last part: which account of negation is correct? As I argued there, accounts of negation are theories concerning a certain relation. More generally, a formal logic (including its semantics) is a theory of all the relations it deals with, and, crucially, the relation of logical consequence. Now, the theory of rational belief sketched above was absolutely neutral as to what sort of theory it was whose belief was in question. The account can be applied to theories in physics, metaphysics, and, of course, logic. Thus, one determines the correct logic by seeing which one comes out best on the standard criteria of theory-choice.[172]

To see how this works, let me sketch an argument to the effect that the most rational logical theory to accept (at present) is a dialetheic one. Given the rudimentary nature of the theory of rationality I have given, and the intricacies of a number of the issues concerned, this can be no more than a sketch; but it will at least illustrate the application of the theory of rationality to logic itself.

First, one cannot isolate logic from other subjects. The applications of logic spread to many other areas in metaphysics, the philosophy of language, and elsewhere. No logic, however pretty it is, can be considered acceptable if it makes a

---

of the ideology of consistency).

[171] Versions of the objection can be found in Lewis (1982), p. 434, and Popper (1963), p. 316f.

[172] The view of logic as a theory, on a par with all other theories, is defended by Haack (1974), esp. ch. 2. She dubs it the 'pragmatist' view, though the name is not entirely happy, since the view is compatible, e.g., with orthodox realism concerning what theory is, in fact, true. Haack also accepts Quine's attack on the analytic/synthetic distinction. But the view is quite compatible with the laws of logic being analytic. We can have theories about what is analytic as much as anything else.

hash of these. In other words, one has to evaluate a logic as part of a package deal. In particular, one cannot divorce logic and truth: the two are intimately related. Thus, to keep things (overly) simple, suppose we face a choice between classical logic plus a consistent account of truth, and a paraconsistent logic plus an account of truth that endorses the T-schema, and is therefore inconsistent. Which is preferable?

First, perhaps the most crucial question concerns the extent to which each theory is adequate to the data, which, in this case, comprises the intuitions we have concerning individual inferences. A consistent account fares badly in this area, at least with respect to the inferences enshrined in the T-schema, which certainly *appear* to be valid.[173]

It may be replied that in other areas the advantages are reversed. For a paraconsistent logic is weaker than classical logic; and hence a paraconsistent logic cannot account for a number of inferences, say those used in classical mathematics, for which classical logic can account. But as we saw in 7.3, a paraconsistent logic *can* account for classical reasoning in consistent domains. The inferences might not be deductively valid; they might, on this account, be enthymematic or non-monotonic; but at least their legitimate use is explained.

What of the other criteria? Perhaps the most important of these is simplicity. As far as truth goes, there is no comparison here. There are many consistent accounts of truth (we looked at two in 5.4), and they are all quite complex, involving (usually transfinite) hierarchies, together with a bunch of *ad hoc* moves required to try to avoid extended paradoxes (the success of which is, in any case, moot, as we saw in 5.4). By contrast, a naive theory of truth, according to which truth is just that notion characterised by the T-schema, is about as simple as it is possible to be.

Again, however, it may be replied that when it comes to other areas, the boot is on the other foot. Classical logic is about as simple as it is possible to be, whilst paraconsistent logics are much more complex, and contain unmotivated elements such as ternary relations. But this difference starts to disappear under scrutiny. Any adequate logic must be at least a modal logic. After all, we need a logic that can account for our modal inferences. But now compare a standard modal logic to a relevant logic, and consider, specifically, their world semantics. There are two major differences between the semantics of a standard modal logic and the world semantics of a relevant logic. The first is that the relevant semantics has a class of logically impossible worlds, over and above the possible worlds of the modal logic. But there would seem to be just as good reason to suppose there to be logically impossible worlds as to suppose there to be physically impossible worlds. Indeed, we would seem to need such worlds to complete all the jobs that possible worlds are fruitfully employed in. For example, if propositional content is to be understood in terms of worlds, then we need impossible worlds: someone who holds that the law of excluded middle fails has a different belief from someone

[173]There are many other pertinent inferences, especially concerning the conditional. A relevant paraconsistent logic certainly out-performs classical logic in this area as well.

who holds that the law of distribution fails. If worlds are to be used to analyse counter-factual conditionals, we need logically impossible worlds: merely consider the two conditionals: if intuitionist logic were correct, the law of excluded middle would fail (true); if intuitionist logic were correct, the law of distribution would fail (false). And so on. Or, to put it another way, since any adequate logic must take account of propositional content, counter-factuals, and so on, if impossible worlds are not used to handle these, some other technique must be; and this is likely to be at least as complex as employing impossible worlds. True, Routley/Meyer semantics also employ a ternary relation to give the truth conditions of conditionals at impossible worlds, and the interpretation of this relation is problematic. But a perfectly good relevant logic can be obtained without employing a ternary relation, simply by assigning conditionals arbitrary truth values at non-normal worlds, as I noted in 4.5.[174]

The other major difference between standard world-semantics for modal logics and relevant semantics brings us back to negation. Standard world semantics employ classical negation; relevant semantics employ some other notion. But the simplest relevant account of negation is the four-valued one of 4.4.[175] This is exactly the same as the classical account in its truth and falsity conditions: $\neg\alpha$ is true (at a world) iff α is false (at that world), and vice versa. The only difference between the two accounts is that the classical one assumes that truth and falsity are exclusive and exhaustive, whilst the four-valued account imposes no such restrictions. This is hardly a significant difference in complexity. And if anything, it is the classical account which is more complex, since it imposes an *extra* condition.

There may well, of course, be other criteria relevant to a choice between the two positions we have been discussing.[176] There may equally be other areas in which one would wish to compare the performances of the two positions.[177] But at least according to the preceding considerations, a paraconsistent logic plus dialetheism about truth, comes out well ahead of an explosive and consistent view. Indeed, there are quite general considerations as to why this is always likely to be the case. Anything classical logic can do, paraconsistent logic can do too: classical logic is, after all, just a special case. But paraconsistent logic has extra resources that allow it to provide a natural solution to many of the nagging problems of classical logic. It is the rational choice.

[174]This gives a relevant logic slightly weaker than B. See Priest (2001), ch. 9.

[175]This is not the account that is employed in the usual Routley/Meyer semantics, which is the Routley $*$. But there are perfectly good relevant logics that use the four-valued account of negation; they are just not the usual ones. See Priest (2001), ch. 9.

[176]The one criterion on which the inconsistent approach clearly does not come out ahead is conservatism, which some people take to be a virtue. Conservativeness is a highly dubious virtue, however. Rationality should not reflect elements of luck, such as who got in first.

[177]Another important issue arises here. Is there a uniquely correct logic for reasoning about all domains (logical monism); or is it the case, as some have recently argued, that different domains of reasoning require different logics (logical pluralism)? For a discussion of these issues, with appropriate references, see Priest (2001c).

8.4 Conclusion

Of course, that is merely how things stand (as I see it) at the moment. The determination of the correct logic is a fallible and revisable business. It may well happen that what it is rational to believe about these matters will change as new theories and new pieces of evidence appear. Indeed, revision is to be expected historically: our logical theories have often been revised in the light of new developments. In contemporary universities, logic is often taught in an ahistorical fashion, which induces a certain short-sightedness and a corresponding dogmatism. A knowledge of the history of logic, as displayed in this volume, and the others in the series, should engender not only a sense of excitement about the development of logic, but a certain humility about our own perspective.[178]

[178] A version of this essay was given in a series of seminars at the University of St Andrews in Michaelmas term, 2000. I am grateful to those participating for many helpful comments and criticisms, but especially to Andrés Bobenrieth, Roy Cook, Agustín Rayo, Stephen Read, Stewart Shapiro, John Skorupski, and Crispin Wright.

BIBLIOGRAPHY

[1] W.Ackermann (1956), 'Begründung einer Strengen Implikation', *Journal of Symbolic Logic* 21, 113-28.

[2] A.R.Anderson and N.D.Belnap (1958), 'A Modification of Ackermann's "Rigorous Implication"' (abstract), *Journal of Symbolic Logic* 23, 457-8.

[3] A.R.Anderson and N.D.Belnap (1975), *Entailment; the Logic of Relevance and Necessity*, Vol. I, Princeton: Princeton University Press.

[4] A.R.Anderson, N.D.Belnap, and J.M.Dunn (1992), *Entailment; the Logic of Relevance and Necessity*, Vol. II, Princeton: Princeton University Press.

[5] F.G.Asenjo (1966), 'A Calculus for Antinomies', *Notre Dame Journal of Formal Logic* 16, 103-5.

[6] E.J.Ashworth (1974), *Language and Logic in the Post-Medieval Period*, Dordrecht: Reidel Publishing Company.

[7] J.Barnes (1984), *The Complete Works of Aristotle*, Princeton: Princeton University Press.

[8] D.Batens (1989), 'Dynamic Dialectical Logic', ch. 6 of Priest, Routley, and Norman (1989).

[9] D.Batens (1999), 'Inconsistency-Adaptive Logics', pp. 445-72 of E.Orłowska (ed.), *Logic at Work: Essays Dedicated to the Memory of Helena Rasiowa*, Heidelberg: Physica Verlag (Springer).

[10] D.Batens (2000), 'A Survey of Inconsistency-Adaptive Logics', pp. 49-73 of D.Batens, C.Mortensen, G.Priest, and J.-P. Van-Bendegem (eds.), *Frontiers of Paraconsistent Logic*, Baldock: Research Studies Press.

[11] I.M.Bocheński (1963), *Ancient Formal Logic*, Amsterdam: North Holland Publishing Company.

[12] D.Bochvar (1939), 'On a Three-Valued Calculus and its Applications to the Analysis of Contradictions', *Mathematičéskij Sbornik* 4, 287-300.

[13] I.Boh (1982), 'Consequences', ch. 15 of N.Kretzman, A.Kenny, and J.Pinborg (eds.), *The Cambridge History of Later Medieval Logic*, Cambridge: Cambridge University Press.

[14] R.Brady (1983), 'The Simple Consistency of a Set Theory Based on *CSQ*', *Notre Dame Journal of Formal Logic* 24, 431-9.

[15] R.Brady (1989), 'The Non-Triviality of Dialectical Set Theory', ch. 16 of Priest, Routley, and Norman (1989).

[16] R.Brady (ed.) (2003), *Relevant Logics and their Rivals*, Vol. II, Aldershot: Ashgate.

[17] B.Brown and P.Schotch (1999), 'Logic and Aggregation', *Journal of Philosophical Logic* 28, 265-87.

[18] A.Church (1951), 'The Weak Calculus of Implication', pp. 22-37 of A.Menne, A.Wilhelmy, and H.Angsil (eds.), *Kontrolliertes Denken, Unterschungen zum Logikkalkül und zur Logik der Einzelwissenschaften*, Munich: Kommissions-Verlag Karl Alber.

[19] N.C.A.da Costa (1974), 'The Theory of Inconsistent Formal Systems', *Notre Dame Journal of Formal Logic* 15, 497-510.

[20] N.C.A.da Costa and E.Alves (1977), 'A Semantical Analysis of the Calculi C_n', *Notre Dame Journal of Formal Logic* 18, 621-30.

[21] N.C.A.da Costa and M.Guillaume (1965), 'Négations Composées et la Loi de Peirce dans les System C_n', *Portugaliae Mathemetica* 24, 201-9.

[22] L.M.De Rijk (ed.) (1970), *Petrus Abealardus; Dialectica*, Assen: van Gorcum & Co.

[23] K.Došen (1992), 'The First Axiomatization of a Relevant Logic', *Journal of Philosophical Logic* 21, 339-56.

[24] J.M.Dunn (1976), 'Intuitive Semantics for First Degree Entailment and "Coupled Trees"', *Philosophical Studies* 29, 149-68.

[25] J.M.Dunn (1988), 'The Impossibility of Certain Second-Order Non-Classical Logics with Extensionality', pp. 261-79 of D.F.Austin (ed.), *Philosophical Analysis*, Dordrecht: Kluwer Academic Publishers.

[26] K.Fine (1974), 'Models for Entailment', *Journal of Philosophical Logic* 3, 347-72.

[27] G.Frege (1919), 'Negation', *Beiträge zur Philosophie des deutschen Idealismus* 1, 143-57; translated into English and reprinted in P.Geach and M.Black (eds.), *Translations for the Philosophical Writings of Gottlob Frege*, Oxford: Basil Blackwell, 1960.

[28] L.Goddard (1998), 'The Inconsistency of Traditional Logic', *Australasian Journal of Philosophy* 76, 152-64.

[29] L.Goldstein (1989), 'Wittgenstein and Paraconsistency', ch. 19 of Priest, Routley, and Norman (1989).
[30] N.D.Goodman (1981), 'The Logic of Contradiction', *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 27, 119-26.
[31] L.Goodship (1996), 'On Dialethism', *Australasian Journal of Philosophy* 74, 153-61.
[32] S.Haack (1974), *Deviant Logic*, Cambridge: Cambridge University Press.
[33] E.Hamilton and H.Cairns (1961), *Plato; the Collected Dialogues*, Princeton: Princeton University Press.
[34] G.Heron (trans.) (1954), *Of Learned Ignorance*, London: Routledge & Kegan Paul.
[35] D.Hilbert (1925), 'On the Infinite', *Mathematische Annalen* 95, 161-90; translated into English and reprinted as pp. 134-151 of P.Benacerraf and H.Putnam (eds.), *Philosophy of Mathematics; Selected Readings*, Oxford: Basil Blackwell.
[36] L.R.Horn (1989), *A Natural History of Negation*, Chicago: Chicago University Press.
[37] D.Hyde (1997), 'From Heaps and Gaps to Heaps of Gluts', *Mind* 106, 641-60.
[38] S.Jaśkowski (1969), 'Propositional Calculus for Contradictory Deductive Systems', *Studia Logica* 24, 143-57.
[39] R.E.Jennings (1994), *The Genealogy of Disjunction*, Oxford: Oxford University Press.
[40] C.E.M.Joad (1927), 'Bertrand Russell; the Man and the Things he Stands for', *The New Leader*, December 9th.
[41] T.Kasulis (1981), *Zen Action; Zen Person*, Honolulu: University of Hawai'i Press.
[42] H.-J. Kim (trans.) (1985), 'Zazenshin: Admonitions for Zazen', ch. 12 of *Flowers of Emptiness*, Lewiston: The Edward Mellen Press.
[43] G.S.Kirk and J.E.Raven (1957), *The Presocratic Philosophers*, Cambridge: Cambridge University Press.
[44] C.Kirwan (1993), *Aristotle; Metaphysics, Books* Γ*,* Δ*,* E, 2nd edition, Oxford: Oxford University Press.
[45] S.C.Kleene (1952), *Introduction to Metamathematics*, Amsterdam: North Holland Publishing Company.
[46] W. & M.Kneale (1962), *The Development of Logic*, Oxford: Clarendon Press.
[47] L.Kolakowski (1978), *Main Currents of Marxism; Vol. I, the Founders*, Oxford: Oxford University Press.
[48] T.S.Kuhn (1977), 'Objectivity, Value Judgment and Theory Choice', ch. 13 of *The Essential Tension*, Chicago, IL: University of Chicago Press.
[49] M.Kwok, M.Palmer, and J.Ramsey (trans.) (1993), *Tao Te Ching*, Dorset: Element Books Ltd.
[50] C.I.Lewis and C.H.Langford (1932), *Symbolic Logic*, The Century Company; 2nd edition, New York: Dover Publications, 1959.
[51] D.Lewis (1982), 'Logic for Equivocators', *Noûs* 14, 431-41.
[52] A.Loparić (1986), 'A Semantical Study of Some Propositional Calculi', *Journal of Non-Classical Logic* 3, 73-95.
[53] A.A.Luce and T.E.Jessop (eds.) (1951), *The Collected Works of George Berkeley*, Vol. IV, London: Thomas Nelson & Sons.
[54] J.Łukasiewicz (1910), *O Zasadzie Sprzeczności u Aristotelsa*, Krakow: Studium Krytyczne.
[55] J.Łukasiewicz (1970), 'On the Principle of Contradiction in Aristotle', *Review of Metaphysics* 24, 485-509.
[56] W.G.Lycan (1988), *Judgment and Justification*, Cambridge: Cambridge University Press.
[57] S.MacKenna (trans.) (1991), *Plotinus; the Enneads*, London: Penguin Classics.
[58] V.H.Mair (1994), *Wandering on the Way; Early Taoist Parables of Chuang Tsu*, New York: Bantam Books.
[59] C.Martin (1986), 'William's Machine', *Journal of Philosophy* 83, 564-72.
[60] C.Martin, (1987), 'Embarrassing Arguments and Surprising Conclusions in the Development of Theories of the Conditional in the Twelfth Century', pp. 377-400 of J.Jolivet and A.de Libera (eds.), *Gilbert de Poitiers et ses contemporains: aux origines de la 'Logica modernorum': actes du 7e Symposium Européen d'histoire de la logique et de la semantique médiévales, Poitiers, 17-22 juin 1985*, Napoli: Bibliopolis.
[61] A.Maurer (1967), 'Nicholas of Cusa', pp. 496-8, Vol. 5, of P.Edwards (ed.), *Encyclopedia of Philosophy*, London: Macmillan.
[62] A.Meinong (1907), *Über die Stellung der Gegenstandstheorie in System der Wissenshaften*, Leipzig: R.Voitlander Verlag.

[63] R.K.Meyer (1978), 'Relevant Arithmetic', *Bulletin of the Section of Logic, Polish Academy of Sciences* 5, 133-7.
[64] R.K.Meyer and R.Routley 'Algebraic Analysis of Entailment, I', *Logique et Analyse* 15, 407-28.
[65] A.V.Miller (trans.) (1969), *The Science of Logic*, London: Allen and Unwin.
[66] S.-K. Moh (1950), 'The Deduction Theorem and Two New Logical Systems', *Methodos* 2, 56-75.
[67] C.Mortensen (1995), *Inconsistent Mathematics*, Dordrecht: Kluwer Academic Publishers.
[68] T.Parsons (1990), 'True Contradictions', *Canadian Journal of Philosophy* 20, 335-53.
[69] G.Peano (1967), 'The Principles of Arithmetic, Presented by a New Method', pp. 83-97 of J. Van Heijenoort (ed.), *From Frege to Gödel; A Source Book of Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press.
[70] L.Peña (1989), 'Verum et Ens Convertuntur', ch. 20 of Priest, Routley, and Norman (1989).
[71] K.Popper (1963), *Conjectures and Refutations*, London: Routledge and Kegan Paul.
[72] G.Priest (1973), 'A Bedside Reader's Guide to the Conventionalist Philosophy of Mathematics', in J.Bell, J.Cole, G.Priest, and A.Slomson (eds.), *Proceedings of the Bertrand Russell Memorial Logic Conference; Denmark, 1971*, Leeds: Bertrand Russell Memorial Logic Conference.
[73] G.Priest (1979), 'Logic of Paradox', *Journal of Philosophical Logic* 8, 219-41.
[74] G.Priest (1987), *In Contradiction; a Study of the Transconsistent*, Dordrecht: Kluwer Academic Publishers.
[75] G.Priest (1989-90), 'Dialectic and Dialetheic', *Science and Society* 53, 388-415.
[76] G.Priest (1991), 'Minimally Inconsistent *LP*', *Studia Logica* 50, 321-31.
[77] G.Priest (1994), 'Is Arithmetic Consistent?', *Mind* 103, 337-49.
[78] G.Priest (1995), *Beyond the Limits of Thought*, Cambridge: Cambridge University Press; second (revised) edition, Oxford: Oxford University Press, 2002.
[79] G.Priest (1998), 'To Be *and* Not to Be—that is the Answer; on Aristotle and the Law of Non-Contradiction', *Philosophiegeschichte und Logische Analyse* 1, 91-130.
[80] G.Priest (1999a), 'What Not? A Defence of Dialetheic Theory of Negation', pp. 101-20 of D.Gabbay and H.Wansing (eds.), *What is Negation?*, Dordrecht: Kluwer Academic Publishers.
[81] G.Priest (1999b), 'Negation as Cancellation and Connexive Logic', *Topoi* 18, 141-8.
[82] G.Priest (1999c), 'Perceiving Contradictions', *Australasian Journal of Philosophy* 77, 439-46.
[83] G.Priest (2000a), 'Truth and Contradiction', *Philosophical Quarterly* 50, 305-19.
[84] G.Priest (2000b), 'Vasil'ev and Imaginary Logic', *History and Philosophy of Logic* 21, 135-46.
[85] G.Priest (2001a), *Introduction to Non-Classical Logic*, Cambridge: Cambridge University Press.
[86] G.Priest (2001b), 'Paraconsistent Belief Revision', *Theoria* 68, 214-28.
[87] G.Priest (2001c), 'Logic: One or Many?', in J.Woods and B.Brown (eds.), *Logical Consequences: Rival Approaches. Proceedings of the 1999 Conference of the Society of Exact Philosophy*, Stanmore: Hermes Science Publishers Ltd.
[88] G.Priest (2003), 'Paraconsistent Logic', pp. 287-393, Vol. 6 of D.Gabbay and F.Guenthner (eds.), *Handbook of Philosophical Logic*, 2nd edition, Dordrecht: Kluwer Academic Publishers.
[89] G.Priest and R.Routley (1989), 'The Philosophical Significance and Inevitability of Paraconsistency', ch. 18 of Priest, Routley, and Norman (1989).
[90] G.Priest, R.Routley, and J.Norman (1989), *Paraconsistent Logics: Essays on the Inconsistent*, Munich: Philosophia Verlag.
[91] G.Priest and R.Sylvan (1992), 'Simplified Semantics for Basic Relevant Logics', *Journal of Philosophical Logic* 21, 217-32.
[92] A.Prior (1960), 'The Runabout Inference Ticket', *Analysis* 21, 38-9.
[93] W.V.Quine (1966), 'Three Grades of Modal Involvement', ch. 13 of *The Ways of Paradox and Other Essays,* New York: Random House.
[94] W.V.Quine and J.Ullian (1970), *The Web of Belief*, New York, NY: Random House.
[95] P.T.Raju (1953-54), 'The Principle of Four-Cornered Negation in Indian Philosophy', *Review of Metaphysics* 7, 694-713.
[96] C.Rauszer (1977), 'Applications of Kripke Models to Heyting-Brower Logic', *Studia Logica* 36, 61-71.
[97] S.Read (1988), *Relevant Logic; a Philosophical Examination of Inference*, Oxford: Basil Blackwell.

[98] S.Read (1993), ‘Formal and Material Consequence; Disjunctive Syllogism and Gamma’, pp. 233-59 of K.Jacobi (ed.), *Argumentationstheorie; Scholastische Forschungen zu den logischen und semantischen Regeln korrekten Folgerns*, Leiden: E.J.Brill.
[99] N.Rescher and R.Brandom (1980), *The Logic of Inconsistency*, Oxford: Basil Blackwell.
[100] N.Rescher and R.Manor (1970-71), ‘On Inferences from Inconsistent Premises’, *Theory and Decision* 1, 179-217.
[101] G.Restall (1992), ‘A Note on Naive Set Theory in *LP*’, *Notre Dame Journal of Formal Logic* 33, 422-32.
[102] G.Restall (1993), ‘Simplified Semantics for Relevant Logics (and some of their Rivals)’, *Journal of Philosophical Logic* 22, 481-511.
[103] G.Restall (1999), ‘Negation in Relevant Logics (How I Stopped Worrying and Learned to Love the Routley Star)’, pp. 53-76 of D.Gabbay and H.Wansing (eds.), *What is Negation?*, Dordrecht: Kluwer Academic Publishers.
[104] T.M.Robinson (1987), *Heraclitus*, Toronto: University of Toronto Press.
[105] R.Routley (1977), ‘Ultralogic as Universal?’, *Relevance Logic Newsletter* 2, 50-90 and 138-75; reprinted as an appendix to *Exploring Meinong’s Jungle and Beyond*, Canberra: Research School of Social Sciences, 1980.
[106] R.Routley (1979), ‘Dialectical Logic, Semantics and Metamathematics’ *Erkenntnis* 14, 301-31.
[107] R.Routley (1980), ‘Three Meinongs’, ch. 5 of *Exploring Meinong’s Jungle and Beyond*, Canberra: Research School of Social Sciences.
[108] R.Routley, V.Plumwood, R.K.Meyer, and R.Brady (1984), *Relevant Logics and their Rivals*, Vol. I Atascadero: Ridgeview.
[109] R. & V.Routley (1972), ‘Semantics of First Degree Entailment’, *Noûs* 6, 335-59.
[110] B.Russell (1905), ‘On Denoting’, *Mind* 14, 479-93.
[111] B.Russell (1997), *The Collected Papers of Bertrand Russell*, Vol. 11, ed. J. Slater, London: Routledge.
[112] P.Schotch and R.Jennings (1980), ‘Inference and Necessity’, *Journal of Philosophical Logic* 9, 327-40.
[113] N.Smart (1967), ‘Eckhart, Meister’, pp. 449-551, Vol. 2, of P.Edwards (ed.), *Encyclopedia of Philosophy*, London: Macmillan.
[114] T.J.Smiley (1959), ‘Entailment and Deducibility’, *Proceedings of the Aristotelian Society* 59, 233-254.
[115] E.Stump (1989), *Dialectic and its Place in Medieval Logic*, Ithaca: Cornell University Press.
[116] R.Sylvan (2000), ‘A Preliminary Western History of Sociative Logics’, ch. 5 of D.Hyde and G.Priest (eds.), *Sociative Logics and their Applications; Essays by the Late Richard Sylvan*, Aldershot: Ashgate Publishers.
[117] K.Tanahashi (ed.) (1985), ‘Mountains and Waters Sūtra’, pp. 97-107 of *Moon in a Dewdrop; Writings of Zen Master Dōgen*, New York: Farrar, Straus and Giroux.
[118] N.Tennant (1984), ‘Perfect Validity, Entailment and Paraconsistency’, *Studia Logica* 43, 181-200.
[119] N.Tennant (1992), *Autologic,* Edinburgh: Edinburgh University Press.
[120] I.Urbas (1990), ‘Paraconsistency’, *Studies in Soviet Thought* 39, 434-54.
[121] A.Urquhart (1972), ‘Semantics for Relevant Logics’, *Journal of Symbolic Logic* 37, 274-82.
[122] N.A.Vasil’év (1912-13), ‘Logica i Métalogica’, *Logos* 1-2, 53-81; translated into English by V.Vasyukov as ‘Logic and Metalogic’, *Axiomathes* 4 (1993), 329-51.
[123] W.Wallace (trans.) (1975), *Hegel’s Logic; being Part One of the Encyclopaedia of the Philosophical Sciences*, Oxford: Oxford University Press.
[124] L.Wittgenstein (1978), *Remarks on the Foundations of Mathematics*, 3rd (revised) edition, Oxford: Basil Blackwell.
[125] L.Wittgenstein (1979), *Wittgenstein and the Vienna Circle*, ed. B.F.McGuinness, Oxford Basil Blackwell.