The Review of Symbolic Logic
Volume 14, Number 1, March 2021

EXCEPTIONAL LOGIC

BRUNO WHITTLE
University of Wisconsin–Madison

Abstract. The aim of the paper is to argue that all—or *almost* all—logical rules have exceptions. In particular, it is argued that this is a moral that we should draw from the semantic paradoxes. The idea that we should respond to the paradoxes by revising logic in some way is familiar. But previous proposals advocate the replacement of classical logic with some alternative logic. That is, some alternative system of rules, where it is taken for granted that these hold without exception. The present proposal is quite different. According to this, there is no such alternative logic. Rather, classical logic retains the status of the 'one true logic', but this status must be reconceived so as to be compatible with (almost) all of its rules admitting of exceptions. This would seem to have significant repercussions for a range of widely held views about logic: e.g., that it is a priori, or that it is necessary. Indeed, if the arguments of the paper succeed, then such views must be given up.

The aim of this paper is to argue that all—or *almost* all—logical rules have exceptions. In particular, I argue that this is a moral that we should draw from the semantic paradoxes.

Of course, responding to the paradoxes by revising classical logic in some way is familiar: see, e.g., Kripke (1975), Priest (1987), Soames (1999), Maudlin (2004), Field (2008) and Beall (2009). But such proposals tend to advocate replacing classical logic with some alternative logic. That is, some alternative system of rules—where, of course, it is taken for granted that these alternatives hold without exception.

The present proposal is quite different. According to this, there is no such alternative logic. Classical logic retains the status of the 'one true logic', but this status must be reconceived so as to be compatible with (almost) all of its rules admitting of exceptions.

This would seem to have significant repercussions for a range of general, widely held views about logic. For example, it is widely held that logic is a priori in the sense that if an argument is logically valid, then we can know a priori that it is truth preserving. However the arguments of the paper challenge this view. It is similarly widely held that logic is necessary, i.e., logically valid arguments are not just truth truth preserving but necessarily so. Indeed, it is common to be presented with a picture on which necessity can naturally be seen as coming in various strengths or 'grades', where it is held that logical necessity is the highest strength or grade. However, if the arguments below succeed, then it would seem that logic is not necessary at all—let alone at the highest grade.

Received: June 13, 2018.
2020 *Mathematics Subject Classification*: Primary 03A05, Secondary 03B80.
Keywords and phrases: logical rules, semantic paradox, classical logic.

doi:10.1017/S1755020320000210

The account of logic defended below is in certain respects similar to that of Hofweber (2008, 2010). According to the latter, logic is merely 'generic': its rules hold under 'normal' conditions, but they can fail under 'abnormal' ones. Although I do not put things in quite these terms, this component of Hofweber's account can certainly be regarded as shared by the present proposal. However, Hofweber's account is also 'quietist': he argues that we cannot say anything more precise about exactly where the exceptions to logic occur. At this point we part company. For the present proposal includes both a philosophical account of where these exceptions lie, and a formal theory that precisely delimits them. Further, the arguments for logical 'exceptionalism' given below are quite different from those of Hofweber.

The structure of the paper is as follows. In §1 I introduce a phenomenon that arises from the semantic paradoxes. I call this the Chrysippus phenomenon, after one of the participants in the example that I use to illustrate it. This is a phenomenon that has been observed before. However, when it is discussed in the literature, it is used to motivate a type of approach very different from that of the present paper. In particular, it is used to motivate 'contextualist' approaches to truth, according to which different tokens of the same type are used to make different statements; for example, as a result of 'true' taking different semantic values in the distinct tokens. This account of the phenomenon is certainly not logically revisionary. However, I argue that the best explanation of the phenomenon is not a contextualist one—rather, it is one that yields logical 'exceptionalism'. Thus, having given this explanation, in §2 I argue from it to logical exceptionalism. In §3 I draw out the wider implications of the proposed account of logic, specifically for the view that logic is a priori, and for the view that it is necessary. In §4 I sketch the bigger picture that this account of the Chrysippus phenomenon extends to, outlining, too, how one can give a formal theory of truth that vindicates this account. (Such a theory is given in §6.) §5 contains further discussion, including of the possibility of trying to make the proposal less radical by combining it with the view that classical logic holds strictly, i.e., without exception, at the level of propositions. I argue, however, that no such account can succeed, in virtue of the fact that there are purely propositional instances of the phenomenon that leads to logical exceptionalism. In §6 I give a formal theory that vindicates the claims of the paper. Finally, in §7, I consider alternative approaches.

There are some additional remarks that I should make by way of introduction. The first of these concern the paper's conclusion. This, as I have said, is that almost all logical rules have exceptions. This conclusion is argued for by way of a lemma to the effect that almost every rule has an instance that does not preserve truth (i.e., an instance with true premises but an untrue conclusion). I argue that the most natural characterization of this situation is that the logically correct, or valid, rules are precisely those of classical logic; it is simply that these valid rules have exceptions. These uses of 'almost', however, might already prompt a worry: am I not simply proposing a very weak alternative to classical logic, namely that consisting just of the exceptionless rules? In fact, though, what would be left if all of the rules with exceptions were removed is too impoverished to count as a genuine alternative logic. For (restricting attention for simplicity to the propositional case) the only exceptionless rules are those where this claim is in some sense vacuous: either the premises are inconsistent (so there are no instances where the premises are all true), or the rule is circular (the conclusion is one of the premises, so the truth of the premises trivially guarantees that of the conclusion). But a 'logic' consisting only of such rules would never enable us to extend our knowledge—for it

would never allow us to go from some truths to a new one. We would be much better off, surely, making our peace with the idea that logical rules admit of exceptions; especially if, as I argue, there is a natural way of developing this idea.

There is another rival characterization that is perhaps less easily dismissed, but that is also compatible with the main claims of the paper. According to this, the answer to the question 'which logic is correct?' is: classical logic minus the exceptions. That is, the valid arguments are those that, as a matter of fact, preserve truth. As I explain, my reason for preferring the answer 'classical logic' is that it retains the idea that the validity of an argument is a matter of its 'form'; on the rival, in contrast, we must look at the particular nonlogical terms an argument contains to determine whether it is valid. Nevertheless, this rival is compatible with much of what I say here: for example, the challenges to the widely held views about the a priority and necessity of logic mentioned above.

There is something in the above that warrants emphasis. I use 'valid' and 'logically correct' as synonyms. These pick out a property that it is, essentially, the main aim of this paper to give an account of. Thus, in particular, I do not use 'valid' to mean: preserves truth under every interpretation. For I argue that the rules of classical logic are valid, despite sometimes failing to preserve truth.

A final remark is as follows. This is a paper about the semantic paradoxes. The aim is not, however, to 'solve' these, in the sense of showing at exactly which point the arguments of the paradoxes go wrong. I of course have a view about this—and it is essentially implicit in the arguments below. However, the emphasis of the paper is not on this question. The aim of the paper is, rather, to consider a phenomenon that the paradoxes give rise to, and to argue that the best explanation of this seems to lead to the striking result that (almost) all logical rules have exceptions.

§1. The Chrysippus phenomenon. Suppose that at some time t Zeno utters: What Zeno says at t is untrue. (Call this utterance Z.) It seems that Z is in some way 'pathological', and so neither true nor false. However, suppose that Chrysippus overhears Zeno and—in sympathy with the conclusion just reached—utters: What Zeno says at t is untrue. (Call this C.) It is very natural to think that C—unlike Z—is simply true. I call the apparent fact that these two tokens of the same type differ in their truth status in this way the *Chrysippus phenomenon*.[1, 2]

How is this possible? How is it possible, that is, for these two utterances to have different truth statuses, despite the fact that they contain exactly the same words? The phenomenon has certainly been noticed before. In particular, it has been used to motivate 'contextualist' approaches, according to which different tokens of the same type can make different statements; for example, as a result of 'true' taking different semantic values. Indeed, it seems generally to have been taken for granted that the only way to accommodate the phenomenon is via contextualism.[3]

[1] Gupta (2001) uses 'Chrysippus intuition'. I prefer 'phenomenon' just to keep the focus on the utterances rather than a mental state about them.

[2] I use 'truth status' rather than 'value' simply to remain neutral on the question of whether things like Z that are neither true nor false have a nonstandard true value, or instead lack one altogether. (I revert to 'value' in discussing formal constructions, but that is purely for reasons of familiarity.)

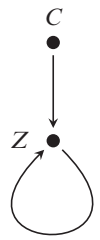
[3] For claims to this effect see, e.g., Burge (1979), pp. 176–77 and Gupta (2001), p. 113.

Of course, if 'true' does take different semantic values in different occurrences, then it is easy enough to see—in principle at least—how the Chrysippus phenomenon is possible. For example, the simplest way of ensuring this result would be by letting 'true' have partial semantic values, $\langle Z^+, Z^- \rangle$ in Z and $\langle C^+, C^- \rangle$ in C,[4] such that $Z \notin Z^+ \cup Z^-$, but $Z \in C^-$.[5]

In fact, however, the best explanation of the phenomenon does not seem to be a contextualist one. It seems rather to be as follows.

> **Structural Explanation.** The reason that Z is neither true nor false—the sense in which it is pathological—is that it is a loop: if one tries to determine whether it is true, one is sent back to that very question. In contrast, C is not a loop: if one tries to determine whether it is true, one is sent to consider whether Z is; one is not sent back to C.

That is, if one asks whether Z is true, one is sent to consider whether the utterance that Z ascribes untruth to is indeed untrue, i.e., one is sent to consider whether this utterance is true. But this utterance is simply Z, and so one is back to the question of whether Z is true—the question we started with. The situation with C is quite different: ask whether this is true, and one is sent to consider whether Z is; one is not sent back to consider whether C is. As one might put it, Z is *in* the paradox, while C is merely *about* it.



Once articulated, this explanation is apt to seem very plausible. But it does not support contextualism: it does not appeal to multiple semantic values, or multiple

[4] If $\langle S^+, S^- \rangle$ is a partial interpretation, then S^+ is the extension, i.e., the things the predicate is true of, while S^- is the anti-extension, the things that the predicate is false of.

[5] Although this is the simplest contextualist means of accommodating the phenomenon, a range of alternatives have been proposed. For example, on the approach of Gaifman (1992, 2000), Z and C have different 'meanings', or involve different 'readings' or 'interpretations' of the same sentence, but these different meanings are not supposed to result from 'true' (or any other word) taking different values in the two cases. On the approaches of Parsons (1974), Burge (1979), Simmons (1993) and Glanzberg (2001) 'true' does take different semantic values in different occurrences. However, the Chrysippus phenomenon is accommodated not by it taking different values in Z and C, but rather by Z being 'neither true nor false' in one sense of 'true' (and 'false'), and C being 'true' in another. The approach of Barwise & Etchemendy (1987) yields something close to the phenomenon—namely, Z being false while C is true—by allowing that the utterances are about different situations (i.e., portions of the world). Only the situation that C is about includes the fact that what Zeno says is untrue, and it is this that explains the difference in truth status. Although I sometimes focus on the example of the simple form of contextualism described in the text, everything that I say would also seem to apply to these different varieties.

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

statements.[6] Rather it explains the phenomenon simply by reference to the structure formed by the utterances and the relation of aboutness.[7]

It is not hard to see why the phenomenon might be thought to require contextualism: since the it entails contextualism when combined with an apparently plausible principle; namely, the following version of the principle of compositionality.[8]

> **Compositionality.** The truth status of an utterance is determined by the semantic values of the words of the utterance, together with the way that these words are combined in the utterance.

But then, since Z and C of course contain the same words, combined in just the same way, it follows from the fact that the utterances have different truth statuses that at least one of these words must take different semantic values in the two utterances. That is, contexualism must hold.

The key point, however, is that Structural Explanation provides us with a *noncompositional* account of the way in which the truth status of Z is determined. Thus, to fix ideas, consider a standard compositional explanation of the truth status of an utterance u of the form $\neg Fa$: this will explain how the truth status of u is determined by combining the semantic values of the nonlogical symbols that u contains. Thus, suppose that the semantic value of F (as it occurs in u) is the set X, while the semantic value of a (in u) is the object o. A standard compositional account of the way in which the truth status of u is determined is either: $o \in X$ and so u is false; or: $o \notin X$ and so u is true.

The account Structural Explanation provides of the determination of the truth status of Z is quite different. According to this, Z has its truth status—neither true nor false—not because the semantic values of its words can be combined to yield this in any such way. Rather, Z has this status *because* it is a loop. Thus, although Compositionality might have seemed inevitable, we in fact have a positive reason to reject it in the case of Z.

Further, the principal reason for wanting to give a compositional account of language, i.e., an account that vindicates Compositionality, is that we want to give one that is systematic. We will see, however, that the proposed account of the Chrysippus phenomenon can be developed into a (completely systematic) formal theory of truth (see §6). Thus, this reason for wanting to maintain Compositionality is also disarmed.

1.1. Tokens and types. On a contextualist account, for example, one according to which 'true' takes different semantic values in different utterances, one cannot in general ascribe truth to sentences, i.e., types: since if these contain 'true', then one must know which semantic value this is to be understood as taking, before one can

[6] Similarly, it does not fall back on the idea that the sense in which C is true is different from the sense in which Z fails to be.

[7] Or more strictly: the relation of being about an utterance's truth status. For it is certainly *not* part of the present proposal that self-referential utterances are always loops, and subsequently neither true nor false (in the way that Z is). On the contrary, many such sentences are true: e.g., tokens of 'this utterance is short' or 'this utterance doesn't use 'true''. These are not loops in the manner of Z: if one asks whether they are true, one does not find oneself returned to that very question. I say more about the notion of a loop in §2.1 below.

[8] For discussion of such principles, see e.g., Szabó (2017).

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

say whether or not the sentence is true. Rather, we must ascribe truth to tokens (or, equivalently, to pairs of sentences and contexts).

If, however, we explain the Chrysippus phenomenon in a noncontextualist way, then there is no obstacle to ascribing truth directly to types. Of course, the example of Z and C shows that (even in the absence of indexicals) we cannot in general reduce truth for tokens to truth for types. However, there would seem to be no objection to ascribing truth directly to types *as well* as to tokens.

And once we do this we see that there are instances of the Chrysippus phenomenon purely at that level. Thus, using T for truth, let λ be the sentence $\neg Tc_\lambda$, where c_λ is an individual constant denoting λ itself.[9, 10] λ is then a loop in just the way that Z is: if one asks whether this sentence is true, one is simply sent back to that very question. That is, λ is pathological in just the way that Z is. It is thus plausible that λ is neither true nor false—just like Z.

However, suppose that b_λ is a distinct constant denoting λ.[11] And let η be $\neg Tb_\lambda$. Then η—just like C—is not a loop: if one asks whether it is true, one is sent to consider whether λ is; one is not sent back to consider η. It would seem, then, that η—just like C—is straightforwardly true.

It might seem surprising that λ and η could differ in their truth status in this way. Isn't η a mere notational variant of λ? Yes. But the point is that notational variants can differ with respect to the property of being a loop (just as they can differ with respect to the property of being self-referential).

On a contextualist approach, the Chrysippus phenomenon—i.e., the fact that Z and C have different truth statuses—requires no sort of revision of logic: any more than the fact that 'I am hungry' has different statuses in different contexts requires such. However, the generalization of the Chrysippus phenomenon just mentioned certainly does give an exception to classical logic, as follows.

By $\beta_1, \ldots, \beta_n \vDash \gamma$ I mean that the truth of $\beta_1, \ldots, \beta_n$ guarantees that of γ, i.e., γ is true in every interpretation in which each of $\beta_1, \ldots, \beta_n$ are.[12] We thus have:

$$\neg Tb_\lambda, b_\lambda = c_\lambda \nvDash \neg Tc_\lambda.$$

For the sentences to the left of the turnstile are true ($b_\lambda = c_\lambda$ is a simple identity, and so is of course true), while that to the right is not. So this instance of the indiscernibility of identicals, i.e.,

---

[9] Throughout I use subscripts to indicate denotation in this way.

[10] I assume that our language contains a relatively rich supply of constants along the lines of c_λ, i.e., denoting sentences that contain the constant, or denoting sentences that themselves contain further constants that denote sentences that contain the constant, etc. This is merely for simplicity, however. One could of course instead make all of the points that follow using general syntactic resources sufficient to generate such patterns of reference, or arithmetical resources and Gödel numbering—or, indeed, even more mundane resources such as 'the sentence written on the board in room 101'.

[11] That is, if we were being strict about use and mention: $b_\lambda = c_\lambda$ but 'b_λ' $\neq$ 'c_λ'. (Just as Hesperus = Phosphorus but 'Hesperus' $\neq$ 'Phosphorus'.) However, I continue with the standard practice of using object language expressions to refer to themselves.

[12] As I said in the introduction, I ultimately distinguish validity from truth preservation in this sense (see §2.3). Nevertheless, it is convenient to follow the standard practice of using the double turnstile for truth preservation, since the main aim of the paper is to argue that almost all classical rules sometimes fail to preserve truth, and this notation gives a concise statement of such exceptions.

$$\frac{\alpha, \tau = \upsilon}{\alpha(\tau/\upsilon)},$$

which is of course classically valid, is not truth preserving.[13]

The advantage of doing things in terms of types is that logic is almost always presented in terms of these, and so by formulating our claims in this way we make their relation to standard logical claims completely straightforward. However, similar points can also be made about tokens.

§2. Logical exceptionalism. The above exception to classical logic is just the tip of the iceberg. But before giving further examples I say a bit more about the notion of a loop. (This is officially defined in 3 of §6.)

2.1. Loops. This notion is intended to capture, and generalize, what is going in the case of λ (and Z). The relevant feature of λ is that if one tries to determine whether it is true, then one is sent back to λ, and only to λ. That is, λ is about its own truth status, and not about that of any other sentence. Thus, if α is Tc_β (for some sentence β), for example, then it is about the truth status of β. Similarly if α is $\neg Tc_\beta$; or $Ws \wedge Tc_\beta$ (using W for white and s for snow); or indeed $\neg Ws \vee Tc_\beta$ or $\exists x(x = c_\beta \wedge Tx)$. Further, if β is itself about the truth status of a sentence γ, then in a natural sense α is also about the truth status of γ: if you say that what Sarah says is true, and I say that what you say is true, then, as one might put it, I have directly said something about your utterance, and by doing so indirectly said something about Sarah's. Thus, if β is Tc_γ, then if α is any of the previous examples, it will be about the truth status not only of β but also of γ; and so on.

In general, sentences are not of course about their own truth statuses (they are typically not about themselves at all). A loop, however, is a sentence that is about its own truth status, and that is not about that of any other sentence. Or, more generally, a loop is a set of sentences, each member of which is about the truth statuses of exactly the members of the set. For example, $\{\alpha, \beta\}$, where α is Tc_β and β is $\neg Tc_\alpha$.[14]

Of course, not all semantic paradoxes involve loops in this sense (e.g., Yablo's paradox does not). I discuss such paradoxes without circularity in §5.1. But, as I explain there, my reason for focusing on paradoxes *with* circularity is that it is only these that give rise to the Chrysippus phenomenon, and hence to widespread exceptions to logical rules—because, essentially, those without circularity do not support the

[13] This sort of exception to classical logic has already been countenanced: by Skyrms (1984). On that approach, the Chrysippus phenomenon is bracketed with more familiar instances of intensionality, e.g., arising from propositional attitudes or modality. It does not yield the wide range of exceptions to classical rules that the present approach does. I discuss Skyrms's approach in §7.4.1.

Another approach that involves this sort of exception to classical logic is that of Hansen (2014). This has a wider range of exceptions than that of Skyrms, but these exceptions are quite different from those of the present approach (e.g., Hansen's approach is classically inconsistent). Hansen's approach is discussed in §7.4.2.

[14] The official definition of §6 restricts the notion to sets. But I continue to describe single sentences (such as λ) as loops, where this can can be seen as shorthand for the claim that the sentence's singleton is a loop.

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

distinction between sentences that are in the paradox, versus those that are merely about it.[15]

2.2. Exceptions. To illustrate the further exceptions to logical rules that the proposed explanation of the Chrysippus phenomenon gives rise to, consider conjunction introduction and elimination.

Start with the latter, i.e., $\alpha \wedge \beta \,/\, \alpha$. We get an exception to this by considering $\lambda \wedge Ws$. Just like C and η, this conjunction is not a loop (and nor does it belong to one): while it is about the truth status of λ, it is not about its own truth status. It would appear, then, that just like C and η, it is simply true. For what the conjunction says (i.e., that λ is untrue and snow is white) is the case; and since it is not a loop this seems sufficient for it to be true (just as with C and η). Thus

$$\lambda \wedge Ws \not\models \lambda$$

which is of course an exception to conjunction elimination.[16]

The fact that this rule can fail to preserve truth—i.e., the fact that there are true conjunctions with untrue conjuncts—is perhaps surprising. Indeed, it might even be held that such a thing is incoherent. In response, though, we should observe that the connection between the truth of a conjunction, and that of its conjuncts, is less intimate than one might have been tempted to think.

Thus, consider for example $Ws \wedge Gg$ (using Gg for grass is green). One might initially think that what this sentence says—its truth condition—is: Ws is true and Gg is true. This is a thought that is encouraged by standard definitions of truth, inspired by the work of Tarski. For these define truth for conjunctions in terms of that for conjuncts.

On reflection, however, it should be clear that the connection between what conjunctions say, and the truth of their conjuncts, is nothing like this tight. For what $Ws \wedge Gg$ says is: snow is white and grass is green. It is certainly not *about* Ws and Gg, i.e., these linguistic items. To drive this point home, we can observe that this conjunction surely says the same thing as its translation into English: snow is white and grass is green. But it would be quite wrong to think that this English sentence is about these formal sentences, introduced for the purposes of this paper. Once this is

[15] I should flag the main respect in which the sketch of this subsection simplifies the official definition of loops: the latter is relative to a given partial interpretation of the truth predicate, where this represents a stage in the construction of the interpretation that we ultimately propose. Our official definition restricts attention to sentences that are in neither the extension nor the anti-extension of the relevant partial interpretation, i.e., sentences that are yet to be assigned a truth status. A loop is then a set of such sentences, each member of which is about the truth statuses of exactly the such sentences that belong to the set. This restriction is needed if we are to treat, e.g., $\alpha = \neg Tc_\alpha \wedge Tc_{Ws}$ in essentially the same way as λ, which is surely desirable. Until §6 I focus exclusively on examples where the relativity of the official definition can safely be ignored. However, even once it is taken into account, we can recover an absolute notion: just say that something is a loop in the absolute sense if it is one at some stage of the construction.

[16] It is sometimes convenient to put things, as I just have, in terms of whether what a sentence says is the case. For the purposes of this paper, however, this is simply a rhetorical device. The claim that what Ws says is the case, for example, can without essential loss be replaced by: Ws says that snow is white, and snow is white; and similarly elsewhere.

appreciated, there does not seem to be any objection of principle to a true conjunction with an untrue conjunct.[17]

Another way to put the point is this. What *would* be incoherent would be for a conjunction to be true, without what its conjuncts say being the case. For example, for $Ws \wedge Gg$ to be true without snow being white, or without grass being green. In the case of our exception to conjunction elimination, however, nothing like that occurs. What the conjuncts of $\lambda \wedge Ws$ say, i.e., that λ is untrue and that snow is white, is indeed the case. The rule fails simply because λ is untrue, despite what it says being the case.

What now about conjunction introduction: $\alpha, \beta \ / \ \alpha \wedge \beta$? Our previous examples of logical exceptions both involved the standard liar sentence λ. But since that is not a conjunction, it won't give a counterexample to this rule. Instead, let ζ be $\neg Tc_{\zeta} \wedge Ws$. In this case, ζ is a loop, just as Z and λ were: it is about its own truth status, and not about that of any other sentence. Consequently, ζ, just like Z and λ, is neither true nor false.

On the other hand, $\neg Tc_{\zeta}$ is not a loop: it is about the truth status of ζ, not about its own. Just as with previous examples, then, $\neg Tc_{\zeta}$ is simply true, giving: $\neg Tc_{\zeta}, Ws \nvDash \neg Tc_{\zeta} \wedge Ws$.

What this example turns on is the fact that the notion of a loop is defined purely in terms of the relation that a sentence stands in to those whose truth status it is about. An alternative course would be to work with a broader notion of 'dependence' that is the union of this relation and that which sentences stand in to their sentential components, their instances, and so on.[18] This approach would in a natural sense be 'upwards strict': if α and β are each true or false, then so would be $\neg\alpha$, $\alpha \wedge \beta$, etc. Thus, while we would still have exceptions to the indiscernibility of identicals and conjunction elimination, for example, we would no longer have one to conjunction introduction. For neither λ nor ζ would any longer count as loops, i.e., on their own; rather, the loops would instead be $\{\lambda, Tc_{\lambda}\}$ and $\{\zeta, Tc_{\zeta}\}$.

I consider this approach in §7.2, but there seem to be two reasons for preferring that which I have focused on. Firstly, sentences that are neither true nor false are in a sense failures. On the alternative, then, more sentences are classified as failures: e.g., Tc_{λ} as well as λ, Tc_{ζ} as well as ζ. But if we can avoid counting these additional sentences as failures—as we will see that we can—it seems desirable to do so. However, a deeper reason for preferring the approach that I have focused on is this. The relation that a sentence stands in to those whose truth statuses it is about is simply very different from that which it stands in to its components and instances (and so on). The former is a semantic relation—it is a function of what the sentence says—while the latter is merely syntactic—one doesn't need to know how a sentence's predicate symbols and constants are interpreted, for example, to know what its components and instances are. Thus, if we define loops, as on the alternative, in terms of the union of these relations, we

[17] It might be objected that we should view our formal sentences are mere abbreviations of English ones. In that case, appeal to the English conjunction might be thought unpersuasive: why not regard each conjunction as being about 'snow is white' and 'grass is green'? Essentially the same point, however, can still be made by appealing to, for example, the Japanese translation of $Ws \wedge Gg$. That Japanese sentence is surely not about English sentences. But since it is about the same things as $Ws \wedge Gg$, that cannot be about these either.

[18] The sentential components of $\neg\alpha$ are α together with its such components; if $*$ is a binary connective, then those of $\alpha * \beta$ are α and β together with theirs.

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

are really running together two quite different things. Our notion of a loop, and our overall account based on it, would thus seem to be less natural if we were to do this.

We have, then, exceptions to both conjunction introduction and elimination. But I promised exceptions to 'almost all' logical rules. Focusing initially on propositional logic, that promise is kept by the following result.

THEOREM. *Let* $\alpha_1, \ldots, \alpha_n, \beta$ *be propositional formulas such that*

$\beta \neq \alpha_i$ *for* $i = 1, \ldots, n$,
$\{\alpha_1, \ldots, \alpha_n\}$ *is classically consistent*.

Then $\alpha_1, \ldots, \alpha_n \vDash \beta$ *has a false instance.*

Here propositional formulas are used schematically to express claims about sentences of our language. Thus, by an instance of $\alpha_1, \ldots, \alpha_n \vDash \beta$ I mean a claim of the form

$$\alpha'_1, \ldots, \alpha'_n \vDash \beta'$$

where $\alpha'_1, \ldots, \alpha'_n, \beta'$ result from uniformly substituting sentences of our language for propositional symbols in $\alpha_1, \ldots, \alpha_n, \beta$ (i.e., substituting every occurrence of a given propositional symbol in these formulas for the same sentence). The argument of the theorem yields an instance where $\alpha'_1, \ldots, \alpha'_n$ are true but β' is not.

This result substantiates the 'almost all' (in the propositional case). For given the result, the only rules that are always truth preserving are those where this claim is in some way vacuous: either because the conclusion is simply one of the premises (i.e., the first condition of the theorem is not met), or because the premises cannot all be true (the second condition is not met). In every other case, we get an exception.

This is a much wider class of exceptions than we get on extant non-classical approaches. Consider, for example, the most celebrated such approach: the strong Kleene theory of Kripke (1975).[19] This theory yields exceptions to all classical propositional logical truths, but it certainly does not give exceptions to classical rules such as conjunction introduction and elimination, modus ponens, double negation introduction and elimination, and so on. Given the above theorem, however, we get exceptions to all of these on the present proposal. This proposal can be similarly contrasted with other extant nonclassical approaches, such as those mentioned in the introduction.

The result is proved in §6, as theorem 2. But the basic idea is as follows.

Proof Sketch. If $\alpha_1, \ldots, \alpha_n \,/\, \beta$ is not classically valid, then of course we can find a false instance. So suppose that it is. Let $p_1, \ldots, p_m$ be the propositional symbols that occur in $\alpha_1, \ldots, \alpha_n, \beta$, and let r be a row of the truth table for $\alpha_1, \ldots, \alpha_n, \beta$ in which each is true. For $\gamma \in \{\alpha_1, \ldots, \alpha_n, \beta\}$, γ' is the result of making the following substitutions in γ: if p_i is true in r, then replace p_i with $\neg Tc_{\beta'} \wedge i = i$; otherwise, replace it with $Tc_{\beta'} \wedge i = i$. Then β' is a loop and so neither true nor false, while each α_i is true. Thus $\alpha'_1, \ldots, \alpha'_n \nvDash \beta'$. □

Similar claims hold in the first-order case. To illustrate, consider existential generalization: $\varphi(x/\tau) \,/\, \exists x \varphi$. To get an exception to this, let ξ be $\exists x(x = c_\xi \wedge \neg Tx)$.

[19] Kripke of course declares himself 'amazed' [1975: 700] that anyone would regard his proposal as a revision of classical logic. Nevertheless, most readers have found this to be the most natural way to take his proposal.

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

ξ is then a loop, and so neither true nor false. On the other hand, $c_\xi = c_\xi \wedge \neg Tc_\xi$ is not a loop; rather it is simply true, giving the desired exception. Other examples can easily be produced, but I do not in this paper give a general result analogous to that above concerning the propositional case.

All of the exceptions that I have given involve sentences that, in one way or other, say of themselves that they are untrue. However, not all sentences that are loops in the relevant sense do this. For example, sentences that say of themselves that they are true, e.g., $\theta = Tc_\theta$, are also loops. Do other such loops generate logical exceptions? Yes, but only some of them.

For an example of a sentence that gives an exception, but that does not ascribe untruth to itself, let $\mu = Tc_\mu \vee \neg Tc_\mu$. This is a loop, and so gives an exception to the law of excluded middle, $\alpha \vee \neg\alpha$. In contrast, the truth teller θ does not yield an exception to classical logic. That is, there is no classically valid argument $\alpha_1, \ldots, \alpha_n \,/\, \theta$, where $\alpha_1, \ldots, \alpha_n$ are true despite the fact that θ (being a loop) is not. For if $\alpha_1, \ldots, \alpha_n$ are true, then what they say is the case. But then, by the classical validity of the argument, what θ says must be the case too, which is just that θ is true: contradicting the fact that θ isn't true. The fact that θ is about truth is in a sense irrelevant here. By just the same token (using B for black) we cannot have a classically valid argument $\beta_1, \ldots, \beta_m \,/\, Bs$, the premises of which are all true. For in this case what they say would be the case, in which case (by the classical validity), snow would have to be black.

This style of reasoning might set off alarm bells: how, given that I am claiming there are exceptions to classical rules, can I rely on classical validity in this way? The point, though, is that while we can have classically valid rules that fail to preserve truth, what we cannot have is a situation where what the premises say is the case, while what the conclusion says isn't. For example, while we have $Ws \wedge \neg Tc_\lambda$ being true without $\neg Tc_\lambda$ being, we certainly do not have snow being white and λ being untrue, without λ being untrue—that would be absurd!

More generally, we get logical exceptions precisely in the cases of loops that are classically entailed by their own untruth (and 'unfalsity', together perhaps with other facts): any other loop would require its own truth or falsity, and so would not yield an exception for the same reason that θ doesn't.

We get exceptions to classical rules only in cases in which the conclusion is neither true nor false. This means that corresponding to any classical rule $\alpha_1, \ldots, \alpha_n \,/\, \beta$, there is a rule that holds strictly—i.e., without exception—on this approach. Namely, that which results from adding $T\tau_\beta \vee F\tau_\beta$ (for some closed term τ_β) to the premises.[20] This rule will in general play a role that is significantly different from that played by the original—at least as traditionally understood. Consider conjunction introduction, for example. On the traditional understanding, this allows us to go from γ and δ to a distinct claim $\gamma \wedge \delta$—even if we do not have any previous knowledge of this claim. The modified rule is quite different: it allows us to pass from γ and δ to their conjunction only if we have already established something about the latter (i.e., that it is either true or false). Nevertheless, the existence of these modifications shows that we retain a system of strict rules that 'shadow' the standard logical ones.

Although the focus of this paper is on logical rules, I should mention how things stand with the most discussed truth-theoretic principle: the truth schema, $T\tau_\alpha \leftrightarrow \alpha$,

[20] Here F means falsity. On this approach, we add a falsity predicate as well as a truth one: see §6.

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

where τ_α is a closed term denoting α. An instance of this— $Tc_\lambda \leftrightarrow \lambda$—is of course classically inconsistent, and so all instances of the schema are certainly *not* true on the present approach. This much is in common with many extant approaches. However, there are two ways in which the present approach goes beyond this. Firstly, it is not merely that some instances of the schema fail to be true: some are false. For example, that just mentioned for c_λ: similarly to $\lambda \wedge Ws$, this instance is not a loop, and so is straightforwardly false in virtue of the fact that what it says (i.e., that λ is true iff it is not) is not the case. Secondly, the sentences on the two sides of an instance of the schema need not have the same truth status. In particular, the left hand sentence can be false while the right hand one is neither true nor false (this is what we have in the case of Tc_λ and λ).[21] Just as with the exceptions to logical rules discussed above, these features of the proposal seem to be naturally and straightforwardly motivated by the Chrysippus phenomenon.

2.3. Which logic is correct?. What follows for the question of which logic is correct? For simplicity, let us again focus on the propositional case. The upshot of the above theorem is that if we insist that logically valid rules hold strictly then there are almost none of these. More precisely, the only logically valid arguments will be those that are (in one of two ways) vacuously truth preserving.

On the other hand, there remains a natural sense in which classical logic is correct, despite the exceptions. For on the proposed account, a sentence can have its truth status determined compositionally or otherwise. Thus, let G and H be unary predicate symbols whose semantic values are the sets X and Y, respectively. Then for Gc_o, for example, to have its truth status determined compositionally is for it to be true if $o \in X$, and false otherwise. For $\neg Gc_o$, it is for this to be true if $o \notin X$, and false otherwise. Similarly, for $Gc_o \wedge Hc_r$ to have its truth status determined in this way is for it to be true if $o \in X$ and $r \in Y$, and false otherwise, and so on.[22] Sentences like λ are such that if one tries to determine their truth status compositionally, one finds that the process never ends. In the case of λ in particular one finds oneself going round in a circle. Because of this, these sentences are neither true nor false; a status that is not determined compositionally in the relevant sense. It is only when dealing with this latter class of sentences that we get exceptions to classical logic. This logic thus remains correct in the following sense: its arguments are truth preserving as long as the sentences involved have their truth statuses determined compositionally.[23]

---

[21] This is the only way in which such pairs can differ in truth status, however. For if $T\tau_\alpha$ is true, then what it says must be the case, i.e., α must be true too. Further, if α is true or false, then $T\tau_\alpha$ will be true or false (respectively) too: see example 2(vi) of §6. The reason is that, given that α is true or false, then the only way that $T\tau_\alpha$ could fail to be is if it belonged to a loop; but any such loop would have to contain α, in which case α would not be true or false after all.

[22] Note that for a sentence to have its truth status determined compositionally is not necessarily for it to be determined on the basis of the statuses of subsentences (or instances thereof) in the familiar way. For example, $\lambda \wedge Ws$ has its status determined compositionally as long as it is true if λ is untrue and snow is white, and it is false otherwise. We do not however have: this conjunction is true if both of its conjuncts are, and false otherwise. (As we have seen, the conjunction is true despite the fact that its first conjunct isn't.) Given that λ is neither true nor false, to determine the truth status of the conjunction on the basis of that of λ would not really be to determine it compositionally, because this latter truth status is not itself so determined.

[23] Indeed, it is sufficient for the conclusion to have its truth status determined in this way (for a sentence whose truth status is determined noncompositionally is neither true nor false, and

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

Here is another way to put it. The exceptions to classical logic given above all involve sentences whose truth status is *not* determined by what they express in the standard way. That is, in the normal case, a sentence α says something p, and α is true if p is the case, and false if p is not the case. By saying that α has its truth status determined by what it says, I mean that it has its truth status determined in this way. Loops such as λ do not have their truth statuses determined by what they say: for λ is neither true nor false, despite the fact that what it says, i.e., that it itself is untrue, is indeed the case. This allows an alternative characterization of the status of classical logic: its arguments preserve truth as long as the sentences involved have their truth statuses determined by what they say.[24]

The natural answer to the question—which logic is correct?—would thus seem to be: classical logic. Similarly, continuing to use 'valid' to mean logically correct, the valid arguments seem precisely to be those of classical logic.

An alternative answer to the question 'Which logic is correct?' would be: classical logic minus the exceptions. That is, the logic consisting of those classical arguments that do in fact preserve truth. This answer, however, would sever the connection between the traditional logical expressions, i.e., the connectives, quantifiers and variables, and validity: for whether an argument is valid would no longer be determined by the meanings of these logical expressions. That is, an argument would no longer be classified as valid or not on the basis of its 'form'. For this reason, the previous answer to the question of which logic is correct seems preferable. However, many of the points to follow, e.g., those about a priority and necessity in the next section, would be unaffected by moving to this alternative answer.

It is instructive to once again compare the present proposal with Kripke's strong Kleene theory. On that theory, classical logic is not retained in either of the senses given above. Thus, on Kripke's approach, $\lambda \vee \neg\lambda$ is neither true nor false. But this truth status is determined compositionally: i.e., using the partial interpretation that Kripke assigns T, together with Kleene's strong scheme. Similarly, on what would seem to be the most natural account of what λ says—given Kripke's proposal— λ *does* have its truth status determined by what it says. For λ would seem to express the proposition $\langle\neg\text{TRUE}(\lambda)\rangle$, where TRUE is the property that applies precisely to the objects in the extension of T, and whose negation applies precisely to those objects in the anti-extension of T. But it would then seem that what λ says is neither the case nor not the case—and so the truth status of λ is determined by what it says.[25]

Of course, it remains true that, even on Kripke's proposal, classical arguments preserve truth as long as all of the sentences involved are either true or false. But this does not seem to constitute a very substantive sense in which classical logic remains correct. For if the status of being neither true nor false is one that can be

so cases where an argument's premises are so determined do not lead to failures of truth preservation).

[24] Again the condition can be weakened to require only that this holds of the conclusion of the argument (cf. note 23).

[25] I should note that, just as Kripke insists that his proposal does not amount to a revision of classical logic, so he suggests that, according to it, sentences that are neither true nor false fail to express propositions [1975: 700–701]. It does however seem rather more natural to regard λ, for example, as expressing the proposition mentioned in the text. For a fuller defence of the claim that liar sentences express propositions, see Whittle (2017).

determined by whether what a sentence says is the case—and one which is determined in a straightforward compositional way—then why should logic disregard this status?

§3. Wider implications. In this section I consider the implications that the above account has for widely held views about logic, specifically, that it is a priori and necessary.

3.1. A priority. Logic is widely held to be a priori in the sense: if an argument is logically valid, then we can know a priori that it is truth preserving. On the present proposal, the valid arguments are those of classical logic. Since these are *not* always truth preserving, we certainly cannot always know—a priori or otherwise—that they are.

Nevertheless, logic would still be a priori in an important sense if the exceptions could be determined a priori. They can't be, however. For we get exceptions in the case of liar sentences, for example, and we cannot, in general, determine a priori whether something is a liar sentence. To illustrate, suppose that at noon on November 8, 2016 Donald Trump says: what the next president is saying at noon on November 8, 2016 is untrue. Call the sentence uttered θ (and, for simplicity, take Trump's talk of 'saying' to be referring to the relevant sentence). As it turned out, θ is a loop, but this was not of course something that could be determined a priori. Thus, we could not determine a priori whether, e.g., $\theta \wedge Ws \mathrel{/} \theta$ is an exception.

3.2. Necessity. Logic is also widely held to be necessary: i.e., its arguments do not just preserve truth, they do so necessarily. Again, since not all classical arguments preserve truth simpliciter, they do not do so of necessity. Thus logic is not necessary in this sense. However, it would still be so in an important sense if the classical arguments that actually preserve truth do so necessarily.

This is false, however: simply suppose that Clinton, rather than Trump, was the one uttering θ at noon on November 8, 2016. $\theta \wedge Ws \mathrel{/} \theta$ would then be truth preserving,[26] but merely contingently so.

§4. The bigger picture. I now sketch the bigger picture that the above account of the Chrysippus phenomenon, and of logic, extends to. This picture is cast into a formal theory in §6.

Since Kripke (1975) we have been used to the idea that the property of truth can usefully be thought of as constructed via an iterative procedure. The basic idea behind Kripke's theory is this.

The nonsemantic facts make Ws true and Bs false. This then gives the first level of semantic facts. These then make Tc_{Ws} and $\neg Tc_{Bs}$ true, and Tc_{Bs} and $\neg Tc_{Ws}$ false. Etc. The property of truth applies to those sentences that are eventually made true by this procedure, and 'disapplies' (its negation applies) to those sentences that are made false by it.

In Kripke's picture, sentences receive one of three values: t, f or n (i.e., neither true nor false). For example, λ receives n. However, it is only assignments of t and f that affect later assignments of values. The sentences that receive n are simply those that

[26] Strictly speaking, this requires a further assumption, e.g., that Trump is saying something straightforwardly true or false at noon on November 8, 2016.

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

are left over when all possible assignments of t and f have been made. As one might put it, we 'actively' assign sentences only t and f, and merely 'passively' assign them n.

If we are to accommodate the Chrysippus phenomenon, then something quite different is required. For consider λ and η ($= \neg Tb_\lambda$). η is made true by something about λ. But not of course by λ being true or false (it is not!) Rather, η is made true by λ being neither true nor false. Thus, the assignment of t to η depends on the assignment of n to λ. So we need 'active' assignments not merely of t and f but also of n.

The natural addition to Kripke's picture is simply this: a rule 'actively' assigning n to loops. This rule would allow us to assign n to λ, which would then allow us to assign t to η.[27]

§5. Further issues. In this section I discuss a number of issues that the above discussion makes salient.

5.1. Paradoxes without circularity. The above proposal is in terms of the notion of a loop. But what about paradoxes—or, more generally, sets of 'ungrounded' sentences—that do not involve any such circularity, such as Yablo's paradox? Why have these been neglected?

The reason is that in such cases there is no comparable distinction between the sentences that are in the paradox versus those that are merely about it and, as a result, these do not give rise to instances of the Chrysippus phenomenon, or to the sort of logical exceptions considered in §2.

To illustrate, consider Yablo's paradox, which arises from the following sentences (see [1993]).

α_1: $\forall m > 1$, α_m is untrue.
α_2: $\forall m > 2$, α_m is untrue.
$\vdots$

In the original instance of the Chrysippus phenomenon, there is a clear distinction between the utterances that are in the paradox and those that are merely about it—and C is true because it is on the 'about' side of this distinction.

There is a similar distinction in the case of more complicated circular paradoxes, i.e., more complicated loops. For example, if β is Tc_γ, while γ is $\neg Tc_\beta$, then β and γ—*in* the paradox—are neither true nor false. In contrast, $\delta = \neg Tc_\beta \wedge \neg Tc_\gamma$—merely *about* it—is simply true.

We find a similar distinction in the case of infinitary circular paradoxes. Thus, suppose that for each natural number m, ζ_m is:

$$m = m \wedge \forall l, \zeta_l \text{ is untrue.}$$

Then each ζ_i ('in') is neither true nor false, while a sentence along the following lines ('about') is true:

$$Ws \wedge \forall m, \zeta_m \text{ is untrue.}$$

[27] As I said above (§2.1), although I have indulged in the simplification of talking of single sentences as loops, these must in general be regarded as sets: see definition 3. The rule would then assign n to the members of loops.

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

In the case of Yablo's paradox, however, there is no such natural distinction to be found: if α_0 is

$$\forall m > 0, \alpha_m \text{ is untrue,}$$

then this stands to α_1 α_2, ... just as α_1 stands to α_2, α_3, ... So any line drawn between α_0 and the sentences that it is about would be arbitrary. And a similar point could be made about any other sentence asserting the untruth (or for that matter unfalsity) of the sentences in Yablo's paradox. Consequently, the Chrysippus phenomenon is absent, and thus so are the sort of exceptions that it gives rise to.

We do get *some* exceptions. Specifically, to logical truths: e.g., $\alpha_1 \vee \neg\alpha_1$ is neither true nor false. But these are the only exceptions that noncircular paradoxes (more generally, noncircular ungrounded sets) give rise to. Hence the focus here on circular paradoxes.[28]

5.2. Logic for tokens. As I said in §1.1, I have focused mainly on types so as to make claims that relate directly to standard presentations of logic. However, very similar points can also be made about tokens. Indeed, the analogue of the theorem of §2 can even be strengthened to remove the condition that the conclusion of the propositional argument (i.e., the argument schema) be distinct from each premise: for, as the original Chrysippus phenomenon shows, when it comes to tokens we can have an exception even to the schema p / p; i.e., the instance of this with premise C and conclusion Z.

5.3. Propositions. We have seen that, at the level of sentences (as well of that of tokens), the Chrysippus phenomenon gives rise to exceptions to almost all logical rules. But one might wonder if the position couldn't be made less radical by combining it with the claim that, at the level of propositions, classical logic holds strictly. That is, might we, by taking the traditional line that paradoxical sentences fail to express propositions, be able to vindicate strict classical logic for propositions?

There are reasons to think not. Suppose first that we take propositions to be structured, as on Fregean or Russellian accounts, for example. Then—quite apart from the issue of whether paradoxical sentences express propositions—there will be instances of the Chrysippus phenomenon at the level of propositions. This can be seen as follows.

Although it is arguable that there will *not* be a proposition p of the form $\langle\neg\text{TRUE}(p)\rangle$ (since such a proposition would have to have itself as a constituent), we can generate a

[28] The sentences of Yablo's paradoxes are *only* about later sentences, i.e., α_i is only about α_j for $j > i$. But a similar point could be made about the version of the paradox that replaces '$>$' with '$\geq$'. Again, there is no clear demarcation between the sentences that are in the paradox versus those that are merely about it: a fact that can be illustrated by considering a distinct sentence that essentially says 'me and all of the sentences in the original paradox are untrue'. The crucial point is that once we are dealing with a set of sentences some members of which are not about some others, then we seem no longer to be in a position to draw a clear line between the sentences that are in the paradox versus those that are merely about it: for the natural way of doing this would be by counting as 'merely about' those that some sentences involved in the paradox are *not* about; but in the case of Yablo's paradox (or its variant) this won't work, because sentences that are clearly in the paradox already have this property.

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

proposition that says of itself that it is untrue.[29] Simply consider a function d that sends a proposition q to the result of replacing all occurrences of the proposition $\langle 0 = 0\rangle$ (say) with q itself. The following is then a proposition that says that it itself is untrue:

$$\langle \neg \text{TRUE}(d(\langle \neg \text{TRUE}(d(\langle 0 = 0\rangle))\rangle))\rangle.$$

Call this LP (for 'liar proposition'). LP is a loop in essentially the same way that Z and λ are. But distinct propositions that say that LP is untrue, such as $\langle \text{LP} \wedge \text{LP}\rangle$, are not. Just like C, $\lambda \wedge Ws$ etc. these distinct propositions are simply true. But we then get logical exceptions precisely analogous to as in the sentential case.

There remains the possibility of giving an unstructured account of propositions, e.g., identifying them with sets of possible worlds. But, even ignoring the counterexamples generated by the fact that there is only one necessary proposition, this account does not seem to allow a very robust vindication of classical logic. For classical logic is the theory of the forms generated by the familiar connectives and quantifiers (together with identity); especially, the consequence relation that holds between these. But sets of possible worlds do not realize this structure in any natural sense: that is an upshot of the fact that the possible worlds account identifies logically equivalent propositions.

Either way, then, the hope for a vindication of classical logic that avoids exceptions seems likely to be dashed. Rather, the exceptions are a robust phenomenon.

§6. A formal theory. In this section I present a formal theory that vindicates the claims made above. In the next section I consider alternative approaches. The theory of this section is a version of the contextualist approach of Gaifman (2000), with the crucial difference that truth applies to sentences rather than tokens (cf. §1.1).[30, 31]

Let L be a first-order language with classical interpretation $\mathfrak{A}$, the domain of which is A. I assume that every member of A is denoted by some closed term of L. We extend L to $\mathcal{L}$ by adding two new unary predicate symbols, T and F. Below, unless otherwise stated, by 'sentence', 'term' etc. I mean sentence of $\mathcal{L}$, term of $\mathcal{L}$ etc. I assume that every sentence belongs to A. The formal theory being presented is a recipe for extending $\mathfrak{A}$ to an interpretation of $\mathcal{L}$, under which T is a truth predicate for $\mathcal{L}$, and F is a falsity predicate for $\mathcal{L}$. We add a falsity predicate as well as a truth one because on this approach falsity is distinguished both from untruth and from having a false negation: λ is untrue but not false; and if $\alpha = Tc_\alpha$, then it is neither true nor false, while $\neg\alpha$ is true.

[29] See Whittle (2017).

[30] Gaifman mentions in passing something like this version of his approach. However, he does not discuss the consequences for logic—i.e., whether we get logical exceptionalism—of this or any other version of his approach.

[31] The fact that this theory focuses exclusively on types means that it does not of course treat the tokens Z and C with which we began the paper. Further, since these are tokens of the same type, yet Z is neither true nor false while C is true, we cannot extend the theory to tokens simply by saying essentially that a token is true iff its type is (as we might in the case of a theory that does not aim to incorporate this version of the Chrysippus phenomenon). Rather, to adequately treat tokens we must give a version of the construction that incorporates them from the ground up. It is, however, straightforward enough to do this: this is in effect what the original theory of Gaifman (2000) does.

We simultaneously construct the interpretations of T and F via a sequence of pairs of partial interpretations: $\{\langle T_\mu, F_\mu\rangle : \mu \in \mathrm{On}\}$.[32] Of course, the idea is that T_μ is an interpretation of T, while F_μ is one of F. I use $\mathcal{I}_\mu$ for $\langle T_\mu, F_\mu\rangle$. Further, if S is a partial interpretation, then S^+ is the extension of S, and S^- is the anti-extension.

In defining $\mathcal{I}_\mu$ there are four cases to consider (the third is the most involved).

(1) $\mu = 0$. Then $T_\mu^+ = F_\mu^+ = \varnothing$, and $T_\mu^- = F_\mu^-$ is the set of members of A that are not sentences.

(2) $\mu = \xi + 1$ is odd.[33] By saying that a sentence is satisfied by a pair of partial interpretations $\mathcal{P} = \langle S, R\rangle$ I mean that it is satisfied by the extension of $\mathfrak{A}$ that interprets T by S and F by R, under Kleene's strong scheme.

A sentence α is *determined* by $\mathcal{P}$ if either α or $\neg\alpha$ is satisfied by $\mathcal{P}$; and *undetermined* by $\mathcal{P}$ otherwise. In contrast, α is *evaluated* by $\mathcal{P}$ if $\alpha \in S^+ \cup S^- \cup R^+ \cup R^-$; and *unevaluated* by $\mathcal{P}$ otherwise.

Intuitively, in this case $\mathcal{I}_\mu$ results from assigning t to all sentences that are unevaluated but satisfied by $\mathcal{I}_\xi$; and assigning f to all sentences that are unevaluated, but whose negations are satisfied, by $\mathcal{I}_\xi$. It is essential to restrict attention to unevaluated sentences here: for, looking ahead, once λ has been assigned n, for example, it will become satisfied, but we do not of course want to assign t to it.

That is, let Φ_ξ be the set of sentences that are unevaluated but satisfied by $\mathcal{I}_\xi$; and let Σ_ξ be the set of sentences that are unevaluated, but whose negations are satisfied, by $\mathcal{I}_\xi$. Then:

(OT) $T_\mu^+ = T_\xi^+ \cup \Phi_\xi$ and $T_\mu^- = T_\xi^- \cup \Sigma_\xi$;
(OF) $F_\mu^+ = F_\xi^+ \cup \Sigma_\xi$ and $F_\mu^- = F_\xi^- \cup \Phi_\xi$.

Thus, unevaluated sentences that are satisfied by $\mathcal{I}_\xi$ are placed into the extension of T and the anti-extension of F. (Those whose negations are satisfied are rather placed into the extension of F and the anti-extension of T.) In general, i.e., at each stage of the construction, anything in the extension of T is in the anti-extension of F, and similarly everything in the extension of F is in the anti-extension of T. The reverse claims however do not hold: sentences like λ that are neither true nor false will be placed into each anti-extension, but neither extension (a fate shared of course by any member of the domain that is not a sentence).

(3) $\mu = \xi + 1$ is even. It is in this case that we 'actively' assign n to sentences (cf. §4). That is, we make assignments of n that affect later assignments of values: e.g., that to λ, which is the basis of the later assignment of t to η (see example 2(ii)). Formally speaking, to assign n to a sentence is to place it in the anti-extension of both T and F.

The following sequence of definitions leads up to that of a loop. The first codifies the notion of a sentence α being directly about the truth status of an unevaluated sentence β. We restrict attention to the case where β is unevaluated because we want to treat, e.g., $\gamma = \neg Tc_\gamma \wedge Tc_{Ws}$ as essentially on a par with λ. That is, once Ws has been assigned t, we ignore the fact that γ is about this truth status. The basic idea is that α is about the truth status of β if β is one of the sentences that must receive a truth status for α to become determined. This means that if α is $\neg Tc_\alpha \wedge Ws$ or $\neg Tc_\alpha \vee Bs$

[32] On is the class of ordinals.
[33] That is, $\mu = \rho + m$ for some limit ordinal ρ and odd $m \in \omega$.

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

(and unevaluated), then it will, in the relevant sense, be about its own truth status. In contrast, if α is $\neg Tc_\alpha \vee Ws$, then it won't be. This seems desirable: for while we want to treat the first two sentences as akin to λ, we do not so want to treat the third, because intuitively it is not a loop in the same way; ask whether this sentence is true and there is no need to return to that question, the matter is settled simply by the fact that snow is white.[34]

DEFINITION 1. *Let $\mathcal{P}$ be a pair of partial interpretations, and let α and β be sentences.* α calls directly β *under $\mathcal{P}$ if there are sentences $\gamma_1, \ldots, \gamma_m$ such that*:

(i) *each γ_i is undetermined by $\mathcal{P}$*;
(ii) *γ_1 is α*;
(iii) *for each $i < m$, γ_{i+1} is an immediate sentential component*[35] *or an instance of γ_i*;
(iv) *γ_n is $T\tau$ or $F\tau$ for some closed term τ with $\tau^{\mathfrak{A}} = \beta$.*

I omit mention of the pair of partial interpretations when this is clear from the context. For example, if δ is $\exists x[(x = c_\zeta \vee x = c_\theta) \wedge Tx]$, and ζ and θ are unevaluated by $\mathcal{P}$, then δ calls directly ζ under $\mathcal{P}$, via the sequence: $\delta\ (=\gamma_1)$; $(x = c_\zeta \vee x = c_\theta) \wedge Tc_\zeta$; Tc_ζ.[36] Similarly, δ calls directly θ, but it does not call directly any other sentences. And, as expected, if λ is unevaluated by $\mathcal{P}$, then it calls directly itself, via the sequence: λ; Tc_λ. λ does not however call directly anything else. The following codifies the general notion of a sentence being about the truth status of some unevaluated sentence.

DEFINITION 2. *Let $\mathcal{P}$ be a pair of partial interpretations, and let α and β be sentences.* α calls β *under $\mathcal{P}$ if there are sentences $\gamma_1, \ldots, \gamma_m$ ($m \geq 2$) such that*: *γ_1 is α*; *for $i < m$, γ_i calls directly γ_{i+1} under $\mathcal{P}$*; *γ_m is β.*

DEFINITION 3. *Let $\mathcal{P}$ be a pair of partial interpretations, and let Γ be a set of sentences.* Γ *is a* loop *under $\mathcal{P}$ if every member of Γ calls exactly the members of Γ under $\mathcal{P}$.*

EXAMPLE 1. *In the following, I assume that all of the sentences mentioned are unevaluated by $\mathcal{P}$. These sets are then loops under $\mathcal{P}$.*

(i) $\{\lambda\}$.
(ii) $\{\alpha\}$, *where α is Tc_α.*
(iii) $\{\beta, \gamma\}$, *where β is Tc_γ and γ is Fc_β.*
(iv) $\{\delta\}$, *where δ is $Tc_\delta \rightarrow 0 = 1$.*
(v) $\Theta = \{\chi_m : m \in \omega\}$, *where χ_i is $\forall x(x \in c_\Theta \rightarrow \neg Tx) \wedge i = i$.*

Intuitively, $\mathcal{I}_\mu$ results from assigning n to every sentence that belongs to a loop under $\mathcal{I}_\xi$. Thus, let Ψ_ξ be $\bigcup\{\Gamma : \Gamma \text{ is a loop under } \mathcal{I}_\xi\}$. Then:

[34] Note that in the following we do not need to stipulate that β is unevaluated: this will follow from the fact that a sentence of the form $T\tau$ or $F\tau$, for τ a closed term denoting β is undetermined, together with the definition of our sequence of partial interpretations.

[35] The immediate sentential component of $\neg\delta$ is δ; and if $*$ is a binary connective, then the immediate sentential components of $\delta * \zeta$ are δ and ζ.

[36] Strictly speaking, this assumes that $\exists$ and $\wedge$ are basic symbols of $\mathcal{L}$. If they are instead defined, then δ will still call directly ζ, but via a different sequence. For simplicity, I assume below that all the logical symbols mentioned are basic symbols of $\mathcal{L}$, but nothing essential turns on this. For it is easy to see that which sets are loops, and thus which sentences are assigned n, is unaffected by which logical symbols are basic and which defined.

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

(ET) $T^+_\mu = T^+_\xi$ and $T^-_\mu = T^-_\xi \cup \Psi_\xi$;
(EF) $F^+_\mu = F^+_\xi$ and $F^-_\mu = F^-_\xi \cup \Psi_\xi$.

(4) μ is a limit ordinal. Then $T^+_\mu = \bigcup_{\xi<\mu} T^+_\xi$ and $T^-_\mu = \bigcup_{\xi<\mu} T^-_\xi$. Similarly for F_μ.

Since we only ever assign values to sentences that were previously unevaluated, we have that if $\mu \leq \xi$, then $T_\mu \leq T_\xi$ and $F_\mu \leq F_\xi$.[37] The fact that there are more ordinals than sentences thus gives:

THEOREM 1. *There is $\rho \in On$ such that for all $\mu \geq \rho$, $T_\mu = T_\rho$ and $F_\mu = F_\rho$.*

According to the proposed theory, the sentences that are true are those that are assigned t in the above construction, i.e., those that belong to T^+_ρ. Similarly, the sentences that are false are those that belong to F^+_ρ. These sets, T^+_ρ and F^+_ρ, are also the proposed interpretations of T and F, respectively. I use $\mathbf{T}$ for T^+_ρ, $\mathbf{F}$ for F^+_ρ, and $\mathbf{I}$ for $\langle \mathbf{T}, \mathbf{F} \rangle$. Thus, the proposed interpretations are classical, or total, rather than partial: as required, if we are to vindicate the claim of §2.3 that classical arguments preserve truth when the sentences involved have their truth statuses determined compositionally.

By the monotonicity of Kleene's strong scheme, if $\alpha \in \mathbf{T}$, then $\mathbf{I} \vDash \alpha$; and if $\alpha \in \mathbf{F}$, then $\mathbf{I} \vDash \neg\alpha$. But of course the converses of these claims do not hold. For example, $\mathbf{I} \vDash \lambda$, but $\lambda \notin \mathbf{T}$. And if β is Tc_β, then $\mathbf{I} \vDash \neg\beta$, but $\beta \notin \mathbf{F}$.

It might be complained that there is something arbitrary about the above construction: for example, why handle the first batch of loops at the second, rather than the first or third stage, say? Indeed, why even separate the assignment of standard values (i.e., t and f) from that of n, as I have, rather than dealing with these together at each stage? It can be proved, however, that the end result of the construction, i.e., $\mathbf{I}$, does not depend on such choices. (For reasons of space I do not give this argument here.) It follows that there is nothing arbitrary about our proposed interpretations of T and F.

EXAMPLE 2.
(i) *All of the sentences mentioned in example 1 belong to $T^-_2 \cap F^-_2$: it is easy to see that none of these or their negations are satisfied by $\mathcal{I}_0$, and thus that they belong to Ψ_1.*
(ii) *$\eta \in T^+_3$. Recall that η is $\neg Tb_\lambda$, where b_λ is distinct from c_λ. Given $\lambda \in T^-_2$, $\mathcal{I}_2 \vDash \eta$. $\eta \in T^+_3$ is then clear as long as $\eta \notin \Psi_1$. But this follows from the fact that if Γ is a loop under $\mathcal{I}_1$, then $\eta \in \Gamma$ would require $\lambda \in \Gamma$ and thus for λ to call η, which is clearly impossible.*
(iii) *$\lambda \wedge Ws \in T^+_3$, $Tc_\lambda \in F^+_3$, $Tc_\lambda \leftrightarrow \lambda \in F^+_3$, and $\forall x(x \in c_\Theta \rightarrow \neg Tx) \in T^+_3$ (where Θ is as in example 1(v)) by similar arguments.*
(iv) *If ζ is $\neg Tc_\zeta \wedge Ws$, then $\zeta \in T^-_2 \cap F^-_2$ but $\neg Tc_\zeta \in T^+_3$. First, $\{\zeta\}$ is a loop under $\mathcal{I}_1$. For ζ calls directly itself via the sequence: ζ; $\neg Tc_\zeta$; Tc_ζ. It is easy to see, however, that ζ does not call directly anything else. It follows that $\zeta \in \Psi_1$ and thus $T^-_2 \cap F^-_2$. Since it is clear that $\neg Tc_\zeta$ does not belong to a loop under $\mathcal{I}_1$, but $\mathcal{I}_2 \vDash \neg Tc_\zeta$, we have $\neg Tc_\zeta \in T^+_3$.*
(v) *Let $\varphi(x,y)$ be a formula of L satisfied precisely by $\langle m, \alpha_m \rangle$, for $m \in \omega$, where α_i is:*

$$\forall x \forall y[(x > i \wedge \varphi(x,y)) \rightarrow \neg Ty].$$

[37] If R and S are partial interpretations, then $S \leq R$ means $S^+ \subseteq R^+$ and $S^- \subseteq R^-$.

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

Then, for each m, $\alpha_m \notin \mathbf{T} \cup \mathbf{F}$. We must show that no α_i or $\neg\alpha_i$ is satisfied by $\mathcal{I}_\mu$; and also that no α_i belongs to a loop under $\mathcal{I}_\mu$ (for any $\mu \in On$). We prove both claims by a simultaneous induction on μ. Thus, suppose that μ is least such that one of the claims fails. If $\mathcal{I}_\mu$ satisfies either α_i or $\neg\alpha_i$, for some i, then it is easy to see that some α_j must be evaluated by $\mathcal{I}_\mu$, which is impossible by the inductive hypothesis. So some α_i must belong to a loop under $\mathcal{I}_\mu$. But this would require α_i to call itself, which it is easy to see is impossible (since α_i will only call α_j for $j > i$). It follows that, for each m, $\alpha_m \notin \mathbf{T} \cup \mathbf{F}$.

(vi) *If $\alpha \in \mathbf{T}$, then $T\tau_\alpha \in \mathbf{T}$ and $F\tau_\alpha \in \mathbf{F}$; and if $\alpha \in \mathbf{F}$, then $T\tau_\alpha \in \mathbf{F}$ and $F\tau_\alpha \in \mathbf{T}$. Suppose $\alpha \in T^+_\mu$ for some even μ. Then $\mathcal{I}_\mu \models T\tau_\alpha$, and so $T\tau_\alpha \in T_{\mu+1}$ as long as $T\tau_\alpha$ does not belong to a loop for some $\xi < \mu$. It is easy to see, however, that any loop that contained $T\tau_\alpha$ would also contain α, which is impossible given $\alpha \in T^+_\mu$. Thus, $T\tau_\alpha \in \mathbf{T}$. Similarly for the other cases.*

Finally, we prove a version of the theorem of §2. I assume that for each formula φ of $\mathcal{L}$, with at most x free, there is a closed term π of L such that $\pi^{\mathfrak{A}} = \varphi(x/\pi)$.[38]

Let $\alpha_1, ..., \alpha_m$ be propositional formulas and let $\alpha'_1, ..., \alpha'_m$ be sentences of $\mathcal{L}$. We say that $\alpha'_1, ..., \alpha'_m$ result by *uniform substitution* from $\alpha_1, ..., \alpha_m$, respectively, if $\alpha'_1, ..., \alpha'_m$ can be obtained from $\alpha_1, ..., \alpha_m$, respectively, by uniformly substituting sentences of $\mathcal{L}$ for propositional symbols.

THEOREM 2. *Let $\alpha_1, ..., \alpha_m, \beta$ be propositional formulas such that*

$\beta \neq \alpha_i$ *for* $i = 1, ..., m$,
$\{\alpha_1, ..., \alpha_m\}$ *is classically consistent.*

Then there are sentences $\alpha'_1, ..., \alpha'_m, \beta'$ of $\mathcal{L}$ such that:

(i) *$\alpha'_1, ..., \alpha'_m, \beta'$ result by uniform substitution from $\alpha_1, ..., \alpha_m, \beta$, respectively*;
(ii) *for $i = 1, ..., m$, $\alpha'_i \in \mathbf{T}$*;
(iii) *$\beta' \notin \mathbf{T}$.*

Proof. We use essentially the argument of the sketch of §2. The claim is clear if β is not a classical consequence of $\alpha_1, ..., \alpha_m$, so suppose that it is. Then let $p_1, ..., p_l$ be the propositional symbols that occur in $\alpha_1, ..., \alpha_m, \beta$, and let r be a row of the truth table for $\alpha_1, ..., \alpha_m, \beta$ in which each of these is true. For $\gamma \in \{\alpha_1, ..., \alpha_m, \beta\}$, γ^* is the result of making the following substitutions in γ: if p_i is true in r, then replace p_i with $\neg Tx \wedge i = i$; otherwise, replace it with $Tx \wedge i = i$. Then let γ' be $\gamma^*(x/\pi)$, where π is a closed term of L that denotes $\beta^*(x/\pi)$, i.e., β'.

It is easy to see that $\{\beta'\}$ is a loop under $\mathcal{I}_1$, and thus $\beta' \notin \mathbf{T}$. Further, since $\{\beta'\} \in T^-_2$, each α'_i is satisfied by $\mathcal{I}_2$, and so $\alpha'_i \in T^+_3 \subseteq \mathbf{T}$. □

§7. Alternative approaches.

7.1. Supervaluationist. The most obvious alternative to the theory of §6 would replace the evaluation scheme used there, i.e., Kleene's strong scheme, with an alternative: specifically, the supervaluationist scheme. The result would be a slightly—

[38] This assumption would be satisfied if formulas of $\mathcal{L}$ are in A (the domain of $\mathfrak{A}$), and L contains a term for a diagonal function. Alternatively, a version of it would be satisfied if we used the language of arithmetic and Gödel numbers in place of expressions of $\mathcal{L}$.

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

but only slightly—less thoroughgoing form of logical exceptionalism. In particular, theorem 2 would no longer hold. However, if we add the condition that β is not a logical truth, then the result does hold (and by a similar argument).

7.2. Upwards strict. In §2.1 I argued that in defining loops, we should consider only the sentences that a given sentence is *about*, rather than those that are merely components of the sentence. And it is this approach that we pursued in §6. Nevertheless, it is instructive to consider what happens if we reject this conclusion.

On this alternative approach, then, definition 1 would be replaced with the following.[39]

DEFINITION 4. *Let $\mathcal{P}$ be a pair of partial interpretations, and let α and β be sentences. α* calls* directly *β under $\mathcal{P}$ if α is undetermined by $\mathcal{P}$ and one of the following conditions obtains.*

(a) *β is undetermined by $\mathcal{P}$ and an immediate sentential component or an instance of α.*
(b) *α is $T\tau$ or $F\tau$ for some closed term τ with $\tau^{\mathfrak{A}} = \beta$.*

Everything else is as in §6.

This change has the effect that one can never have, e.g., a conjunction that is neither true nor false, but whose conjuncts are true. More generally, we have the following. If α is a propositional formula, then $C(\alpha)$ is the set of its components.[40] I use $\vDash_{\mathrm{SK}}$ for the strong Kleene consequence relation.

THEOREM 3. *Let $\alpha_1, \ldots, \alpha_m, \beta$ be propositional formulas such that $C(\beta) \cap \{\alpha_1, \ldots, \alpha_m\} \vDash_{\mathrm{SK}} \beta$. Then $\alpha_1, \ldots, \alpha_m \vDash \beta$ holds on this approach.*

Conversely, however:

THEOREM 4. *Let $\alpha_1, \ldots, \alpha_m, \beta$ be propositional formulas such that*

$$C(\beta) \cap \{\alpha_1, \ldots, \alpha_m\} = \varnothing,$$
$$\{\alpha_1, \ldots, \alpha_m\} \text{ is classically consistent.}$$

Then $\alpha_1, \ldots, \alpha_m \vDash \beta$ has a false instance on this approach.

Both results can be strengthened a bit, but for simplicity I omit the details. As before, similar claims hold for the first-order case.

One disadvantage of this approach is that the relation of calling* directly, and ultimately the final result of the evaluation procedure, is sensitive to which connectives we take as basic. To illustrate, consider a Curry sentence $\alpha = Tc_\alpha \to 0 = 1$. Regardless of whether we take $\to$ as basic, α will belong to a loop under $\mathcal{I}_1$ (i.e., on this alternative approach). However, which other sentences belong to this loop depends on whether we take $\to$ as basic—as, ultimately, does the issue of which sentences are assigned *n*.

[39] The approach of Hansen (2014) in effect uses this definition. However, since the rest of that approach is importantly different from that of §6, the overall approach is also different from that of this subsection. I discuss Hansen's approach in §7.4.2.

[40] That is, $C(\alpha)$ is the closure of $\{\alpha\}$ under the relation of being an immediate sentential component.

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

Specifically, if $\rightarrow$ is basic, then the loop that contains α will be $\{\alpha, Tc_\alpha\}$. In this case, $\neg Tc_\alpha$ will belong to T_3^+ (i.e., be assigned t); while $Tc_\alpha \wedge 0 \neq 1$ will belong to F_3^+ (i.e., be assigned f). In contrast, if $\rightarrow$ is instead defined in terms of $\neg$ and $\vee$ (so that α is $\neg Tc_\alpha \vee 0 = 1$), then the loop containing α will rather be $\{\alpha, \neg Tc_\alpha, Tc_\alpha\}$. Thus, $\neg Tc_\alpha$ will instead be assigned n. ($Tc_\alpha \wedge 0 \neq 1$ will still be assigned f.) But if $\rightarrow$ is instead defined in terms of $\neg$ and $\wedge$ (so α is $\neg(Tc_\alpha \wedge 0 \neq 1)$), then the loop will rather be $\{\alpha, Tc_\alpha \wedge 0 \neq 1, Tc_\alpha\}$. This means that $Tc_\alpha \wedge 0 \neq 1$ is assigned n (while $\neg Tc_\alpha$ is assigned t). And so on.

As I said (in note 36), things are quite different on the approach of §6: under that, which sets are loops, and thus which sentences are assigned which values, does not depend on which symbols are taken as basic. This would seem to be an advantage of that approach.

7.3. Downwards strict. To get a sort of opposite of this, we need to change the way in which standard values are assigned.[41] According to (OT) and (OF), unevaluated sentences are assigned t (or f) if they (their negations) are satisfied. To get an approach that is downwards strict, one would instead assign standard values on the basis of sentential components and instances. Thus, $\neg\alpha$ is assigned t if α has been assigned f; and f if α has been assigned t. $\alpha \wedge \beta$ is assigned t if α and β have been; and f if one of α or β has been. $\forall x\varphi$ is assigned t if, for every $a \in A$, there is a closed term τ denoting a such that $\varphi(x/\tau)$ has been assigned t; $\forall x\varphi$ is assigned f if some instance of it has been. The rest of the approach is as before.

Let $\Gamma \cup \{\delta\}$ be a set of propositional formulas. Γ *downward entails* δ if $\Gamma \vDash_{SK} \delta$ and for some $\gamma \in \Gamma$, $\delta \in C(\gamma)$. We then have the following.

THEOREM 5. *Let α_1, ..., α_m, β be propositional formulas. Then $\alpha_1, \ldots, \alpha_m \vDash \beta$ holds strictly on this approach iff $\{\alpha_1, \ldots, \alpha_m\}$ downward entails β.*

Again, similar claims hold for first-order logic. Further, we can combine the approaches of this and the last subsection to get one that is both upwards and downwards strict (in the senses exemplified by those approaches).

7.4. Two extant proposals. Finally, I consider two approaches from the literature that bear similarities to the present one.

7.4.1. Skyrms. As I mentioned in note 13, Skyrms (1984) advocates a treatment of the Chrysippus phenomenon that involves exceptions to rules of classical logic: specifically, instances of intensionality, i.e., cases where the substitution of coreferential terms results in a change in truth status. It is instructive to compare this approach to that proposed here.

The main difference is that, on Skyrms' approach, the only exceptions to logical rules are such instances of intensionality. And this is an essential feature of the approach, since its central innovation is to think of the semantic value of the truth predicate not as the set of objects that it applies to (or as the pair of this set together with the set of objects that it disapplies to), but rather as a set of pairs of objects together with terms denoting them (or a pair of such sets); where the idea is that $T\tau$ is true iff $\langle \tau^{\mathfrak{A}}, \tau \rangle$ is in

[41] The approach of this subsection is a version of that of Gaifman (1992), but with again the difference that truth applies to sentences rather than tokens.

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

the set.[42, 43] Compound sentences then have their truth status determined from atomic ones in the obvious way.

This allows for the Chrysippus phenomenon as follows. Consider λ and η. One can have η but not λ being true by placing $\langle \lambda, b_\lambda \rangle$, but not $\langle \lambda, c_\lambda \rangle$, into the anti-extension of T.

The main problem with this proposal, however, is that we have seen (§2) that there are instances of the Chrysippus phenomenon—i.e., instances of the same phenomenon that λ and η give rise to—that do not involve distinct coreferential names (or distinct tokens of names). For example, the instances discussed in §2 in connection with conjunction introduction and elimination. What is really responsible for the phenomenon, therefore, is not anything to do with distinct coreferential names in particular. Rather, what is responsible is the very natural distinction between sentences that are in a paradox versus those that are merely about it. And this is a distinction that occurs both in cases that involve multiple coreferential names, and in cases that do not. It seems, therefore, that Skyrms' proposal does not get to the heart of the matter.[44]

7.4.2. Hansen. Finally, I consider the approach of Hansen (2014). This involves a wider range of exceptions to classical rules than that of Skyrms (and, unlike on that approach, these are not tied to the presence of distinct coreferential names). In terms of motivation and general shape, there are a number of ways in which Hansen's approach is similar to the proposal of this paper. However, there are also significant differences—which, it seems to me, favour the approach proposed here. The most important of these are as follows.

---

[42] Although Skyrms thinks of the semantic value of T as constituted by pairs of this form, there is of course a sense in which this is optional: the same information would be carried more simply by the set of terms τ such that $T\tau$ is true.

[43] It is not part of the proposal as stated, but the idea would be to extend this to instances of the Chrysippus phenomenon involving different tokens of the same type (e.g., our original case of Z and C) by further augmenting the semantic value of T. For example, by allowing it to contain triples of objects, terms and utterances. (In fact, this sort of extension of Skyrms' proposal would fix a flaw in that which he envisages which results from his apparent failure to realize that one can have instances of the phenomenon involving different tokens of the same type, even without the involvement of indexicals: e.g., the case of Z and C.)

[44] There is another significant drawback with the specific formal theory of Skyrms (1984), which is that it is completely unsystematic. Thus, this theory allows for interpretations of $\mathcal{L}$ in which η is true but λ is not. However, it also allows for interpretations in which η and λ are both untrue—the interpretations under which η is true are generated simply by putting $\langle \lambda, b_\lambda \rangle$ into the extension of T 'by hand'. Similarly, let b'_λ be a new name (distinct from c_λ and b_λ), and let η' be $\neg Tb'_\lambda$. Then Skyrms' theory also allows for interpretations in which, e.g., η is true while η' and λ are not. What is desirable (both simpliciter and by Skyrms' lights) is an interpretation on which η and η' are both true while λ is not. However, the theory provides no deterministic procedure which results in such an interpretation, as opposed to one of the other interpretations just mentioned. This is of course in stark contrast to the theory of §6 which does include such a procedure (i.e., which yields an interpretation on which η and η' are true while λ is neither true nor false). The failure of Skyrms' approach to provide such a procedure seems problematic for a number of reasons. To give just the most obvious: it seems that he has failed to in fact propose a definite extension of our original interpreted language.

Definition of loops. Hansen in effect uses the notion of calling* directly, and thus the definition of loops, of §7.2. As a consequence, his approach is upwards strict in the sense of theorem 3. However, this definition of loops has the disadvantages that we have noted. Firstly, it seems to ride roughshod over the distinction between the things that a sentence is about, versus those that are its components (see §2). And, secondly, it means that the approach is sensitive to which logical symbols we take as basic (see §7.2).

Special treatment of quote names. Another difference (which is less obviously a disadvantage) is that Hansen singles out quote names for special treatment. Essentially, quote names are privileged in the following way: when considering a loop, for example, sentences that contain quote names are assigned values after those that don't, with the result that sentences with quote names stand a better chance of receiving a standard value.

A simple example: let α be T'β', where β is Tc_α (and where c_α is not a quote name). On the approach of §6, α and β are both assigned n. On Hansen's approach, however, while β will be neither true nor false, α will be false.

The main selling point of this feature of Hansen's approach is that it means that we can always truthfully describe a sentence's truth status. This is in contrast to the present approach on which, for example, any sentence that says that one of the sentences in Yablo's paradox is neither true nor false is itself neither true nor false.

The disadvantage of this feature, it seems to me, is that truth statuses are no longer determined purely by the structure of the referential network that the sentences of our language form. Rather, they depend also on what might seem to be rather arbitrary distinctions between the means by which the links of this network are forged.

Nevertheless, if desired one could straightforwardly enough incorporate this aspect of Hansen's approach into the present one (although for reasons of space I omit the details).[45] That is, this difference is independent of the others discussed in this subsection, and which do seem to me to clearly favour the approach proposed in this paper.

Classical inconsistency. The main difference between Hansen's approach and the present one, however, is that it is classically inconsistent in the sense: there are classically inconsistent sets of sentences, each member of which is true on Hansen's approach. This is due to the following two features of Hansen's approach:[46] (i) sentences that receive standard values have these determined compositionally from the values of their sentential components; and (ii) in such cases, n is treated identically to f.

In some cases, this seems to give the right result. Thus, to illustrate, consider $\neg\alpha$, where α is Fc_α. On Hansen's approach (as on the present one), α is neither true nor

[45] Indeed, for a version of the approach of Gaifman (2000) that incorporates such an idea—and which, when transposed from tokens to types would give a version of the approach of §6 that incorporates such an idea—see 2000, pp. 113–18).

[46] For simplicity, I restrict attention here to the relationship between the truth status of a sentence and the statuses of its sentential components, but similar remarks apply to the relationship between the truth status of a sentence and those of its instances, those of instances of its sentential components, etc.

https://doi.org/10.1017/S1755020320000210 Published online by Cambridge University Press

false, while $\neg\alpha$ is simply true. But since Hansen insists on deriving the *t* that $\neg\alpha$ receives from the *n* that its sentential component α receives, he is forced to say that whenever a sentence β is neither true nor false, and $\neg\beta$ receives a standard value, then $\neg\beta$ is true. But while this gives the right result in the case of $\neg\alpha$, it gives disastrous ones in other cases. For example, $\neg\lambda$, i.e., $\neg\neg Tc_\lambda$, comes out as true, despite the fact that λ is not true. But how can something be not not true, when it *is* in fact not true? This also gives a classical inconsistency: since η, i.e., $\neg Tb_\lambda$, and $b_\lambda = c_\lambda$ are both true. It is hard not to feel that something has gone badly wrong.

The mistake, I would suggest, is thinking that $\neg\alpha$'s *t* must be derived from its sentential component α's *n*. Rather, $\neg\alpha$ is true because, once α has been assigned *n*, what $\neg\alpha$ says is the case (i.e., $\neg\alpha$ becomes satisfied). Seeing things this way—as on the present approach—allows $\neg\alpha$ to be true in a way that does not bring the undesirable truth of $\neg\lambda$ in its wake. As far as I can tell, then, the main differences between Hansen's approach and the present one favour the latter.[47]

We have seen, then, that logic is exceptional, if not in quite the way that was expected.[48]

BIBLIOGRAPHY

Barwise, J., & Etchemendy, J. (1987). *The Liar*: *An Essay on Truth and Circularity*. Oxford: Oxford University Press.

Beall, J. (2009). *Spandrels of Truth*. Oxford: Clarendon Press.

Burge, T. (1979). Semantical paradox. *Journal of Philosophy*, **76**, 169–198.

Field, H. (2008). *Saving Truth from Paradox*. Oxford: Oxford University Press.

Gaifman, H. (1992). Pointers to truth. *Journal of Philosophy*, **89**, 223–261.

———. (2000). Pointers to propositions. In Chapuis, A., & Gupta, A., editors. *Circularity, Definition, and Truth*. New Delhi, India: Indian Council of Philosophical Research, pp. 79–121.

Glanzberg, M. (2001). The liar in context. *Philosophical Studies*, **103**, 217–251.

Gupta, A. (2001). Truth. In Goble, L., editor. *The Blackwell Guide to Philosophical Logic*. Oxford: Blackwell, pp. 90–114.

Hansen, C. S. (2014). Grounded ungroundedness. *Inquiry*, **57**, 216–243.

[47] There is one other difference that I should mention: Hansen's approach seems to give the wrong result in the case of infinite loops. For example, recall Θ of example 1(v) of §6. The theory of that section gives what would seem to be the right result concerning $\forall x(x \in c_\Theta \rightarrow \neg Tx)$ (which does not belong to Θ): it is simply true (like C, η etc.) After all, just as in the cases of C, η and so on, this sentence seems to be merely about—rather than in—the paradox. Hansen's approach, however, essentially counts all infinite paradoxes as akin to Yablo's, and so this sentence is assigned *n* along with the members of Θ. This seems undesirable, since the fundamental distinction is not between finite and infinite paradoxes (or sets of ungrounded sentences), but rather between circular and noncircular ones. Thus, the present approach, by treating the paradox involving Θ essentially as it treats the standard liar paradox that involves λ, and not in the way that it treats Yablo's paradox, seems to get things right in this regard.

[48] For comments and discussion, I would like to thank Matti Eklund, Stephan Krämer, Stephan Leuenberger, Joshua Schechter, Martin Smith, Joel Velasco, Gareth Young, audiences at Glasgow, Hong Kong, Texas Tech and Uppsala, and a referee for this journal. This work was supported by the Arts and Humanities Research Council [grant number AH/M009610/1].

Hofweber, T. (2008). Validity, paradox and the ideal of deductive logic. In Beall, J., editor. *Revenge of the Liar*: *New Essays on the Paradox*. Oxford: Oxford University Press, pp. 145–158.

———. (2010). Inferential role and the ideal of deductive logic. *The Baltic International Yearbook of Cognition, Logic and Communication*, **5**, pp. 1–26.

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, **72**, 690–716.

Maudlin, T. (2004). *Truth and Paradox*: *Solving the Riddles*. Oxford: Clarendon Press.

Parsons, C. (1974). The liar paradox. *Journal of Philosophical Logic*, **3**, 381–412.

Priest, G. (2006). *In Contradiction*. Oxford: Clarendon Press.

Simmons, K. (1993). *Universality and the Liar*: *An Essay on Truth and the Diagonal Argument*. Cambridge: Cambridge University Press.

Skyrms, B. (1984). Intensional aspects of self-reference. In Martin, R. L., editor. *Recent Essays on Truth and the Liar Paradox*. Oxford: Clarendon Press, pp. 119–131.

Soames, S. (1999). *Understanding Truth*. Oxford: Oxford University Press.

Szabó, Z. G. (2017). Compositionality. In Zalta, E. N., editor. *The Stanford Encyclopedia of Philosophy*. Summer 2017 edition. plato.stanford.edu/archives/sum2017/entries/compositionality.

Whittle, B. (2017). Self-referential propositions. *Synthese*, **194**, 5023–5037.

Yablo, S. (1993). Paradox without self-reference. *Analysis*, **53**, 251–252.

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF WISCONSIN–MADISON
MADISON, WI 53706, USA
E-mail: bwhittle@wisc.edu