THE PHILOSOPHICAL SIGNIFICANCE OF PARADOXES

Greg Restall[*]

Philosophy Department

The University of St Andrews

gr69@st-andrews.ac.uk

DECEMBER 2023

In this essay, I will examine the significance of a range of paradoxes (such as semantic, set-theoretic, sorites paradoxes) for a number of different philosophical issues concerning logic, including the choice of a logical system, the epistemology of logic, and the boundary – if there is one – between logical and non-logical concepts.

The semantic, set-theoretic and sorites paradoxes are seemingly valid arguments from seemingly true premises to seemingly untrue conclusions. These arguments are *paradoxes* because there seems to be no *obvious* resolution to any of them. Each paradox gives rise to some small set of seemingly true principles which are jointly unsatisfiable. As such, paradoxes are fertile ground for the development of new ideas, because each different resolution, rejecting some one of the seemingly true principles to resolve the tension, has what seems to be good reasons in its favour—the other principles, each of which also seems to be true.

There are a small number of *logical* notions that play a role in many paradoxical proofs (the conditional, negation, etc.), and there are also a number of different *non-logical* notions (the truth predicate, set/class membership, reference, as well as vague predicates), which also play a role in these paradoxes. The question naturally arises for logic—is there any reason to keep logical constants and quantifiers fixed, or should we revise them in the wake of the paradoxes?

This is one significant divide in the history of discussion of the paradoxes, between those approaches that are logically revisionist, and those that are not. In this essay, I will examine revisionist responses to the paradoxes from perspectives of *proof-first* and *model-first* approaches to logic, since both approaches have given rise to different revisionary proposals.

---

[*]Thanks to audiences at Logic Lunch at the Arché Philosophical Research Centre at St Andrews, at the Euphilo Seminar, Patrick Greenough with the students in our *Paradoxes* module in Semester 2, 2023, and especially Miriam Bowen, Jann Paul Engler, Viviane Fairbank, Filippo Ferrari, Ben Martin, Sabina Domínguez Parrado, Matteo Plebani, Thomas Randriamahazaka, Lorenzo Rossi, Sophie Nagler, Gil Sagi, Giulia Schirripa, Francisca Silva, Sebastian Speitel and Shawn Standefer, for helpful discussions about this draft and the ideas discussed therein. ¶ This is a draft of chapter to appear in Elke Brendel, Massimiliano Carrara, Filippo Ferrari, Ole Hjortland, Gil Sagi, Gila Sher, and Florian Steinberger (eds.), *The Oxford Handbook of the Philosophy of Logic*.

Finally, with these different revisionary approaches on the table, together with the logically "conservative" approach of keeping logic fixed and responding to the paradoxes by attending to the so-called *non-logical* notions in use in the paradoxical derivation, we can examine in more detail the different reasons you might have (whether the approach is *exceptionalist* or *anti-exceptionalist* [26, 52]) for choosing a revisionist or conservative approach. We will see how all four combinations of revisionist or conservative, and exceptionalist or anti-exceptionalist are possible avenues for responding to the paradoxes.

1. THE PARADOXES

Here is one version of the Liar Paradox:

> Consider a sentence that says of itself that it is not true:
>
> (λ): (λ) is not true.
>
> Suppose first that this sentence is not true. Then, since this is what it says, it is true after all. So supposing that (λ) is not true, we can conclude that (λ) is true. But if (λ) is true, then what (λ) says is the case. And what (λ) says is that it is not true. So (λ) is not true. But this contradicts what we had already concluded, namely that (λ) is true. [15, p. 321]

One delightful but infuriating feature of the Liar paradox is that it is so *simple*. We can reach a contradictory conclusion from a small number of inferences, each of which seems unproblematic when considered on its own. I spell the principles out in detail here, and then I will show how they combine to produce the argument to the contradictory conclusion.

$$\frac{\neg A \quad A}{\bot}\,\neg E \qquad \frac{\begin{matrix}[A]^1\\ \Pi\\ \bot\end{matrix}}{\neg A}\,\neg I^1 \qquad \frac{a = b \quad Fa}{Fb}\,{=}E \qquad \frac{A}{T\ulcorner A\urcorner}\,TI \qquad \frac{T\ulcorner A\urcorner}{A}\,TE$$

Each of these rules has the virtue of isolating one single connective, operator or predicate, and describing some facet of what we can *do* with that notion, either as an *elimination* rule, which shows what follows from the use of the notion, or an *introduction* rule, which spells out how we might be in a position to conclude a claim using that notion.

Negation elimination ($\neg$E), according to which, a statement A and its negation $\neg A$ are contradictory.

We can understand the $\neg$E rule as follows: if we manage to prove $\neg A$ and also prove A (from various assumptions), we can treat the proof up to this point as a *refutation* of one or other of our assumption we have appealed to on the way to this dead end. We mark refutations, in this sense, with the special conclusion marker

$\bot$, which need not be understood as a formula, but could be reserved as a special kind of 'punctuation mark'.[1]

Negation introduction ($\neg$I), according to which, you can prove a negation $\neg A$ by first assuming the negand A and deriving a contradiction from this assumption.

So, if I do reach a dead end in my reasoning (by reaching a contradiction), I can look among the assumptions and choose one and blame *that*. Since A cannot hold (along with the other assumptions), then given the other assumptions, we have $\neg A$.[2]

Identity elimination ($=$E), the principle of "indiscernibility" of identicals. If a and b are identical, then any feature of a is a feature of b.

The identity rule is about *identity*, and nothig else. The F in this rule stands in for *any* predicate, but we will apply $=$E only in the case of the truth predicate, in the liar paradoxical proof.

Introduction for the truth predicate (TI): you can infer $T\ulcorner A\urcorner$ from A.

Elimination for the truth predicate (TE): you can infer A from $T\ulcorner A\urcorner$.

The truth rules are simple, too, except they use the notion of *quotation*. For every sentence A we take ourselves to have a singular term $\ulcorner A\urcorner$, which is a device by which we can *mention* the formula A rather than using it, as we do when A occurs in a proof unadorned by these corner quotes. While A is the kind of thing we can assume in a proof, and we can prove or disprove, conjoin with something else, etc., $\ulcorner A\urcorner$ is something of which we can predicate properties. In our formal grammar, A is a sentence which we can use to assert something. $\ulcorner A\urcorner$ is a singular term which refers to an object, namely, the sentence A.

With the grammar understood, the truth rules allow us to introduce quotation names and the truth predicate in one inference, and to reverse this inference. The connections between using a sentence and predicating truth *of* it is familiar: if $2+2=4$ then $\ulcorner 2+2=4\urcorner$ is true. Conversely, if $\ulcorner 2+2=4\urcorner$ is true, then $2+2=4$.

* * *

Using these five rules alone, we can start from the assumption that we have a sentence λ which says of itself that it is not true (i.e., some λ where $\lambda = \ulcorner\neg T\lambda\urcorner$), and we prove a contradiction using only the five inference principles we have seen

[1] This is how the negation rules are treated in Tennant's *Natural Logic* [68], as one example.

[2] With this understanding, $\bot$ need not be a formula and could be taken to be a structural feature of proofs, but nothing will hang on this interpretation of $\bot$ in this chapter. (Neil Tennant treats the contradiction marker in this way in his *Natural Logic* [68]. There it is written '$\sharp$'.)

above:

$$
\dfrac{
\dfrac{
\dfrac{
\dfrac{
\dfrac{\lambda = \ulcorner\neg T\lambda\urcorner \quad [T\lambda]^1}{T\ulcorner\neg T\lambda\urcorner}\,{=}E
}{\neg T\lambda}\,TE
\quad [T\lambda]^1
}{\bot}\,\neg E
}{\neg T\lambda}\,\neg I^1
\qquad
\dfrac{
\lambda = \ulcorner\neg T\lambda\urcorner
\quad
\dfrac{
\dfrac{
\dfrac{
\dfrac{
\dfrac{\lambda = \ulcorner\neg T\lambda\urcorner \quad [T\lambda]^2}{T\ulcorner\neg T\lambda\urcorner}\,{=}E
}{\neg T\lambda}\,TE
\quad [T\lambda]^2
}{\bot}\,\neg E
}{\neg T\lambda}\,\neg I^2
}{T\ulcorner\neg T\lambda\urcorner}\,TI
}{T\lambda}\,{=}E
}{\bot}\,\neg E
$$

So, if a proof of a contradiction counts as a refutation of the assmptions granted in the proof, then either we are to reject the initial (and only remaining) assumption, that $\lambda = \ulcorner\neg T\lambda\urcorner$, or one of the five inference principles is to be rejected. This is the lesson of such a paradoxical argument, and the benefit of such a sharp and precise proof to such a repugnant conclusion is that we have a map of the territory of different responses to the liar paradox. The most obvious way to classify responses to the paradoxes is in which kind of rule is rejected. In the case of the *liar* paradox, is fault to be found in the rules concerning the *truth predicate* (in TI or TE, or in the assumption that there is such a λ where $\lambda = \ulcorner\neg T\lambda\urcorner$), or is the blame to be laid on the so-called *logical* principles (here, the negation rules or the identity rule)? So-called 'classical' theories of truth in which either self-reference is banned [65, 66], or the truth rules TI/E are either restricted or reinterpreted [22,23,27,28] differ from 'revisionary' theories of truth where the truth rules hold unrestrictedly and restrictions are placed on either the negation rules or the identity rule [2, 3, 16, 32, 46]. However, before we attempt to understand the costs and benefits of taking up a position on any particular territory on this map, we would do well to consider other paradoxes, for similar considerations apply the different notions in use in those paradoxe.

* * *

Curry's paradox [13,39] is formulated not using negation, but using the *conditional*. In Curry's original formulation [13], an arbitrary conclusion is proved given an item that serves as a *fixed point* for operations on propositions. One way to provide a fixed point is to use the truth predicate and the identity predicate, as we did with the liar paradox. Here, instead of using the concept of *truth*, I will consider of *property possession*, to give an example of another way that fixed points might arise.[3]

[3] I describe all of this in terms of *properties* and *exemplification* rather than *classes* and *membership*, as is familiar from Russell's paradox concerning the class of all classes that are not members of themselves, but the reasoning would be the same in either property-theoretic or class-theoretic dress. Classes satisfy an extra condition, *extensionality*, not satisfied by properties. If C and D are classes with exactly the same members, then they are identical; while P and Q might be distinct properties exemplified by exactly the same items. The class of equiangular triangles in the Euclidean plane is

Some very plausible rules of property exemplification go like this: if $A(t)$ holds, then t exemplifies the property $\langle x : A(x)\rangle$. We write this, '$t\ \varepsilon\ \langle x : A(x)\rangle$'. And conversely, if t exemplifies the property $\langle x : A(x)\rangle$ then we also have $A(t)$. These are the introduction and elimination rules for property exemplification. The green things are all and only the things that exemplify the property of being green.

$$\dfrac{\begin{matrix}[A]^1\\ \Pi\\ B\end{matrix}}{A \to B}{\to}I^1 \qquad \dfrac{A \to B \quad B}{B}{\to}E \qquad \dfrac{A(t)}{t\ \varepsilon\ \langle x : A(x)\rangle}\varepsilon I \qquad \dfrac{t\ \varepsilon\ \langle x : A(x)\rangle}{A(t)}\varepsilon E$$

We work with a very simple property. We choose an arbitrary proposition p and consider this property: being a thing such that if it exemplifies itself, then p. For short, we call this property c, which is short for $\langle x : x\ \varepsilon\ x \to p\rangle$. The εE and εI rules when applied to $c\ \varepsilon\ c$ have a very interesting structure. We have:

$$\dfrac{c\ \varepsilon\ c}{c\ \varepsilon\ c \to p}\varepsilon E \qquad \dfrac{c\ \varepsilon\ c \to p}{c\ \varepsilon\ c}\varepsilon I$$

With these instances of the ε rules, we can reason as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{[c\ \varepsilon\ c]^1}{c\ \varepsilon\ c \to p}\varepsilon E \quad [c\ \varepsilon\ c]^1}{p}{\to}E}{c\ \varepsilon\ c \to p}{\to}I^1 \qquad \dfrac{\dfrac{\dfrac{\dfrac{[c\ \varepsilon\ c]^2}{c\ \varepsilon\ c \to p}\varepsilon E \quad [c\ \varepsilon\ c]^2}{p}{\to}E}{c\ \varepsilon\ c \to p}{\to}I^2}{c\ \varepsilon\ c}\varepsilon I}{p}{\to}E$$

The result is a proof of the proposition p, which was whatever proposition we cared to choose in the first place. We can prove anything we like, from no premises at all. This proof has a remarkably similar structure to the liar paradoxical proof, but it uses *none* of the same inference rules. This means that the lessons learned concerning the paradoxes cannot be confined to the behaviour of an individual logical constant (say, *negation*) or a single predicate (say, the truth predicate). If Curry's paradox, concerning property abstraction, is the same kind of problem as the liar paradox concerning truth, under a different guise, then we have an opportunity to refine our diagnosis, to give a more general account that can give insight into a wider range of settings.

This point can be pushed *much* further. Class theoretic paradoxes can be formulated in languages in which the only notions in play are the identity predicate and class abstraction itself [25, 55].[4] Thankfully, some of diagnoses of the liar paradox

exactly the same as the class of equilateral triangles in the Euclidean plane. However, the property of being an equilateral triangle need not be identical to the property of being an equiangular triangle. Since extensionality is not involved in Curry's paradox, we need not worry about whether we are reasoning about classes or properties.

[4] The *Hinnion–Libert* paradox is easy to state: consider the class $\mathfrak{H}_p$, defined as $\{x : \{y : x \in x\} = \{y : p\}\}$. Using only $=$E, $\in$I/E and a simple extensionality rule, we can derive p, with no other logical vocabulary entering into the proof [55].

given above do generalise in a relatively natural way, beyond a focus on negation or on the truth predicate alone.[5]

* * *

Consider a coloured strip, shading gradually and evently from red on the left to green on the right [30]. Divide the strip into ten thousand evenly sized patches, labelled 1 to 10 000 from left to right. For each n from 1 to 10 000, consider the claim p_n, that *patch number* n *looks red* (*to me*). The first claim, p_1, is true. The last such claim, $p_{10\,000}$ is false, since patch 10 000 does not look red to me, it looks green. The claim $p_1 \to p_2$ also seems true, because patches 1 and 2 look indistinguishable to me. Each claim of the form $p_n \to p_{n+1}$ seems just as true as $p_1 \to p_2$, since the strip shades evenly from red to green, with no sharp changes in observable colour, and we chose so many subdivisions, that patch differs from its neighbours by at most a tiny difference, and my powers of visual discrimination are limited. So, each of the premises of this argument seem true, but the conclusion seems false.

$p_1, p_1 \to p_2, p_2 \to p_3, \ldots, p_{9\,999} \to p_{10\,000}$.
Therefore, $p_{10\,000}$.

Unfortunately, we can reach the seemingly false conclusion from the seemingly true premises using the single inference rule $\to$E. Our range of options seems very narrow indeed. If we are to reject the conclusion then either we accepte that the argument is valid and we find some way to deny that the premises all hold, or we reject the validity of the argument.

Rather than exploring the range of options available in response to the sorites paradox, we will take stock and with the different semantic paradoxes and the sorites paradox in mind, we will return to start to map out the territory of different responses to such paradoxes, and their philosophical significance.

The first thing to notice is that these paradoxes are *genuinely* paradoxical. There is not an obvious fall-back position to take, there is not an obvious contender for an invalid principle or false premise. Furthermore, for the everyday user of the concept of truth, or property possession, or the possessor of vague predicates, the everyday response is to be puzzled by the paradoxes, to recognise that we can get into trouble if we reason in *that* way, and to retreat to the more everyday uses of these notions where paradox does not threaten, insofar as such a retreat is possible. The *general* issue of charting the boundaries of the extensions of vague predicates is only pressing for the *philosopher* who wants to know how use these notions in general, to understand whether the rules for inference are unrestrictedly valid, and if

[5]For further discussion of some of the considerations around the costs and benefits of uniformity in diagnoses of the paradoxes of self-reference, see the critical discussion [47, 53]. (Although published earlier than [47] paper, [53] was written partly in response to an earlier version Priest's paper, which was presented at the 1991 *Australasian Association for Logic* conference.)

not, how they could be replaced with other principles which fare better.[6] For those of us who care about the rules of logic, or of truth, or properties, or the semantics of other predicates in our language (whether vague ones or not), addressing the paradoxes at this level of generality is a pressing matter. For everyday language users, it is not.

So, to keep our philosopher hats on, let's look at these inferences and see how we might characterise different solutions that have been offered for these paradoxes. One broad distinction that can be drawn is between the inferences in these paradoxical arguments that trade on the behaviour of the logical connectives, quantifiers and identity, and the inferences and premises involving particular *non-logical* notions – truth, class, property, and vague predicates. The distinction between the logical and the non-logical may be substantial (perhaps the logical notions are topic neutral and the non-logical ones are not [1]; perhaps the logical principles are formal in a way that the other principles are not [34], perhaps there are other ways to draw this boundary [61]), but maybe the differences, if any, are a matter of degree and not of kind. Even if there is no substantial difference between the kinds of features represented by logical vocabulary as opposed to non-logical vocabulary, perhaps certain kinds of logical vocabulary (such as the conditional and the universal quantifier) plays a distinctive role in our own conceptual schemes that deserves marking out as distinct [18, 40]. We need not take a stand on the logic/non-logic boundary except to note that the distinction is worth marking on pragmatic grounds, since various approaches to the paradoxes can be usefully characterised in terms of the way they treat concepts on either side of this boundary.

2. LOGICALLY "CONSERVATIVE" RESPONSES

Logically conservative approaches do not revise the negation or conditional or identity rules in our paradoxical proofs. They take each paradoxical proof to teach us some lesson concerning the non-logical concepts exploited therein. This is a constraint on our theory of truth, sets, types, etc., and our account of vague predicates.

TRUTH If we do not tinker with the negation rules and the identity rule in the liar paradoxical proof, then if we are to resist the contradictory conclusion, we must reject either the premise to the effect that there is something (λ) that is the claim that λ is not true, or we must reject one of the rules TI or TE. Responses in this vein have a long history. Medieval analyses of the liar paradox fall in this category, though it is anachronistic to call them 'classical' in any sense. Still, for logicians such as John Buridan or Thomas Bradwardine, it is the T *introduction* rule that is to be rejected. There are liar propositions that say of themselves that they are not true, and it is correct to conclude that indeed, they *are* not true. However, it doesn't follow from

[6] Of course, the boundaries of certain vague predicates, such as 'person', 'alive', 'murder', 'intention', and so on, raise important ethical and legal issues, which are of concern not only for philosophers. Nonetheless, we find a way to navigate with these terms without a settled theory of *how* they work.

that conclusion that the original liar sentence is true [51]. The liar proposition is self-contradictory (and hence false) while the other judgement, to the effect that the liar is not true is *not* self-contradictory. For some judgement to be true, it is necessary for *everything* it says to hold, and the liar sentence not only says that it is not true, it also (as all sentences implicitly do) says of itself that it *is* true, and so, it cannot be true. It is simply false, and the derivation is blocked at the TI step.

Now, this medieval analysis of the liar paradox has fallen out of favour, and other approaches to the truth predicate have taken centre stage. The most famous is Tarski's own account of the truth predicate, according to which a truth predicate can be added to any sufficiently defined language, provided that we respect the distinction between the original language and the new metalanguage, introduced with the addition of the T-predicate. Here, we can keep the TI and TE rules, provided that we respect the synactic distinction: the sentence A must come from the original language and the truth predication $T\ulcorner A\urcorner$ is not a sentence from the original language but from the metalanguage. The two rules can survive as a *definition* for how we assign the extension of T in the newly expanded language [65, 66]. No paradox threatens, because the distinction between language and metalanguage ensures that there either there is no sentence λ satisfying the constraint that $\lambda = \ulcorner\neg T\lambda\urcorner$, or if there is, it must be a sentence purely of the metalanguage and not of the object language, and so there is no guarantee that the TI and TE rules apply to it, since these are guaranteed only when T is applied to sentences of the object language.

This restriction of truth to some metalinguistic device that applies only to sentences of a prior language is, perhaps, overly restrictive [32]. We seem to be able to apply the truth predicate more liberally to sentences where no rigid distinction between language and metalanguage is in force, and where there are reference cycles [32], or even unending chains of reference [82]. So, richer classical accounts of the logic of truth seek to provide a semantics for the truth predicate in which sentences like λ are allowed, and for which many, but not all, of the instances of the rules TI and TE are admitted. This can be attempted in various ways [22,23,27,28], and the details do not need to detain us here. Suffice it to say that there are a range of different priniciples proposed to allow for *many* of the instances of TI and TE, without admitting all of them. Finding *prinicpled* reasons to reject some instances while accepting others is a difficult matter, and it is fair to say that no consensus view has emerged.

SETS The situation for a classical response to the paradoxes is very different in the case of Russell's Paradox concerning sets and set memberships. Here, the strong set-theoretic principle (the *set comprehension* scheme), to the effect that for any predicate $\phi(x)$ there is a set $\{x : \phi(x)\}$ consisting of all and only the objects satisfying the predicate, was immediately rejected by the mathematical community once Russell's paradox came to light. The orthodoxy in the mathematical community settled around a different account of set rather quickly, and it was one which systematically and decisively rejected not just *some* instances of the set comprehension scheme, but to reject it in its entirety, in favour of much more modest principles of set formation [24, 45], now understood as the cumulative hierarchy. You cannot conjur a set out of thin air merely as the extension of any predicate (that way lies paradox), but the ground is much safer if we have more modest principles of set construction, starting with the empty set, the formation of unions, subsets, power sets, and an infinite set – all principles that have stood the test of time in mathematical reasoning about sets. The result is a theory of sets that, while there is disagreement around the edges concerning how far the theory should be developed, the resulting family of axioms (Zermelo Frankel set theory, with the Axiom of Choice, or ZFC) is the centre of a striking consensus of a consistent view of sets and their features.

This does not, of course, mean that there are no questions when it comes to the correct theory of sets, if indeed there is a single best correct theory to be found. Although, we do not think of the family of all non-self-membered sets as forming a *set* – it cannot be, since on the ZFC picture of sets, *no* sets are self-membered, so this is the family of *all* sets whatsoever, and on the picture of the cumultative hierarchy, there is no set of all sets – we do talk of the *class* of all such sets. It is straightforward to think of classes as a layer 'above' the cumultaive hierarchy of sets, but once we do this, it is hard to see where to stop, and have levels of classes containing earlier classes, etc. If our job is to do everyday mathematics with sets, there is no need to spend any time at these dizzying heights, but if our desire is to understand how to understand sets, it would be good to have a principled answer to the question of where the hierarchy of sets (and classes, if there is one) *stops*, and why it does so.

PROPERTIES AND TYPES However, perhaps talk of *classes* as super-sized sets is not the correct way to think of the family of all sets. We have another way to think about such things, and this is through talk of *properties*. We saw, in the discussion of Curry's paradox that we also cannot identify a unique property $\langle x : \phi(x) \rangle$ corresponding to any extension: at least, we cannot do so and treat properties themselves as objects which can also *have* properties, as the inferences εI and εE implicitly allow when understood in a first-order language. The consensus classical treatment of such *property* paradoxes is relatively straightforward. We may talk of properties if we like, and allow that for any predicate $\phi(x)$ there is a property of being an x such that $\phi(x)$. However, this 'there *is* a property' should not be understood as a first-order quantifier, but is better understood one level up, as a *second-order* quantifier. The following second-order property abstraction scheme:

$$\exists X \forall x (Xx \leftrightarrow \phi(x))$$

is not only *consistent*, it is a valid theorem of second-order logic. If we think of properties (or at least, properties in-extension) as the domain of the second order quantifier, then there we have retained the spirit of the judgements εI and εE, while rejecting the paradoxical consequences, for just as with Tarski's account of truth, the level distinction between first-order and second-order quantifiers ensures that the paradoxical property (of being an x where $x \ \varepsilon \ x$ implies p) cannot arise, since properties are not the kinds of things that can apply to *themselves*. Second-order properties apply only to first-order objects.

Insofar a second-order logic is accepted, there is a great deal that one can do in this vocabulary, consistently, and without being subject to paradox [59], retaining the spirit of the property application conditions consistently, at the cost of a type distinction. However, some small worries remain, because the type discipline of second-order (or higher-order) logic is hard to maintain in general, since properties themselves have features, and some features seem to be shared *across* levels. A relation R is said to be *symmetric* if whenever Rxy then Ryx. This definition makes sense whether we think of R as a *property*$_1$ (applying only to first-order objects) or a *property*$_2$ (applying to *properties*$_1$), or an even higher order property. It seems that *symmetry* could be undersood as a level-ambiguous *feature* of properties, which can be had at any level of the hierarchy, but what kind of property is *that*?

As with sets and truth, we can find a coherent classical core of the idea, where in this case (as with truth) the safety is found by a type or level distinction, but how far from that safe centre we should stray is an open question. We will see that the situation is nothing like this when it comes to vagueness.

VAGUENESS A logically "conservative" approach to the sorites paradox will accept the validity of the sorites argument and hence to reject the set of premises by means or other (in the case where the conclusion is to be rejected, at least), since the argument form is nothing more than an extended *modus ponens*. Recall, though, that in sorites arguments of the form discussed above, each of the premises are designed to be (individually, at least) hard to deny. In any classical two-valued model, each sentence p_n ("patch n looks red to me") is assigned either *true* or *false*, and lest they *all* be true, there will be some pair of adjacent patches where the model represents one as looking red to me and the other as *not* looking red to me. However, the setup is designed to make each patch indistinguishable from its immediate neighbour. So, any two-valued counterexample seems to diverge from the datum of indistinguishability of adjacent patches.[7]

So, if diverging from the datum of indistinguishability is unpalatable, one natural reaction involves expanding the picture of semantic evaluation to allow for more than the two values of "true" and "false": logics with truth-value gaps, or a whole panoply of degrees of truth might provide ways to understand the sorites paradoxes [5, 57, 63]. To do this, is to be revisionary, to at least some degree, on logical grounds, so I will defer discussion of these approaches to the next section.

On the other hand, we may attempt to retain the classical two-valued semantic picture but preserve the intuition of indistinguishability (or the so-called *tolerance* principles concerning vague predicates), by taking the relationship between truth *simpliciter* and truth-in-a-model to be looser than a one-to-one correspondence between boolean models and reality. Perhaps the everyday description "looks red to me" should not be modelled by a two-valued *cut* on the domain of the model,

[7] One option, of course, is to reinterpret the datum as *epistemicists* do, by distinguishing x *being red*, x *looking red* and *knowing that* x *looks red*, and exploiting these differences in creative ways [75].

but by a range of equally acceptable cuts [17, 29, 71, 80, 81]. Perhaps each two-valued model represents a way to represent a *sharp* analogue to the vague predicate "looks red". Truth in any individual model adheres to classical logical strictures, but truth simpliciter (however this is to be defined) may diverge from those strictures in ways which might respect tolerance judgements better than truth-in-a-model does. For example, if there is a range of equally acceptable models, then the tolerance conditionals $p_n \to p_{n+1}$ are each not true in *all* of the acceptable models, they are each true in the *overwhelming majority* of such models, and perhaps this is enough for a kind of acceptability. However, any account that takes a range of evaluations to be a model of truth simpliciter comes at its own costs. In any case where p is true in some acceptable valuations and $\neg p$ is true in others, then $p \lor \neg p$ is true in all of them, but neither p nor $\neg p$ is also true in all of them. So, if we identify truth simpliciter with truth in *all* acceptable models (as the *supervaluationist* would have it), then we must revise the logic of disjunction, by rejecting the claim that $A \lor B$ is true iff A is true or B is true. Similarly, p is true in *some* acceptable valuations, and $\neg p$ is true in some acceptable valuations, but $p \land \neg p$ is true in *none* of them, so if we identify truth simpliciter with truth in *some* acceptable valuation (as the *subvaluationist* would have it), then we reject the claim that $A \land B$ is true iff A is true and B is true. This is not the end of the story for the supervaluationist, but it is agreed that providing an account of how a *range* of valuations is to be related to coherent account of logical consequence is a difficult task [71].

Logically conservative approaches are to be distinguished from logically revisionary approaches, which point to the logical inference steps as opposed to the non-logical inference steps in the argument. Although there is a core of consensus among classical responses to the paradoxes, there are enough complications in each of the logically conservative approachs to make the option of looking elsewhere an attractive one. Since logical consequence can be understood both proof-theoretically and model-theoretically, and since *both* proof-theoretic and model-theoretic considerations have been raised to give an 'answer' to the questions raised by the paradoxes, I will examine both approaches in turn, starting with approaches that take their cue from *models*.

3. LOGICAL REVISION: *MODEL* FIRST

Return to the sorites paradox, and the difficulty that we had in assigning an evaluation to the vague predicate in a single model that respects the tolerance conditions that $p_i \to p_{i+1}$ never be *false*. A natural thought in the light of vagueness is to take it that some of the p_i are simply *true*, others are simply *false*, but some claims, in some indeterminate middle, are in some kind of third zone between truth and falsity. Let's model this by allowing for formulas to take an intermediate value—n. Having an extra semantic value to give to sentences like the liar would also allow us to give a different response to the *semantic* paradoxes, so we focus on this for the moment. If we retain the traditional two-valued interpretation of truth conditions for the connectives like this:

- $m(A \wedge B) = 1$ iff $m(A) = 1$ and $m(B) = 1$,
  $m(A \wedge B) = 0$ iff $m(A) = 0$ or $m(B) = 0$.
- $m(A \vee B) = 1$ iff $m(A) = 1$ or $m(B) = 1$,
  $m(A \vee B) = 0$ iff $m(A) = 0$ and $m(B) = 0$.
- $m(A \rightarrow B) = 1$ iff $m(A) = 0$ or $m(B) = 1$,
  $m(A \rightarrow B) = 0$ iff $m(A) = 1$ and $m(B) = 0$.
- $m(\bot) = 0$.

but we allow for atoms to receive the third value n, we can retain these clauses. For example, if $m(p) = n$ then $m(\neg p) = n$ and $m(p \vee \neg p) = m(p \wedge \neg p) = n$ too.

With these three-valued valuations at hand, the task of defining what it is for a model to provide a counterexample to an argument becomes more complicated. Each valuation m provides a *tripartite* verdict for formulas, and so, we have more options for a definition of validity. Different options will give different analyses of the paradoxes. Start with validity in Kleene's (strong) three-valued logic, defined by setting $X \vDash_{K3} A$ iff whenever $m(B) = 1$ for each $B \in X$, then $m(A) = 1$ too [31]. K3-validity takes *truth* to consist in bearing the value 1, and validity is preservation of truth. On this view, if A and $\neg A$ are both neither true nor false, then so is $A \vee \neg A$. So, on this analysis, if p_i is a statement saying that a given borderline item on the strip is red, then $p_i \vee \neg p_i$ is also a borderline statement. The approach to the *logic* of the connectives is now highly non-classical. However, it does not provide a *counterexample* to the sorites paradox, for modus ponens is K3-valid for the conditional. In the case the strip's statements p_i trend from 1 to n to 0 as we go from left to right, at the very least the conditionals $p_i \rightarrow p_{i+1}$ are never *false* (i.e. 0), since we never have a conditional whose antecedent is 1 and consequent 0, but many of the conditionals will be evaluated as n, so K3 is no help if we wish to render the sorites argument *invalid*.

However, the reasoning does break down in the case of the liar paradox or Curry's paradox, if we evaluate the proofs by the canons of K3-validity. In particular, $\neg I$ is K3-invalid in the following sense: We may indeed have $X, A \vDash_{K3} \bot$, if no valuation makes each member of X, and A have the value 1. This does not mean that we must have $X \vDash_{K3} \neg A$. On the K3 approach, the liar paradoxical reasoning breaks down. It is consistent to assign $T\lambda$ the value n, in which case it has the same value as its negation, and the same could go for Curry paradoxical reasoning: $\rightarrow I$ is similarly K3-invalid.

However, K3 is not the only way to interpret these tables. We could instead take an argument to have a counterexample when in some valuation where the conclusion is assigned the value 0 and the premises are each assigned a value *other than* 0. This is equally a generalisation of classical logic, and on this perspective, $A \vee \neg A$ is now valid (it is never assigned 0), but have $A \wedge \neg A$ no longer entails $\bot$.

An assignment giving A the value n now (weakly) satisfies $A \land \neg A$, in the sense that this formula is never assigned the value 0. The resulting logic is Priest's *Logic of Paradox*, LP [46]. In LP, liar-paradoxical argument also breaks down, not at the $\neg I$ step, but rather at $\neg E$, since $A, \neg A \not\vdash_{LP} \bot$. Here, the liar paradox does not fall into a truth-value *gap*. It is rather in the *glut*, which may be understood as an overlap between truth and falsity [49].

Instead of pursuing the different applications of gaps and gluts [4, 6, 14], we will pause to consider one of the most elegant formal properties of such models. We have seen that the trivialising proofs fail to be K3 and LP valid. This should give us *some* hope that expanding our scope to these models gives us some way of modelling semantic concepts like a robust truth predicate without collapse into triviality. Some paradoxical proofs fail in K3 and in LP, but this does not mean the strong principles of truth or property abstraction are paradox-free on these lights. To show that these logics are safe for unrestricted truth or property abstraction, we need to do more. We can turn to an ingenious model construction technique, due independently to Brady [7], Gilmore [21], Martin and Woodruff [35], and Kripke [32], that shows that no trivialising argument like this is possible, given these three-valued valuations, interpreted either in the K3 manner or the LP manner. We can construct a *model* in which $T\ulcorner A \urcorner$ is always assigned the same value as A, even in the presence of fixed point sentences such as λ.

The details are subtle, but it is not too difficult to sketch the key ideas of the construction.[8] The most important insight is that we can think of our semantic values, 1, n and 0 as "ordered" with $n \sqsubset 1$ and $n \sqsubset 0$, but with 0 and 1 incomparable by $\sqsubset$. The key idea is that n is *less specific* as a value than either 0 or 1. Each connective *respects* this order in the following way. If we think of valuations as similarly ordered (so $m_1 \sqsubseteq m_2$ iff $m_1(p) \sqsubseteq m_2(p)$ for each atom p) then this ordering extends to the entire language: $m_1(A) \sqsubseteq m_2(A)$ for *every* formula A, made up out of $\land, \lor, \neg, \to$ and $\bot$.

Using this fact, we can construct models for semantic notions like the truth predicate that validate the T*I*/*E* rules. In particular, this construction makes models for the truth predicate that ensure that $m(T\ulcorner A \urcorner) = m(A)$ for each formula A. We define our model m like this: start with a model m_0 that interprets the original language however we like, except that each T-sentence has the value n. A process of *refining* the interpretation of T, step-by-step, by assigning $T\ulcorner A \urcorner$ at a new stage whatever we assigned A at the *previous* stage, will make the T predicate more and more refined, without 'un-refining' anything else in the language (because each of the logical connectives respects the refinement ordering). This process must have a limit, where any new stage is exactly as refined as the previous one, and no new refinement is possible. This is our model m where $m(T\ulcorner A \urcorner) = m(A)$, a model in which the TI and TE rules are satisfied.

This construction is independent of our choice of *logic*. All that is required is in models, the connectives be appropriately respectful of the ordering $\sqsubseteq$. There is

[8] For a quick overview of the proof, see my *Proofs and Models in Philosophical Logic* [56], Section 3.2.

nothing in this construction about the appropriate notion of *counterexample* that is used. This construction applies to the logics K3 and LP, since these are merely two different ways to interpret models liek these. Since in any such model, $T\lambda$ must be assigned n, from the point of view of K3, this is a model in which the liar is neither true nor false. But for LP, this is a model in which the liar is both true and false.

The model similarly, does not simply tweak the behaviour of one or other connective. It treats them all equally, in the sense that all have to respect the ordering $\sqsubseteq$. This means that the logic of negation and the conditional are both revised. This is essential to the construction.[9] Defenders of truth value gaps and truth value gluts each have reason to take the semantics of the conditional delivered here to be less than we could hope for. For K3, we have $\not\models_{K3} p \to p$, since we can assign p the value n. Yet, even given truth-value gaps there seems to be a sense in which *if* the liar is true, then the liar is true. It is just that this sense of "if" cannot be encoded by the conditional of K3. Similarly, in LP, we have counterexamples to *modus ponens*. Models m where $m(p) = n$ and $m(q) = 0$ ensure $p, p \to q \not\models_{LP} q$.

There is no good way to repair the truth tables while keeping (a) the K3 or LP notions of counterexamples, (b) keeping the "classical" behaviour of the truth table of the conditional for the inputs 0 and 1, and (c) preserving monotonicity over $\sqsubseteq$. So, there is a strong constraint, given by monotonicity, on the evaluation conditions for conditionals in logics with gaps and with gluts.

The lacuna with conditionals might call into question the entire enterprise of 3-valued interpretations. One way to understand our options is to notice that not only does the *conditional* fare badly with respect to monotonicity, but so does the notion of *counterexample* for K3 and LP. Given that $\to$, with its traditional evaluation table *is* monotonic, we should not be surprised if does not fit well with K3 and LP validity, at least when understood as a *conditional*.

However, inspired by the truth table for $\to$, we *could* say that a model m is a counterexample to the argument from X to A when $m(B) = 1$ for each $B \in X$, and $m(A) = 0$. That's when we have refuted the argument from X to A. *This* notion is monotonic, because it will not be revised away as any formulas evaluated n are refined into either 0 or 1, since each premise is settled as 1 and the conclusion is settled as 0. For *this* notion of (in)validity, once a model refutes an argument, any refinement of that model refutes the argument, too.

The picture of validity that results when we take a counterexample in a three-valued model to require that the premises be assigned 1 and conclusions 0, is called ST-*validity*, for Strict/Tolerant validity [11]. On one interpretation, we can think of a formula assigned 1 as being true, when measured to a *strict* standard of evaluation, while a formula assigned 0 fails to be true, even when measured to a *tolerant* standard. The formulas assigned n are those that are not strictly true but not tolerantly untrue. In other words, they are are tolerantly true but not strictly true.

[9] However, more complex constructions can be developed to provide more complicated models in which connectives with other behaviour are defined [8, 9, 16].

Since negation flips 1 to 0 and back, we can think of the formulas assigned 0 as strictly *false*. The formulas assigined 1 or n are (at least) *tolerantly* true, while the formluas assigned n or 0 are (at least) *tolerantly* false An ST-counterexample is a *clear* counterexample: an assignment according to which the premises are strictly true and the conclusion is strictly false, so for an argument to be valid, whenever the premises are strictly true, the conclusion is at least tolerantly true. Hence, ST validity.

On this view of counterexamples, both reflexivity ($\succ A \to A$) and *modus ponens* ($A, A \to B \succ B$) remain *valid*, $A \to A$ never has the value 0 in any model, so it has no counterexample. In no model can we assign A and $A \to B$ the value 1 while assigning B the value 0. In fact *every* classically valid argument $X \succ A$ is valid, under ST semantics. If we had some m where $m(X) = 1$ and $m(A) = 0$ then simply refine m into a classical evaluation m' assigning 1 or 0 to each atom, and by monotonicity $m'(X) = 1$ and $m'(A) = 0$, and this remains a two-valued counterexample to our argument. This is a way to preserve *all* of the valid inferences of classical logic, while keeping fixed points for the truth predicate.

However, we have no proof of a contradiction from this starting point, because our model construction still applies. We can assign $T\ulcorner A \urcorner$ and A the same values, and $T\lambda$ (and its negation) the value n, with no contradiction arising. How, then, does the paradoxical proof break down?

The proof fails at the last $\neg$E step. In our paradoxical reasoning, we can indeed show that from $\lambda = \ulcorner \neg T\lambda \urcorner$, that $\neg T\lambda$. (In any model in which $\lambda = \ulcorner \neg T\lambda \urcorner$ is assigned 1, $T\lambda$ and $\neg T\lambda$ are assigned n: so this step is ST-valid in our models.) We similarly have the proof from $\neg T\lambda$ to $T\lambda$ for exactly the same reasons. However, we cannot chain these two proofs together to infer $\bot$, even though the inference from $T\lambda, \neg T\lambda$ to $\bot$ is *also* ST-valid! In general, ST-validity cannot be chained together. We can have $A \vDash_{ST} B$ and $B \vDash_{ST} C$ without having $A \vDash_{ST} C$.[10]

Here we have one model-theoretic construction which can be interpreted in three different ways, depending on the treatment of the intermediate value n. If it is treated as a *gap* we have K3, if it is a *glut* we have LP, and if it can not feature in a counterexample at all, whether as a premise or a conclusion, we have ST. The one construction gives three very different diagnoses of the liar paradox, and related semantic paradoxes.

* * *

Let's end this section of revisionary model-theoretic interpretations of the paradoxes by applying the ST interpretation to the sorites paradox, because the dichotomy between strict and tolerant truth provides a distinctive way to read the sorites reasoning. As before, with K3, the sorites argument is *still* valid, even by ST lights, since if the premises are all strictly true, the conclusion is also strictly true and hence, also tolerantly true. However, this is not the only thing we can say

[10] For one example, suppose B receives the value n on *every* evaluation, while A is 1 and C is 0.

concerning the paradox by ST lights. We could diagnose the appeal of the paradoxical reasoning by saying that by ST lights, the premises are all at least *tolerantly* true (if the truth values are first 1, then n then 0 when evaluated from left-to-right), since we *never* have a tolerance conditional that has a strictly true antecedent and a strictly true consequent. The tolerantly true premises do not lead to a tolerantly true conclusion, however, so the one model gives a different kind of *failure*, which is not ST invalidity, but is a failure of preservation of tolerant truth.

Now, this is an appealing range of semantic options, which gives us a range of different perspectives from which to view the semantic paradoxes. However, they do not come without their own costs. We have seen that K3 and LP are without a well behaved conditional connective. In ST logical consequence is not transitive. Our Liar derivation tells us that $\lambda = \ulcorner\neg T\lambda\urcorner \succ T\lambda$ and $\lambda = \ulcorner\neg T\lambda\urcorner, T\lambda \succ$. We do *not* conclude from this that $\lambda = \ulcorner\neg T\lambda\urcorner$ must fail. The derivation tells us that – given $\lambda = \ulcorner\neg T\lambda\urcorner$ – $T\lambda$ is not *strictly* true (it cannot be assigned 1) but is *tolerantly* true (it is not assigned 0).

This is *one* example of how we might find a single logical principle which fails in our paradoxical reasoning, despite the fact that the Curry-paradoxical proof and the liar-paradoxical proof share no logical vocabulary. In ST, logical consequence is not transitive, and our proofs fail at points where two proofs are *composed* in a '*Cut*' step. In the case of the Curry proof, this is the last $\rightarrow$E step, for we can prove, from property principles alone that $c \mathrel{\varepsilon} c \rightarrow p$ is at least *tolerantly* true, and we can also prove that $c \mathrel{\varepsilon} c$ is tolerantly true. However, the tolerant truth of $c \mathrel{\varepsilon} c$ and $c \mathrel{\varepsilon} c \rightarrow p$ is not enough to prove the tolerant truth of p, and so the derivation is blocked at the last step.

The one logical principle rejected, on ST grounds, in the paradoxical reasoning is not, really, any of the connective rules, when understood in isolation. In one sense it is a *structural rule*, the composition of proofs. This is a good place to consider other ways that revisions to logical principles have been proposed, and that is to consider *other* structural rules.[11]

[11] It is worth mentioning that the transitivity of logical consequence, or the *Cut* rule, is not the only structural rule that has been interrogated as a suspect in paradoxical proofs. There is a dual logical principle, the reflexivity of logical consequence, (or the *Identity* rule in the sequent calculus), which can also be blamed [19]. It is harder to see, in a natural deduction proof setting, where the identity rule has been applied, but it corresponds to the unrestricted availability of assumptions. If a natural deduction proof from premises X to conclusion C corresponds to a derivation of the sequent $X \succ C$, then the underivability of the sequent $A \succ A$ means that the formula A standing as its own assumption and its own conclusion, cannot be a well-formed natural deduction proof. For then non-reflexive treatment of the liar paradox, the natural deduction 'proof', fails at the first time $T\lambda$ is assumed, since the sequent $T\lambda \succ T\lambda$ fails. Space does not allow for a further discussion of this approach, save to say that in the three-valued setting, a non-reflexive treatment can be modelled by the dual TS (Tolerant/Strict) notion of consequence. A counterexample to an argument is a valuation in which the premises are at least tolerantly true (assigned n or 1) while the conclusion fails to be strictly true (so is assigned either n or 0). When a paradoxical sentence, like $T\lambda$ is assigned n, the model serves as a counterexample to the argument from that sentence to itself.

4. LOGICAL REVISION: *PROOF* FIRST

As we have seen, a closer analysis of the paradoxical proofs shows that despite the lack of any shared inference rules, there are certain logical principles at work in these paradoxical proofs. One option is the behaviour of proof composition, as brought to light by considering ST. But there are others.

The most noticeable logical principle in common to both proofs is known as the principle of *contraction*. In both the liar and the Curry proof, two occurrences of the one assumption are discharged at once. In the liar reasoning, the assumption $T\lambda$ is made twice, to prove the contradiction $\bot$. Both instances are discharged, to conlcude $\neg T\lambda$. For the Curry paradox, the assumption $c\ \varepsilon\ c$ is made twice, to prove p. Both instances are discharged, to prove $c \in c \to p$. If we restrict assumption discharge by demanding that only one assumption be discharged at any time, then these paradoxical proofs would fail.

Not only is contraction present in both of our paradoxical derivations: it can be shown that if we prohibit it, then adding the T and ε rules to our proof rules for the connectives and quantifiers *cannot* result in a proof of a contradiction or of an arbitrary formula. A more detailed analysis of proofs can show that in the absence of contraction, there is no *no* proof of contradictions using the logical rules and the T or ε rules. Proofs using the T rules and the ε rules may be *normalised*, using these reductions:

$$\dfrac{\dfrac{\begin{array}{c}\Pi\\ A\end{array}}{T\ulcorner A\urcorner}\,TI}{A}\,TE \quad\rightsquigarrow\quad \begin{array}{c}\Pi\\ A\end{array} \qquad\qquad \dfrac{\dfrac{\begin{array}{c}\Pi\\ A(t)\end{array}}{t\ \varepsilon\ \{x : A(x)\}}\,\varepsilon I}{A(t)}\,\varepsilon E \quad\rightsquigarrow\quad \begin{array}{c}\Pi\\ A(t)\end{array}$$

In these reductions, the intermediate formulas (here $T\ulcorner A\urcorner$ or $t\ \varepsilon\ \{x : A(x)\}$) may be no more complex than the formulas on either side. In the case where $t\ \varepsilon\ \{x : A(x)\}$ is $c\ \varepsilon\ c$ (from the Curry paradox) the formula is inferred from $c\ \varepsilon\ c \to p$ which is *more* complex than the introduced formula. Reducing the proof does not involve cutting out a local maximum in complexity. The reduction simplifies the proof by making the proof strictly smaller.

The truth and membership rules, then, are well behaved from a proof-theoretic perspective on one measure: normalising these detours shrinks the proof. In the presence of contraction, normalisation steps sometimes *enlarge* proofs, but if we ban contraction, then simplifying a detour *always* shrinks a proof, and so, we can totally eliminate detours in proofs, even in the presence of the T and ε rules.

So, without contraction we know that if there is a proof (even using our T or ε rules) for $X \succ A$, then there is a proof for $X \succ A$ with no detours. However, it is not difficult to show that normal proofs satisfy the subformula property. So, there can be no paradoxical derivations of arbitrary conclusions using these rules, since there is no normal proof of an atomic formula p (or $\bot$) from no premises. Since p and $\bot$ have no subformulas at all, no introduction or elimination rules could feature in any such proof satisfying the subformula property. So, the addition of our truth or class rules cannot interact with our logical vocabulary in this devastating way, if we reject CONTRACTION.[12]

* * *

It is worth saying *something* about the connection between contraction and the sorites paradox. Again, as with ST, the response to the paradox is not that the sorites paradoxical argument as stated above is invalid. As before with ST, also in the absence of contraction, →E remains *valid*. However, this does not mean that the distinctions drawn in the absence of contraction are without use in diagnosing the oddity of the sorites reasoning. For though the argument is valid, this does not mean that the sorites *conditional*:

$$p_1 \wedge \bigwedge_{i=1}^{9\,999} (p_i \to p_{i+1}) \to p_{10\,000}$$

must be true, for here, the antecedent formula must be used *many* times to obtain the consequent, and *demanding* that this come for free from the validity of the *argument* is equivalent to applying contraction *many* times. Though the argument is *logically* valid, this does not mean that granting each of the premises individually (which might not be equivalent to granting them *together*) must be enough to justify granting the conclusion. More must be done here to develop a coherent picture of the semantics of vague predicates on contraction-free lines understood proof-theoretically, but there seems to be some hope the distinctions drawn on a contraction-free account of logic might be a worthwhile perspective on vagueness.[13]

5. ON ANTI-EXCEPTIONALISM: FURTHER SUBDIVIDING THE TERRITORY

So far, this essay has examined the significance of responses to the paradoxes for the development of *logic*. Either commonly accepted logical principles are kept fixed, and difficult decisions are made concerning the behaviour of truth, set/class membership, vague predicates,and more, or the paradoxes are taken to be a motivation for the revision of commonly accepted logical principles. Some who take the

[12] It is worth mentioning here that the normalisation of proofs involving the naïve T or ε rules is at the heart of yet another account of the paradoxical concepts, Neil Tennant's account of relevance and paradox in his *Core Logic* [69, 70]. For Tennant, logical consequence is *relevant* in the following sense. If the argument from X to A is valid in the Core Logic sense, then each member of X and A pays its way: for no proper subset X' of X is the argument from X' to A valid, and neither is the argument from X to $\bot$ (when A is not $\bot$, anyway). Relevant validity in this sense is not transitive, since from $p, \neg p$ we can prove $\bot$ and from $\bot$ we can prove q, but it is not the case that from $p, \neg p$ we can prove q. On this account, when it comes the Curry paradoxical proof, we can successfully derive $c \mathrel{\varepsilon} c \to p$ in the left branch of our proof, and we can successfully derive $c \mathrel{\varepsilon} c$ in the right branch, as these are both normal proofs. However, adding these two derivations together in an elimination step constructs a non-normal proof of p, and this is the step that Core Logic rejects. Proof composition fails at the steps that can create non-normal proofs, and the paradoxical derivation is rejected at the final →E step.

[13] For a start on such an exploration, I recommend John Slaney's "A Logic for Vagueness" [62].

revisionary path do so by proposing different *models*, providing counterexamples to traditional logical principles, while others propose a different understanding of *proof*, which attempt to isolate the error in the paradoxical derivation in some logical or structural principle. Each such approach, whether conservative or revisionary, brings with itself commitments concerning the roles of proofs and models in the development of the metatheory of logical consequence, and the significance of proofs and models for questions of meaning and of epistemology. However, these are not the only ways that the paradoxes have philosophical significance: the very idea of proposing some *revision* of logical theories raises important questions concerning the relationship between logic, semantics, epistemology and methodology [10, 50, 60].

One way to reflect on the norms governing potential logical revision is to relate our topic to the contemporary discussion of whether logic is to be understood as a distinctive body of knowledge with its own norms, or whether it is continuous with the rest of science. This is the debate over exceptionalism and *anti-exceptionalism* [26,52,74]. Anti-exceptionalists understand the development of logic to be of a piece with the development of other scientific theories, and so, the epistemology of logic is not to be understood as exceptional in any way. Just as scientific theories are regularly revised in the face of recalcitrant evidence, logical theories may be revised in the same way, and just as scientific claims are justified in a holistic manner, logical claims may be justified on holistic grounds, as the best available account of the phenomena under consideration. As Hjortland describes anti-exceptionalist accounts of logic, they not only take logical theories to be continuous with the rest of the sciences, and justified in the same way as other scientific theories, they also, as a result, reject the traditional account of logical principles as analytic and *a priori* [26, p. 631], since claims of logic are open to revision and may perhaps be justified only on abductive grounds.

It is natural to think that in the debate over the paradoxes, it is the revisionists who would be anti-exceptionalists, and the conservatives, exceptionalists. This is not the case. The boundary between exceptionalists and their rivals is orthogonal to the boundary between conservatives and revisionists in response to the paradoxes. I will end this essay by explaining this difference between the distinctions, in order to clarify the possible lessons we might learn for the connections between logic and epistemology.

* * *

We can see that not all logical 'conservatives' are exceptionalists by attending to the case of Timothy Williamson, who is not only a paradigm case of an anti-exceptionalist, but one who has used their account of the abductive justification of logical principles to argue that one should be conservative, rather than revisionary, about logical principles in the light of the paradoxes. Williamson's conservative response to the paradoxes is motivated by his distinctive understanding of what logic *is* [76–79]. Williamson proposes a deflationism about logic: logical laws are general truths, in which all non-logical content is generalised away. To say that the law of non-contradiction holds is not, at first, or at root, to say that contradictions are not true *of necessity*, or that it is *provable* $\neg(p \wedge \neg p)$, but rather, simply, that $\forall p \neg(p \wedge \neg p)$. This is a general truth in which any non-logical content has been universally quanitified away. (See his "Is Logic about Validity?" in this volume.) Logical truths are simply true generalities, and our methods for discovering which generalities hold is not a distinctive faculty of deduction or reflection on meaning or necessity, but encompasses all the methods we use to ascertain truths of whatever sort. Logical truths are not truths about a special domain: they are everyday truths, but with a very wide generality.

It follows that logical truths understood in this way, being maximally general, and in a sense, about everything whatsoever, play a special role in our theorising. If we have to choose between revising a view about *logic* and a view about, the behaviour of a single predicate (the truth predicate, for example), the more manageble theoretical change is to revise the latter, rather than the former, since it will be connected to fewer commitments [79]. So, Williamson argues that an abductive methodology for logic leads to a logically conservative response to the paradoxes.[14]

Of course, anti-exceptionalists *can* be revisionists concerning logic. Graham Priest's defence of LP, discussed above, is motivated on abductive grounds [46, 49]. While Priest's account of logic differs substantially from Williamson's, in that Priest takes it that a logical theory is fundamentally about deductive *validity*. (See Chapter 10 and 11 of Doubt Truth to be a Liar [48, Chapters 10 and 11] for a presentation of Priest's methodology.) Alongside this different account of what a logical theory is fundamentally about, comes a different judgement about the costs and benefits of logical revision. If logic is about *validity*, it is unsurprising that revising our account of this one notion in response to the paradoxes may seem, at least *prima facie* more manageable than revising our theories of truth, reference, property ascription, class membership, vague predicates, and so on.[15]

* * *

Let's turn to the other side of the coin, to consider *exceptionalism*. What could make logic *not* continuous with the sciences? It is not to say that logic is wholly unlike the sciences, for logic (like mathematics), plays some kind distinctive role as a part of scientific theorising. When we talk about logic we have some sense of what we are talking about, and it is distinct from when we are talking about other

[14] Note, although Williamson's view is 'conservative' in that he takes the lesson of our paradoxical arguments to be a revision of our pre-theoretical commitments about *non*-logical notions, rather than logic, his view is logically *revisionary* in another sense. For Williamson, since $\exists x \exists y\, x \neq y$ is true, and since it contains no non-logical terms, it is a thesis of logic. This is a non-classical logical commitment, in that the view departs from the judgement of classical first-order predicate logic.

[15] Space does not permit me to discuss other anti-exceptionalist revisionists further, but it seems to me that the epistemology of logic for other revisionist programmes of Zach Weber [72, 73], and Hartry Field [16] both count as broadly anti-exceptionalist, though the details in each case differ significantly.

features of theorising. Here, the relevant distinction between the exceptionalist and the anti-exceptionalist is one of the epistemology of logic, and whether logic has any distinctive epistemic role to play in theorising.

Here we must be careful, however, to distinguish between different kinds of justification and different kinds of claims. Of course, the development of theories *about* logic are scientific theories in their own right, but the epistemic status of a theory *of* logic (which may well use abductive justifications, like any other form of justification in the development of a scientific theory) is not the same thing as the epistemic status of any of the individual deliverances of that theory. The epistemic status of some formalisation of of a mathematical theory as a whole is not the same thing as the epistemic status of the judgement that $2+2=4$; the epistemic status of someone's deduction of the conclusion q from the premises p and *if* p *then* q is not the same as the status of their theory *of* logic, if indeed they have a theory of logic. The epistemic justification of a theory as a whole is not to be identified with the justification of some claim that may happen to be a part of the theory, but which could be made independently of the rest of the theory. With that scene-setting done, what can we say about exceptionalist views of logic?

One well-worked-out contemporary framework which finds a special epistemic role for logical concepts is that of the intuitionist type-theorist, Per Martin-Löf. On his account of logic, core logical concepts like the connectives, quantifiers, identity and the like are *defined* by means of their inference rules, and a proof from some premises to a conclusion gives you the means to transform warrant from the premises into warrant for the conclusion [36–38]. Here, the account of logic is one that takes the inference rules to be *definitions* of the concepts involved, and that takes a proof of a claim to provide the means to know the conclusion *a priori*. For theorists like these, the definitions ground and justify the inference rules used in the paradoxes, and on this account, they are definitions. The paradoxical derivations are evidence that we cannot take the rules for truth (or class membership) as defining those concepts, since their addition is *not* a conservative extension to our underlying vocabulary, and so they cannot count as a definition, unlike the logical concepts which do provide a conservative extension. This is a very different account of the epistemology of inference, at the level of the individual *proof*, rather than restricting the epistemology of logic to the level of the logical theory as a whole.[16]

While it is most easy to understand the scope for exceptionalism on epistemic grounds for those who take proof to play a distinctive epistemic role, it is possible for someone who as a representationalist or model-theoretic semantics to be *equally* exceptionalist concerning logical validity. If, for example, you had a view

[16]The distinction here is a sharp one. Neither Williamson nor Priest give an account of the role of *proof* in demarcating the boundary of what counts as logic. In Williamson's defence of the logical validity of $\exists x \exists y\ x \neq y$ there is no account of what it might be to *prove* $\exists x \exists y\ x \neq y$ in some system of proof which is sound and complete for this notion of logic. Priest, too, gives short shrift to proofs as a means for characterising logical consequence, and he does not look to proof as having any distinctive epistemic role [48, §11.3].

of propositional content according to which propositions were (or were correctly modelled by) sets of possible worlds,[17] such that *any* set of possible worlds counted as a proposition [33, 64], then it seems like a very quick step from here to considering a notion of entailment as *defined* by way of subsethood between propositions, and logical connectives as defined by the set theoretic operations of intersection, union and complement, so that the traditional logical notions count as wholly analytically definable on some underlying structure.

However we understand this kind of exceptionalism, and whatever its prospects, it seems clear that it fits naturally with the conservative response to the paradoxes, since logical revision is not at stake. However, even our last option—the exceptionalist revisionist—is not only conceptually possible, but has been defended in recent years. This is easiest to see in the case of the proof-first epistemology of Martin-Löf. You might be drawn to general picture, according to which the proofs generate warrants, and allow for the a priori knowledge of the conclusions, but think that nonetheless, the paradoxes show us that we have misidentified exactly what the deliverances of logic are. In the light of the non-classical response to the paradoxes, you may wonder whether Martin-Löf's commitment to the structural rules of contraction, cut and identity, is itself unproblematic. In the light of the paradoxes, you may reject one of these principles, and adopt an account of proof differing from the orthodoxy.

Uwe Petersen has adopted position of this form [41, 42, 44]. For Petersen, the paradoxes show us that the contraction rule cannot be unrestrictedly valid, and that on a contraction-free understanding of the structural basis of inference, the logical notions can be introduced by *definition* by their inference rules. Along with Martin-Löf, Petersen can agree that the logical concepts are given by definition, and that logical deduction has itsa own epistemic power. In his discussion of what is involved in rejecting the classical account of the behaviour of logical connectives, Petersen isolates the distinctive role of the structural rules. "My point is that while the theoretical constants may well be something like "free conventions", the structural rules are not. They contain ontological assumptions." [43, p. 1593] According to Petersen, we discovered that the contraction rules fail. The paradoxical derivations are (if we treat them correctly) evidence enough for this conclusion. However, the logical concepts (negation, conditional, quantifiers, etc.) may still be understood as freely and conventionally introduced relative to the underlying basis of the structural rules.[18]

Perhaps the a similar account of the exceptional nature of logic, *despite* logical revision, could be told model-theoretically, though this is not simply a matter of

[17] Or states or scenarios or points or whatever else—the details of what these items are does not matter here.

[18] As far as I can ascertain it, the research team of Pablo Cobreros, Paul Egré, David Ripley and Robert van Rooij [12, 58] do not take a stand the exceptionalism debate, but there is a way of defending their position that takes the paradoxes to be good grounds to reject the *Cut*, but then to take the logical vocabulary to be analytic in the sense of being totally determined by their defining inference principles, relative to this basis.

taking propositions to be three-valued analogues to the implicitly two-valued logic of propositions understood as sets of possible worlds. For logical principles to be understood as arising out of underlying structure by means of definition, you need to not only explain why the underlying structure is one in which paradox-resistent concepts (like the logical concepts given a three-valued semantics above) *can* be defined, but you also must give an account of how paradox prone concepts *cannot* be defined.[19] (Perhaps the fixed-point construction discussed above, and the role of the specificity ordering '$\sqsubseteq$' might give some insights here, but I must leave this for another time and place.)

Regardless of whether a revision is to be understood by way of models or by way of proofs, the resulting perspective on how the revision is to be understood is analogous to another kind of revision of a theory that had been understood to have an *a priori* epistemic status, the revision from Euclidean to non-Euclidean geometry in the 19th and 20th Century. One response to such a revision is to admit that geometric claims are not knowable *a priori*, but are all subject to *a posteriori* justification in the light of observation. However, such a response to the revision of geometric claims is altogether too quick. After all, even though in the light of evidence it was clear that the parallel postulate might turn out to be false in physical three-dimensional space, that does not mean that all geometric judgements must suffer the same fate. It might be that some geometric judgements (for example, the claim that for any two points there is exactly one line passing through those points) are knowable *a priori*, because they help constitute the concepts of point and line. Such was the approach of the neo-Kantians and their account of the nature of spatial judgements [20, p. 71]. On this account, some spatial judgements were synthetic *a priori*, even when the mathematicians showed us that non-Euclidean spaces were coherent and consistent, and Einstein, further, showed that space might actually *be* non-Euclidean. The neo-Kantian account of this conceptual revision proceded in this way: The evidence (in this case, the findings of science, as per Einstein) showed that their original account of what we could know *a priori* was mistaken. Upon reflection, in the light of the new evidence, they agreed that the parallel postulate was *not* as fundamental as the others, and that it was not ever known *a priori*, since we now see that it might well be false. The other postulates still, though, are still known *a priori* and help constitute what it was to be able to make spatial judgements, and play a role in identifying what it is for judgements to be about points and lines and not about something else.

A similar account, then, might be made by the exceptionalist but revisionist logician. The structural properties governing proof and deduction arise out of the norms governing assertion, denial, and inference, or whatever is involved in the practice of making judgements at all. We might *think* that all of the structural rules

[19] It is not enough to say that a proposition is (modelled by) a tripartite division on states. You must explain why not every such tripartite division counts as a proposition, since if it *does*, then for any proposition, we could define its *exclusive* negation, a proposition that is valued 1 just where the original proposition is not valued 1, and the liar paradox would return, when couched in terms of exclusive negation.

of identity, weakening, contraction and *Cut* hold for this account of deduction (just as we might think that the parallel postulate hold for spatial judgements) but we learn in the light of the paradoxes that either *Cut* (for the defender of ST) or *contraction* (for the contraction-free theorist) or some other structural rule may be jettisoned coherently. The traditional rules for the connectives and the quantifiers are definitional for the concepts they introduce, and the paradoxes (both semantic paradoxes and the sorites) show us that while we might have been led to believe[20] that the rules of *Cut* and *contraction* hold in virtue of the very idea of a practice of assertion and denial, these are extra commitments about the practice of deduction may be rejected. On this view, there remains a principled boundary between the domain of the logical, which is analytic and topic-neutral, and the non-logical, while something that had been *thought* to be a princple of logic is rejected.

REFERENCES

[1] RODERICK BATCHELOR. "Topic-Neutrality". *Mind*, 120(477):1–9, 2011.

[2] JC BEALL. *Spandrels of Truth*. Oxford University Press, 2009.

[3] JC BEALL AND DAVID RIPLEY. "Non-Classical Theories of Truth". In *The Oxford Handbook of Truth*. Oxford University Press, 2018.

[4] NUEL D. BELNAP. "How a Computer Should Think". In G. RYLE, editor, *Contemporary Aspects of Philosophy*, pages 30–55. Oriel Press, 1977.

[5] MERRIE BERGMANN. *An Introduction to Many-Valued and Fuzzy Logic: Semantics, Algebras, and Derivation Systems*. Cambridge University Press, Cambridge; New York, 2008.

[6] S. BLAMEY. "Partial Logic". In D. GABBAY AND F. GUENTHNER, editors, *Handbook of Philosophical Logic*, volume III. D. Reidel, Dordrecht, 1986.

[7] ROSS T. BRADY. "The consistency of the axioms of abstraction and extensionality in a three-valued logic". *Notre Dame Journal of Formal Logic*, 12:447–453, 1971.

[8] ROSS T. BRADY. "The Non-Triviality of Dialectical Set Theory". In GRAHAM PRIEST, RICHARD ROUTLEY, AND JEAN NORMAN, editors, *Paraconsistent Logic: Essays on the Inconsistent*, pages 437–470. Philosophia Verlag, 1989.

[9] ROSS T. BRADY. *Universal Logic*. CSLI, Stanford, 2006.

[10] MATTHEW CARLSON. "Logic and the Structure of the Web of Belief". *Journal for the History of Analytical Philosophy*, 3(5):1–26, 2015.

[11] PABLO COBREROS, PAUL EGRÉ, DAVID RIPLEY, AND ROBERT VAN ROOIJ. "Tolerant, Classical, Strict". *Journal of Philosophical Logic*, 41(2):347–385, 2012.

[12] PABLO COBREROS, PAUL EGRÉ, DAVID RIPLEY, AND ROBERT VAN ROOIJ. "Reaching Transparent Truth". *Mind*, 122(488):841–866, 2013.

[13] HASKELL B. CURRY. "The Inconsistency of Certain Formal Logics". *Journal of Symbolic Logic*, 7:115–117, 1942.

[14] J. MICHAEL DUNN. "Partiality and its Dual". *Studia Logica*, 65:5–40, 2000.

[15] MATTI EKLUND. "Deep Inconsistency". *Australasian Journal of Philosophy*, 80(3):321–331, 2002.

[16] HARTRY FIELD. *Saving Truth From Paradox*. Oxford University Press, 2008.

[17] KIT FINE. "Vagueness, Truth and Logic". *Synthese*, 30(3–4):265–300, 1975. Reprinted in *Vagueness: A Reader* [30].

[20] As Restall was led to believe in 2005 [54].

[18] SUKI FINN. "The Adoption Problem and Anti-Exceptionalism about Logic". *The Australasian Journal of Logic*, 16(7):231, 2019.

[19] ROHAN FRENCH. "Structural Reflexivity and the Paradoxes of Self-Reference". *Ergo, an Open Access Journal of Philosophy*, 3, 2016.

[20] MICHAEL FRIEDMAN. *Dynamics of Reason: the 1999 Kant Lectures at Stanford University*. CSLI Publications, Stanford, 2001.

[21] PAUL C. GILMORE. "The Consistency of Partial Set Theory without Extensionality". In *Axiomatic Set Theory*, volume 13 of *Proceedings of Symposia in Pure Mathematics*, pages 147–153, Providence, Rhode Island, 1974. American Mathematical Society.

[22] ANIL GUPTA AND NUEL BELNAP. *The Revision Theory of Truth*. MIT Press, 1993.

[23] VOLKER HALBACH. *Axiomatic Theories of Truth*. Cambridge University Press, 2011.

[24] MICHAEL HALLETT. *Cantorian Set Theory and Limitation of Size*. Number 10 in Oxford Logic Guides. Clarendon Press, Oxford, 1984.

[25] ROLAND HINNION AND THIERRY LIBERT. "Positive Abstraction and Extensionality". *The Journal of Symbolic Logic*, 68(3):828–836, 2003.

[26] OLE THOMASSEN HJORTLAND. "Anti-exceptionalism about logic". *Philosophical Studies*, 174(3):631–658, 2017.

[27] LEON HORSTEN. *The Tarskian Turn: Deflationism and Axiomatic Truth*. The MIT Press, 2011.

[28] PAUL HORWICH. *Truth*. Basil Blackwell, Oxford, 1990.

[29] DOMINIC HYDE. "From heaps and gaps to heaps of gluts". *Mind*, 106(424):641–660, 1997.

[30] ROSANNA KEEFE AND PETER SMITH. *Vagueness: A Reader*. Bradford Books. MIT Press, 1997.

[31] S. C. KLEENE. *Introduction to Metamathematics*. D. van Nostrand, Princeton, 1950.

[32] SAUL KRIPKE. "Outline of a Theory of Truth". *The Journal of Philosophy*, 72(19):690–716, 1975.

[33] DAVID K. LEWIS. *On the Plurality of Worlds*. Blackwell, Oxford, 1986.

[34] JOHN MACFARLANE. *What Does it Mean to Say that Logic is Formal?* PhD thesis, University of Pittsburgh, 2000.

[35] R. L. MARTIN AND P. W. WOODRUFF. "On Representing "true-in-L" in L". *Philosophia (Israel)*, 5:213–217, 1975.

[36] PER MARTIN-LÖF. *Notes on Constructive Mathematics*. Almqvist and Wiksell, Stockholm, 1970.

[37] PER MARTIN-LÖF. "Analytic and Synthetic Judgements in Type Theory". In *Kant and Contemporary Epistemology*, pages 87–99. Springer Netherlands, 1994.

[38] PER MARTIN-LÖF. "On the Meaning of the Logical Constants and the Justification of the Logial Laws". *Nordic Journal of Philosophical Logic*, 1:11–69, 1996.

[39] ROBERT K. MEYER, RICHARD ROUTLEY, AND J. MICHAEL DUNN. "Curry's Paradox". *Analysis*, 39:124–128, 1979.

[40] ROMINA PADRO. *What the Tortoise Said to Kripke: the Adoption Problem and the Epistemology of Logic*. PhD thesis, City University of New York, New York, 2015.

[41] UWE PETERSEN. "Logic Without Contraction as Based on Inclusion and Unrestricted Abstraction". *Studia Logica*, 64(3):365–403, 2000.

[42] UWE PETERSEN. *Diagonal Method and Dialectical Logic: Tools, materials, and groundworks for a logical foundation of dialectic and speculative philosophy, Book One: Tools for Dialectic*. Der Andere Verlag, Osnabrück, 2002.

[43] UWE PETERSEN. *Diagonal Method and Dialectical Logic: Tools, materials, and groundworks for a logical foundation of dialectic and speculative philosophy, Book Three: Groundworks for Dialectical Logic*. Der Andere Verlag, Osnabrück, 2002.

[44] UWE PETERSEN. "$\mathrm{L}^{\mathrm{i}}\mathrm{D}^{\mathrm{Z}}_{\lambda}$ as a basis for PRA". *Archive for Mathematical Logic*, 42(7):665–694, 2003.

[45] MICHAEL D POTTER. *Set theory and its philosophy: a critical introduction*. Oxford University Press, Oxford; New York, 2004.

[46] GRAHAM PRIEST. "The Logic of Paradox". *Journal of Philosophical Logic*, 8(1):219–241, 1979.

[47] GRAHAM PRIEST. "The Structure of the Paradoxes of Self-Reference". *Mind*, 103(409):25–34, 1994.

[48] GRAHAM PRIEST. *Doubt Truth to be a Liar*. Oxford University Press, 2006.

[49] GRAHAM PRIEST. *In Contradiction: a study of the transconsistent*. Oxford University Press, Oxford, Second edition, 2006.

[50] W. V. O. QUINE. "Two Dogmas of Empiricism". *The Philosophical Review*, 60(1):20–43, 1951.

[51] STEPHEN READ. "The Liar Paradox from John Buridan back to Thomas Bradwardine". *Vivarium*, 40(2):189–218, 2002.

[52] STEPHEN READ. "Anti-Exceptionalism about Logic". *The Australasian Journal of Logic*, 16(7):298–318, Nov 2019.

[53] GREG RESTALL. "Deviant Logic and the Paradoxes of Self Reference". *Philosophical Studies*, 70(3):279–303, 1993.

[54] GREG RESTALL. "Multiple Conclusions". In PETR HÁJEK, LUIS VALDÉS-VILLANUEVA, AND DAG WESTERSTÅHL, editors, *Logic, Methodology and Philosophy of Science: Proceedings of the Twelfth International Congress*, pages 189–205. KCL Publications, 2005.

[55] GREG RESTALL. "Assertion, Denial and Non-Classical Theories". In KOJI TANAKA, FRANCESCO BERTO, EDWIN MARES, AND FRANCESCO PAOLI, editors, *Paraconsistency: Logic and Applications*, pages 81–99. Springer, 2013.

[56] GREG RESTALL. *Proofs and Models in Philosophical Logic*. Cambridge University Press, 2022.

[57] DAVID RIPLEY. "Contradictions at the Borders". In RICK NOUWEN, ROBERT VAN ROOIJ, ULI SAUERLAND, AND HANS-CHRISTIAN SCHMITZ, editors, *Vagueness in Communication*, pages 169–188, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[58] DAVID RIPLEY. "Conservatively Extending Classical Logic with Transparent Truth". *The Review of Symbolic Logic*, 5(2):354–378, 2012.

[59] STEWART SHAPIRO. *Foundations without Foundationalism: A case for second-order logic*. Oxford University Press, 1991.

[60] STEWART SHAPIRO. "We Hold These Truths to Be Self-Evident: but What Do We Mean by That?". *The Review of Symbolic Logic*, 2(1):175–207, March 2009.

[61] GILA SHER. *The Bounds of Logic*. MIT Press, Cambridge, Massachusetts, 1991.

[62] JOHN SLANEY. "A Logic for Vagueness". *The Australasian Journal of Logic*, 8:100–134, 2010.

[63] NICHOLAS J. J. SMITH. *Vagueness and Degrees of Truth*. Oxford University Press, Oxford, 2008.

[64] ROBERT STALNAKER. *Inquiry*. Bradford Books. MIT Press, 1984.

[65] ALFRED TARSKI. "The concept of truth in the languages of the deductive sciences". *Prace Towarzystwa Naukowego Warszawskiego, Wydzial III Nauk Matematyczno-Fizycznych*, 34, 1933. expanded English translation in [67], pp. 152–278.

[66] ALFRED TARSKI. "The Semantic Conception of Truth: and the Foundations of Semantics". *Philosophy and Phenomenological Research*, 4:341–376, 1944.

[67] ALFRED TARSKI. *Logic, Semantics, Metamathematics, papers from 1923 to 1938*. Hackett Publishing Company, 1983. Edited by John Corcoran.

[68] NEIL TENNANT. *Natural Logic*. Edinburgh University Press, Edinburgh, 1978.

[69] NEIL TENNANT. "A New Unified Account of Truth and Paradox". *Mind*, 124(494):571–605, 2015.

[70] NEIL TENNANT. *Core Logic*. Oxford University Press, 2017.

[71] ACHILLE C. VARZI. "Supervaluationism and its Logics". *Mind*, 116(463):633–675, 2007.

[72] ZACH WEBER. *Paradoxes and Inconsistent Mathematics*. Cambridge University Press, Oct 2021.

[73] ZACH WEBER. *Paraconsistency in Mathematics*. Cambridge Elements. Cambridge University Press, 2022.

[74] JOHN WIGGLESWORTH. "Logical anti-exceptionalism and theoretical equivalence". *Analysis*, 77(4):759–767, 2017.

[75] TIMOTHY WILLIAMSON. *Vagueness*. Routledge, London; New York, 1994.

[76] TIMOTHY WILLIAMSON. *The Philosophy of Philosophy*. Blackwell Publishing, Malden, MA; Oxford, 2007.

[77] TIMOTHY WILLIAMSON. "Logic, Metalogic and Neutrality". *Erkenntnis*, 79(S2):211–231, 2013.

[78] TIMOTHY WILLIAMSON. *Modal Logic as Metaphysics*. Oxford University Press, Oxford, UK, 2013.

[79] TIMOTHY WILLIAMSON. "Semantic paradoxes and abductive methodology". In BRAD ARMOUR-GARB, editor, *Reflections on the Liar*, pages 325–346. Oxford University Press, Oxford, 2017.

[80] CRISPIN WRIGHT. "On the Coherence of Vague Predicates". *Synthese*, 30(3–4):325–365, 1975.

[81] CRISPIN WRIGHT. *The Riddle of Vagueness*. Oxford University Press, 2021.

[82] STEPHEN YABLO. "Paradox Without Self-Reference". *Analysis*, 53:251–252, 1993.