Dissolving the paradoxicality paradox

William Nava

Abstract

Non-classical solutions to semantic paradox can be associated with conceptions of paradoxicality understood in terms of entailment facts. In a K3-based theory of truth, for example, it is *prima facie* natural to say that a sentence ϕ is paradoxical iff $\phi \vee \neg\phi$ entails an absurdity. In a recent paper, Julien Murzi and Lorenzo Rossi exploit this idea to introduce revenge paradoxes for a number of non-classical approaches, including K3. In this paper, I show that on no understanding of 'is paradoxical' (for K3) should both rules needed for their paradox be expected to hold unrestrictedly. Just which rule fails, however, depends on various factors, including whether the derivability relation of a target system of reasoning is arithmetically definable.

Keywords: semantic paradoxes; revenge paradoxes; non-classical logic; validity paradoxes

1 Introduction

The Liar paradox teaches us that classical logic, self-reference, and the so-called transparency (or 'naivety') of truth are incompatible: together they entail everything. Accordingly, proposed solutions to it and related paradoxes fall into three main camps: those that block self-reference;[1] those that restrict the transparency of truth; and those that weaken classical logic. The latter group can be categorized based on which classical rule(s) they restrict so as to avoid paradox. My focus in this paper will be on K3-based approaches, which restrict the law of excluded middle (see Kripke [1975], Field [2008]).

Murzi and Rossi [2020] (henceforth: M&R) have recently introduced four new paradoxes, one for each of four different non-classical approaches, including K3. Like the Liar, these new paradoxes involve a predicate—'is paradoxical' in the case of K3—governed by apparently intuitive (or 'naive') rules. M&R show that K3 is trivial when augmented with both the naive truth rules and the rules governing the paradoxicality predicate. They claim that further weakening K3 renders it too weak to be tenable,[2] and argue that the motivation for the naive paradoxicality rules is of a kind with the motivation for the naive truth rules, such that it would be ad hoc to insist on one but not the other. If M&R are right, then the K3 approach will not do for a unified treatment of the paradoxes and K3 theorists must simply accept that naivety (for both truth and paradoxicality) must go. My primary aim in this paper is to defuse this argument.[3]

[1] Self-reference scenarios familiar from Kripke [1975], as well the fact that a modicum of arithmetic or syntax generates self-referentially behaving sentences, make this something of a nuclear option. I will not continue to consider it.

[2] This, it should be said, is not obvious. It depends on whether there is a weakening of K3 weak enough to avoid M&R's triviality proof but still strong enough to recapture classical logic in classical contexts. This is an interesting question, though one I won't try to answer.

[3] It should be relatively straightforward to adapt most of the arguments I'll make in this paper to M&R's paradox for the LP approach. On the other hand, their paradoxes for substructural approaches raise quite different issues and

Let me begin by clarifying the dialectical situation. The concept of paradoxicality is not univocal. Rather, that generic concept may be specified in a number of different ways to express distinct (but perhaps related) properties of sentences. M&R offer no reason to think K3 theorists are unable to express *any* such property; nor that a predicate expressing any such property must conform to their proposed rules. Rather, they introduce two cases of intuitive reasoning meant to illustrate an expressive need for a *specific* notion of paradoxicality that does conform to their rules. For this reason, offering up a purportedly better account of paradoxicality that does not conform to their rules would be a non-sequitur. To refute their argument, one needs to show that their case for the coherence of *their specific notion* fails; and that K3 theorists can maintain, without fear of expressive inadequacy, that any notion conforming to M&R's proposed rules is incoherent.[4]

On the other hand, it must be stressed that by the lights of K3 theorists, M&R's own triviality proof is already a refutation of their notion of paradoxicality. Unless M&R can independently motivate the coherence of their notion, there is no reason to think the problem is with K3 and not with their notion. For this reason, undermining M&R's case for the coherence of the latter is *all* the K3 theorist must do. She need not additionally demonstrate its incoherence. For, without independent motivation for its coherence, that has already been accomplished by the triviality proof.

Roadmap: §2 sets up the language and proof system I'll work with, introduces M&R's paradoxicality rules, and outlines the proof of K3's triviality when augmented with the paradoxicality and truth rules. In §3, I discuss M&R's motivations for their rules, including their two cases, the significance of the sequent $\phi \vee \neg\phi \Rightarrow \bot$, and the intended interpretation of '$\Rightarrow$'. With this preliminary discussion out of the way, the key argument is made in §4. The argument is an instance of reasoning by cases. §4.1 considers the supposition that, given some system of reasoning S and arbitrary sentence ϕ, it is not the case that whenever ϕ is paradoxical, it is *derivably* paradoxical. In this case, M&R's motivation for their elimination rule fails. In §4.2, I suppose that if ϕ is paradoxical, then it is derivably so. In this case, the elimination rule must fail if S's derivability relation is arithmetically definable, whereas the motivation for the introduction rule fails if it is not. It is thereby shown that on no understanding of 'is paradoxical' do M&R's cases successfully motivate both of the rules needed to derive triviality. In §4.3, I show that interpreting '$\Rightarrow$' in terms of validity instead of derivability does not rescue M&R's rules. §5 concludes with some remarks on the intuitive reason we should expect the elimination rule for paradoxicality (or, more precisely, the rule it is a special case of) to fail in the cases that it does.

2 The paradox

I'll work with a first-order language, $\mathcal{L}$, which includes an identity predicate, $=$; the logical connectives and quantifiers: $\neg, \vee, \forall$; a sentential absurdity constant $\bot$; and a name-forming function, $\ulcorner \cdot \urcorner$, from sentences ϕ to the closed terms $\ulcorner \phi \urcorner$ that serve as their names. Terms, formulae, and sentences are defined in the usual way. For all theories under consideration, I'll assume that they interpret

I make no claims as to whether or how they can be avoided.

[4]For this reason, Rosenblatt [forthcoming]'s pass at addressing the paradoxicality paradox is at best incomplete. Rosenblatt offers a plausible alternative notion of paradoxicality for which one of M&R's rules fails, but this does not show that there is anything wrong with M&R's notion. Though he also challenges the motivation for M&R's introduction rule, in doing so he makes substantive assumptions about paradoxicality that M&R need not take on. As I'll argue in §4.2, there *are* coherent notions of paradoxicality—including the one arguably closest to what M&R have in mind—on which their introduction rule is very plausible.

enough arithmetic or syntax such that, for all open formulae $\phi(x)$, there is a term $t_\phi := \ulcorner\phi(t_\phi/x)\urcorner$ (where $\phi(t/x)$ is the result of substituting every free occurrence of x in $\phi(x)$ with t).

Sequents are expressions of the form $\Gamma \Rightarrow \phi$, where Γ (the *antecedent*) is a set of sentences and ϕ (the *consequent*) a sentence.[5] Sequents above a horizontal line are *premises*, those below are *conclusions*, and the premise-to-conclusion expressions *rules*. What follows is an axiomatization of classical logic (CL), with quantifier rules and equality rules omitted for simplicity. $\Gamma \Rightarrow \phi\ (\psi)$ means that the given rule holds for both $\Gamma \Rightarrow \phi$ and $\Gamma \Rightarrow \psi$ and a double horizontal line means that the rule works with premise and conclusion flipped.

$$\frac{}{\phi \Rightarrow \phi}\ \text{SRef} \qquad \frac{\Gamma \Rightarrow \psi}{\Gamma, \phi \Rightarrow \psi}\ \text{SWeak} \qquad \frac{\Gamma \Rightarrow \phi \qquad \Delta, \phi \Rightarrow \psi}{\Gamma, \Delta \Rightarrow \psi}\ \text{Cut}$$

$$\frac{\Gamma \Rightarrow \phi\ (\psi)}{\Gamma \Rightarrow \phi \vee \psi}\ \vee\text{-I} \qquad \frac{\Gamma \Rightarrow \phi \vee \psi \qquad \Delta_0, \phi \Rightarrow \chi \qquad \Delta_1, \psi \Rightarrow \chi}{\Gamma, \Delta_0, \Delta_1 \Rightarrow \chi}\ \vee\text{-E}$$

$$\frac{\Gamma \Rightarrow \phi}{\Gamma \Rightarrow \neg\neg\phi}\ \neg\neg\text{-I/E} \qquad \frac{\Gamma, \phi \Rightarrow \bot}{\Gamma \Rightarrow \neg\phi}\ \neg\text{-I} \qquad \frac{\Gamma \Rightarrow \phi \qquad \Delta \Rightarrow \neg\phi}{\Gamma, \Delta \Rightarrow \bot}\ \neg\text{-E} \qquad \frac{\Gamma \Rightarrow \bot}{\Gamma \Rightarrow \phi}\ \bot\text{-E}$$

Given the standard definitions of $\phi \wedge \psi$ as $\neg(\neg\phi \vee \neg\psi)$ and $\phi \to \psi$ as $\neg\phi \vee \psi$, the following familiar rules are derivable:

$$\frac{\Gamma \Rightarrow \phi \qquad \Delta \Rightarrow \psi}{\Gamma, \Delta \Rightarrow \phi \wedge \psi}\ \wedge\text{-I} \qquad \frac{\Gamma \Rightarrow \phi \wedge \psi}{\Gamma \Rightarrow \phi\ (\psi)}\ \wedge\text{-E} \qquad \frac{\Gamma, \phi \Rightarrow \psi}{\Gamma \Rightarrow \phi \to \psi}\ \to\text{-I} \qquad \frac{\Gamma \Rightarrow \phi \qquad \Delta \Rightarrow \phi \to \psi}{\Gamma, \Delta \Rightarrow \psi}\ \to\text{-E}$$

Let $\mathcal{L}_\mathcal{T}$ be $\mathcal{L}$ augmented with a truth predicate, T. The 'naive' rules governing T are:

$$\frac{\Gamma \Rightarrow \phi}{\Gamma \Rightarrow T\ulcorner\phi\urcorner}\ T\text{-I/E} \qquad \frac{\Gamma \Rightarrow \neg\phi}{\Gamma \Rightarrow \neg T\ulcorner\phi\urcorner}\ \neg T\text{-I/E}$$

For any derivation system Z, let ZT be Z with the T rules added. I trust the reader is familiar with the derivation of the triviality of CLT from the Liar sentence, λ, provably equivalent to $\neg T\ulcorner\lambda\urcorner$. What is important for present purposes is that the derivation makes crucial use only of SWeak, SRef, Cut, $\neg$-I, $\neg$-E, $\bot$-E, T-I, and T-E. Solution types correspond to which rule is dropped from CLT to avoid triviality, where K3T is the solution dropping $\neg$-I.[6]

Let $\mathcal{L}_{\mathcal{TP}}$ be $\mathcal{L}_\mathcal{T}$ augmented with a paradoxicality predicate, Par. And let K3TP be K3T with the following rules added:

$$\frac{\Gamma, \phi \vee \neg\phi \Rightarrow \bot}{\Gamma \Rightarrow Par\ulcorner\phi\urcorner}\ Par\text{-I} \qquad \frac{\Gamma \Rightarrow Par\ulcorner\phi\urcorner \qquad \Delta \Rightarrow \phi \vee \neg\phi}{\Gamma, \Delta \Rightarrow \bot}\ Par\text{-E}$$

M&R's central result is the triviality of K3TP.[7] The proof relies on a sentence, ρ, provably equivalent to $T\ulcorner\rho\urcorner \to Par\ulcorner\rho\urcorner$. Standard Curry reasoning derives $\rho \vee \neg\rho \Rightarrow Par\ulcorner\rho\urcorner$. From this together with

[5] M&R use multisets rather than sets, in order to accommodate non-contractive approaches. Since I am setting these aside, I will work with sets.

[6] Note that without $\neg$-I, $\to$-I is no longer derivable. This is important, since one may also derive triviality using $\to$-I/E in place of $\neg$-I/E, via a Curry sentence κ provably equivalent to $T\ulcorner\kappa\urcorner \to \bot$.

[7] M&R also add two 'recapture rules':

$$\frac{\Gamma, \phi \Rightarrow \psi}{\Gamma, \phi \vee \neg\phi \Rightarrow \phi \to \psi}\ \to\text{-I}_w \qquad \frac{\Gamma, \phi \Rightarrow \bot}{\Gamma, \phi \vee \neg\phi \Rightarrow \neg\phi}\ \neg\text{-I}_w$$

However, given a material '$\to$', both these rules are derivable in K3.

$\rho \vee \neg\rho \Rightarrow \rho \vee \neg\rho$ (given by SRef), we can derive $\rho \vee \neg\rho \Rightarrow \bot$ (by Par-E), from which we can then, by Par-I, derive $\Rightarrow Par\ulcorner\rho\urcorner$. From this and the definition of ρ, we can derive $\Rightarrow \rho$, from which we get $\Rightarrow \rho \vee \neg\rho$ by $\vee$-I. This together with $\Rightarrow Par\ulcorner\rho\urcorner$ derives $\Rightarrow \bot$ by Par-E.

3 The motivation

3.1 Initial considerations

To a K3T theorist, Par-I/E will seem immediately suspect. Consider that if we replace $Par\ulcorner\phi\urcorner$ with $\neg(\phi \vee \neg\phi)$ in Par-I/E, the result is $\neg$-I/E restricted to sentences of the form $\phi \vee \neg\phi$. In other words: the introduction and elimination conditions for $Par\ulcorner\phi\urcorner$ are just those which $\neg$-I/E respectively give for negations of $\phi \vee \neg\phi$. But K3T theorists reject $\neg$-I outside of classical contexts; and it is a familiar part of the view that sentences of the form $\phi \vee \neg\phi$ can fail to be classical. To a K3T theorist, then, Par-I/E look like a smuggling in of a negation that conforms to both $\neg$-I/E for instances of $\phi \vee \neg\phi$. That is, the proposed account of 'ϕ is paradoxical' appears to be something like 'NOT($\phi \vee \neg\phi$)' for some such 'NOT'.[8]

So long as their argument is otherwise sound, M&R can surely bite this bullet. They claim to demonstrate an expressive need for a predicate satisfying Par-I/E. If it turns out that this predicate is inferentially equivalent to a negation that satisfies $\neg$-I/E for instances of $\phi \vee \neg\phi$, then, well, there's an expressive need for *that.* And if such a negation is incompatible with the K3T approach, so much the worse for the expressive adequacy of the latter.

This response is fine, as far as it goes. But it highlights the importance of M&R's burden to demonstrate such an expressive need. Without it, K3T theorists have a good story to tell about why any concept governed by Par-I/E is incoherent.

M&R's case for Par-I/E begins with the observation that any extension, S, of K3T, is closed under all rules of CL for all ϕ such that its LEM instance ($\Rightarrow \phi \vee \neg\phi$) is derivable in S. Because of this, the derivability of $\Rightarrow \phi \vee \neg\phi$ seems to be a good characterization, for K3T theorists, of the classicality (and so, perhaps, unparadoxicality) of ϕ. As for paradoxicality, M&R say that "ϕ is *paradoxical-in-S* if and only if $\bot$ follows in S from the assumption that ϕ satisfies [LEM]". Murzi and Rossi [2020, p. 162] In other words: ϕ is paradoxical-in-S iff $\phi \vee \neg\phi \Rightarrow \bot$ is derivable in S.[9]

This proposal ought to strike one as *prima facie* unmotivated. If the derivability of $\Rightarrow \phi \vee \neg\phi$ is a good account of ϕ's unparadoxicality, then the *underivability* of $\Rightarrow \phi \vee \neg\phi$ is the most natural account of its paradoxicality. Indeed, M&R seem at times to assume that for all ϕ, either $\Rightarrow \phi \vee \neg\phi$ is derivable in a given extension of K3T or $\phi \vee \neg\phi \Rightarrow \bot$ is. As we'll see in §4.1, this is not the case, at least given certain rather weak assumptions.

Be that as it may, as I noted in §1, all M&R need is that there be *some* (coherent) notion of paradoxicality that satisfies Par-I/E. To be sure, the derivability of $\phi \vee \neg\phi \Rightarrow \bot$ indicates some

[8]This is not to say that K3T theorists should respond to M&R's paradox by rejecting Par-I. While K3T theorists reject $\neg$-I *for negation*, they very well might accept a rule like $\neg$-I for some other concept. The point is that they would never accept a concept that conforms to analogues of *both* $\neg$-I/E, even when restricted to instances of $\phi \vee \neg\phi$.

[9]Since we're working with sequent calculi, one might naturally translate "$\bot$ follows in S from the assumption that ϕ satisfies LEM" as the condition that $\Rightarrow \bot$ be derivable from $\Rightarrow \phi \vee \neg\phi$ in S; or as the condition that $\phi \vee \neg\phi$ derive $\bot$ in a sentential deduction system that corresponds to the sequent proof procedure of S in the obvious way. Given uncontested structural rules, the two readings are equivalent to the derivability of $\phi \vee \neg\phi \Rightarrow \bot$ in S.

unusual property of ϕ in extensions of K3T; and even if it did not, we surely have independent reason to be able to express derivability claims in general.

But this is just the problem: for all that's been said thus far, we have no motivation for anything other than a derivable-in-S predicate. Assuming the set of S's derivations is decidable, S's derivability relation is arithmetically definable, so there is no need to introduce a new predicate into the language. Furthermore, any paradoxicality predicate defined in terms of a definable derivability predicate will fail to satisfy one of Par-I/E. Assuming the consistency of arithmetic, this follows from M&R's triviality proof.[10]

3.2 The cases

M&R are aware that a definable derivability predicate will not serve their purposes. For this reason, they insist that we must introduce a primitive paradoxicality predicate. The predicate is supposed to facilitate the reasoning featured in the following two cases (where S is, again, some extension of K3T):

> **The logic student**: Lois is a logic student who is learning how to reason in S. She (mistakenly) assumes $\lambda \vee \neg\lambda$. As a result, she carries out the Liar reasoning in S and derives $\perp$. She concludes that assuming that λ satisfies LEM trivializes S. As she puts it, λ is paradoxical: that is, Lois asserts $Par\ulcorner\lambda\urcorner$.
>
> **Misguided reasoning**: Clark reasons in S and assumes that everything that Lois says is paradoxical. Lois asserts that ϕ. As a result, Clark infers that ϕ is paradoxical. However, Clark also proves that ϕ satisfies LEM, and hence all of the principles of classical logic. From his claim that ϕ is paradoxical (that is, such that $\phi \vee \neg\phi$ entails $\perp$), and his proof of $\phi \vee \neg\phi$, Clark concludes $\perp$.
>
> – Murzi and Rossi [2020, p. 163]

Since M&R's argument for a primitive predicate satisfying Par-I/E rests almost entirely on the need to accommodate the reasoning in these two scenarios, it is important to be clear about what's going on in them.

First: the cases are relative to a system of reasoning S.[11] We should not infer from this that paradoxicality is a system-relative notion. Rather, what is going on is as follows. Clearly, the sequent $\phi \vee \neg\phi \Rightarrow \perp$ is in some sense meant to characterize the paradoxicality of ϕ. Now, if $\phi \vee \neg\phi \Rightarrow \perp$ is not derivable *in S* then, so long as we're reasoning in S, we ought not be able to infer that ϕ is paradoxical—even if that sequent is derivable in some other system. This does not mean that the predicate 'is paradoxical' is itself indexed to a system.

Now, one can reply that all that matters is what we can infer about paradoxicality in the system S we *ought* to reason in. I will address this reading of the paradox in §4.2. Until then, I'll proceed leaving S unspecified.

[10] Which one fails depends on exactly which extension of K3T we are defining derivability for and which we are reasoning in. See Field [2017, §8] for a related discussion.

[11] M&R call S a theory, not a system. However, given that a theory is just a set of sentences, it's not clear how one can 'reason in' a theory. So as to avoid letting anything turn on this, let a *system* S be a pair of a derivation procedure (i.e. a set of axioms and rules), D, and a theory, T, such that T is a superset of the set of theorems of D. We can then talk of reasoning in S and of the derivability relation of S even when that derivation procedure doesn't derive all the truths of S (i.e. doesn't derive every element of T from $\varnothing$).

Second, note that though we've been working with sequent calculi, Lois and Clark are reasoning with sentences, not sequents. The obvious interpretation here is that they are reasoning in a system, S', equipped with a sentential derivation procedure in which Γ derives ψ iff $\Gamma \Rightarrow \psi$ is derivable in S. I will proceed with this interpretation and use '$\Gamma \vdash \psi$' as a metalinguistic abbreviation for 'Γ derives ψ in S'' (given some implicit system S). Obviously, $\Gamma \vdash \psi$ iff $\Gamma \Rightarrow \psi$ is derivable (in S).[12]

With this in mind, two initial observations about the reasoning in the cases are in order. The first is that if 'The logic student' is to motivate Par-I, it must be that Lois *derives* (in S') $Par\ulcorner\lambda\urcorner$. If she merely inferred it in some informal way not restricted to what's derivable in S', the move from $\phi \vee \neg\phi \Rightarrow \bot$ to $\Rightarrow Par\ulcorner\phi\urcorner$ sanctioned by Par-I would not be motivated by the case.

Second, the reasoning in 'Misguided reasoning' only seems at all compelling if we assume that if ϕ is paradoxical, then $\phi \vee \neg\phi \vdash \bot$. It would otherwise be entirely mysterious why Clark goes on to derive $\bot$. It is presumably to emphasize this that M&R included the parenthetical in the description of the case.

It seems to follow from the cases, then, that $\vdash Par\ulcorner\phi\urcorner$ iff $\phi \vee \neg\phi \vdash \bot$. This is unsurprising since, given Cut and SRef, Par-E is inferentially equivalent to the bottom-up direction of Par-I (which I'll call Par-E*).[13] So M&R's two rules can alternatively be given as:

$$\frac{\Gamma, \phi \vee \neg\phi \Rightarrow \bot}{\Gamma \Rightarrow Par\ulcorner\phi\urcorner} \; Par\text{-I/E}^*$$

The instances where $\Gamma = \varnothing$ correspond to the equivalence between $\vdash Par\ulcorner\phi\urcorner$ and $\phi \vee \neg\phi \vdash \bot$ suggested by the cases.

3.3 Side antecedents

Taken literally, however, 'The Logic student' and 'Misguided reasoning' do not in fact support Par-I/E*.[14] At best, they might support their restrictions to instances in which $\Gamma = \varnothing$:

$$\frac{\phi \vee \neg\phi \Rightarrow \bot}{\Rightarrow Par\ulcorner\phi\urcorner} \; Par\text{-I}_w/\text{E}^*_w$$

Though M&R's triviality proof relies only on an instance of Par-I that is also an instance of Par-I$_w$, the same is not true of Par-E*. M&R's proof makes crucial use of an instance with side antecedents. It will be instructive to examine the instance of the extra strength of Par-E* relative to Par-E$^*_w$ that is relevant for M&R's proof. This will be more perspicuous if we work with a model-theoretic

[12] Though it does not really fit the reasoning in the cases, I will consider interpretations of '$\Rightarrow$' in terms of *validity* instead of derivability in §4.3.

[13] The derivations proving the equivalence:

$$\dfrac{\dfrac{\Gamma \Rightarrow Par\ulcorner\phi\urcorner}{\Gamma, \phi \vee \neg\phi \Rightarrow \bot}\,Par\text{-E}^* \qquad \Delta \Rightarrow \phi \vee \neg\phi}{\Gamma, \Delta \Rightarrow \bot}\,\text{Cut} \qquad \dfrac{\Gamma \Rightarrow Par\ulcorner\phi\urcorner \qquad \dfrac{}{\phi \vee \neg\phi \Rightarrow \phi \vee \neg\phi}\,\text{SRef}}{\Gamma, \phi \vee \neg\phi \Rightarrow \bot}\,Par\text{-E}$$

[14] It will sometimes be useful for purposes of presentation, especially throughout this subsection, to work with Par-E* rather than Par-E. Since Cut and SRef are not under contention, this is harmless.

account of K3T, in which interpretations map sentences of our language to $\{0, \frac{1}{2}, 1\}$. Interpretations are constrained as follows:[15]

$|\neg\phi| = 1 - |\phi|$
$|\phi \vee \psi| = max\{|\phi|, |\psi|\}$
$|\phi \wedge \psi| = min\{|\phi|, |\psi|\}$
$|T\ulcorner\phi\urcorner| = |\phi|$
$|\bot| = 0$

$\Gamma \models \psi$ iff every interpretation in which $|\gamma|$=1 for every $\gamma \in \Gamma$ is an interpretation in which $|\psi| = 1$. So $\models \psi$ iff $|\psi| = 1$ in every interpretation. For the purposes of this subsection, let us suppose we are working with a system S whose derivation procedure is complete, such that $\Gamma \models \psi$ iff $\Gamma \Rightarrow \psi$ is derivable (iff $\Gamma \vdash \psi$).

Note that I have not given a clause for the interpretation of Par sentences. This is because I want to compare the constraints that would have to be placed on interpretations of Par sentences in order to validate Par-E^*_w versus those needed to validate Par-E.

Since $\phi \rightarrow \psi$ is defined as $\neg\phi \vee \psi$, the sentence featuring in M&R's triviality proof, ρ, is provably equivalent to $\neg T\ulcorner\rho\urcorner \vee Par\ulcorner\rho\urcorner$. It is easy to check that in every interpretation, as a result of the truth, negation, and disjunction clauses, $|Par\ulcorner\rho\urcorner| = 1$ iff $|\rho| = 1$; and $|Par\ulcorner\rho\urcorner| \neq 1$ iff $|\rho| = \frac{1}{2}$ (this is so regardless of whether we admit of interpretations in which $|Par\ulcorner\rho\urcorner| = \frac{1}{2}$). So, for every interpretation, $|\rho| \neq 0$.

Now, suppose we wish to validate Par-E^*_w. Here's the instance for ρ:

$$\dfrac{\Rightarrow Par\ulcorner\rho\urcorner}{\rho \vee \neg\rho \Rightarrow \bot}\,{\scriptstyle Par\text{-}\mathrm{E}^*_w}$$

Given the equivalence of '$\models$' claims and the derivability of corresponding sequents, this instance of Par-E^*_w requires that either $\models Par\ulcorner\rho\urcorner$ and $\rho \vee \neg\rho \models \bot$, or $\not\models Par\ulcorner\rho\urcorner$. Now consider the instance of Par-E^* M&R need for their proof:

$$\dfrac{\rho \vee \neg\rho \Rightarrow Par\ulcorner\rho\urcorner}{\rho \vee \neg\rho \Rightarrow \bot}\,{\scriptstyle Par\text{-}\mathrm{E}^*}$$

Since in every interpretation $|\rho| \neq 0$, it follows that $|\rho \vee \neg\rho| = 1$ iff $|\rho| = 1$ iff $|Par\ulcorner\rho\urcorner| = 1$. As a result, $\rho \vee \neg\rho \models Par\ulcorner\rho\urcorner$, *regardless* of how Par sentences are interpreted. So this instance of Par-E^* requires that $\rho \vee \neg\rho \models \bot$. Therefore, adopting Par-E^* forces us to accept that $\rho \vee \neg\rho \models \bot$ (i.e. that $\rho \vee \neg\rho$ never takes value 1), whereas we can allow interpretations in which $|\rho \vee \neg\rho| = 1$ while still validating Par-E^*_w. Since the Clark scenario only justifies Par-E^*_w, the instance of Par-E^* M&R need for their proof is unwarranted.

One might object to this as follows. Suppose $\rho \vee \neg\rho \not\models \bot$. Plausibly, $Par\ulcorner\rho\urcorner$ is meant to be an object language rendering of $\rho \vee \neg\rho \vdash \bot$. If so, it follows that $\neg Par\ulcorner\rho\urcorner$, and so that $|Par\ulcorner\rho\urcorner| = 0$ in every interpretation. But then $|\rho| = \frac{1}{2}$ in every interpretation, so $\rho \vee \neg\rho \models \bot$ (vacuously). Contradiction. So $\rho \vee \neg\rho \models \bot$, regardless of whether we adopt Par-E^*.

[15] Interpretations should also be constrained by quantifier rules and rules of arithmetic or syntax sufficient to generate self-reference, but we need not make this explicit for present purposes. Setting that aside, we may take there to be an interpretation for every mapping of sentences of $\mathcal{L}_{\mathcal{TP}}$ to elements of $\{0, \frac{1}{2}, 1\}$ that satisfies the listed constraints.

The problem with this objection is, of course, the move from $\neg Par\ulcorner\rho\urcorner$ to *in every interpretation*, $|Par\ulcorner\rho\urcorner| = 0$. The argument can adapted to show that (supposing one interprets $Par\ulcorner\rho\urcorner$ as the object language rendering of $\rho \vee \neg\rho \vdash \bot$) $\neg Par\ulcorner\rho\urcorner$ is *in fact* true and that we must therefore reject ρ (and its negation) and $\rho \vee \neg\rho$ (and its negation). But the generalization to all interpretations is unwarranted. Similarly, supposing instead that $\rho \vee \neg\rho \vdash \bot$ is meant to be equivalent to $\vdash Par\ulcorner\rho\urcorner$ implies that from $\rho \vee \neg\rho \not\models \bot$ we may infer $\not\models Par\ulcorner\rho\urcorner$. That falls short of implying $\models \neg Par\ulcorner\rho\urcorner$, which is what is needed for the objection.

And so, if we restrict ourselves to rules warranted by M&R's cases, the argument so far suffices to block the full *Par*-E.

A natural thought is that M&R can respond to this by simply generalizing 'Misguided reasoning' so as to warrant the full *Par*-E. The generalization might go as follows:

> **Misguided reasoning (generalized)**: Clark reasons in S and assumes (for whatever reason) that Γ derives that ϕ is paradoxical. However, Clark also derives $\phi \vee \neg\phi$ from Δ. From his claim that Γ entails that ϕ is paradoxical (that is, that Γ entails that $\phi \vee \neg\phi$ entails $\bot$), and his derivation of $\phi \vee \neg\phi$ from Δ, Clark concludes that Γ, Δ derives $\bot$.

Though less immediately intuitive, Clark's reasoning here seems to be a generalization of his reasoning in the original 'Misguided reasoning'. This generalized form is just as unobjectionable as the special case used by M&R (the argument goes), so we have just as much reason to adopt *Par*-E as we did to adopt *Par*-E$_w$.

There is at least one sense in which this generalized form is *not* as unobjectionable as the special case: we've just worked through a counterexample to this general form—viz. where ϕ is ρ and $\Gamma = \Delta = \{\rho \vee \neg\rho\}$—that does not arise for the special case. But appealing to this to motivate rejection of *Par*-E arguably begs the question in the present context. Supposing 'Misguided reasoning (generalized)' really does seem (upon sufficient reflection) like good reasoning, then the presence of counterexamples in extensions of K3T would support M&R's claim that K3T is expressively inadequate, rather than challenge the coherence of 'Misguided reasoning (generalized)'.

What we must now do, then, is investigate whether or not we have compelling reason to think that there is a coherent sense of 'paradoxical' for which 'The logic student' and 'Misguided reasoning (generalized)' both do seem like good reasoning. It is important to stress that without some such reason, the K3T theorist can and should maintain that M&R's own triviality proof shows any concept governed by *Par*-I$_w$/E to be incoherent.

4 What is paradoxicality?

4.1 If the truth of $Par\ulcorner\phi\urcorner$ does not imply its derivability

Part of the difficulty in evaluating M&R's cases (and the rules they're meant to justify) is that they only tell us how we ought to *derive* paradoxicality claims in S. But we cannot evaluate proposed derivation rules for a concept without an independent account of what that concept is. It is unhelpful to reply that it is the derivation rules that give the meaning of the concept. For whether these rules correspond to a coherent concept at all is the question at issue. In any case, we're able to debate the derivation rules for say, negation, in part because we have an independent concept of negation against which we can evaluate the adequacy of the rules—whether or not those rules, when adequate, 'give the meaning' of negation. But in the case of paradoxicality, it's not clear what this independent concept is.

To get past this difficulty, we should first recall that the cases require that $\vdash Par\ulcorner\phi\urcorner$ iff $\phi\vee\neg\phi\vdash\bot$ (where, again, $\Gamma\vdash\psi$ iff $\Gamma\Rightarrow\psi$ is derivable in S). So any sense of 'paradoxical' that is to do the work of justifying the reasoning in the cases cannot violate this equivalence.

With this observation in place, we can proceed by considering the inferential relationship between the truth (in S)[16] of $Par\ulcorner\phi\urcorner$ and $\vdash Par\ulcorner\phi\urcorner$. Assuming the derivation procedure of S is sound, $Par\ulcorner\phi\urcorner$ whenever $\vdash Par\ulcorner\phi\urcorner$. But what about the other direction? If $\vdash Par\ulcorner\phi\urcorner$ whenever $Par\ulcorner\phi\urcorner$, then, since $\vdash Par\ulcorner\phi\urcorner$ iff $\phi\vee\neg\phi\vdash\bot$, it follows that $Par\ulcorner\phi\urcorner$ iff $\phi\vee\neg\phi\vdash\bot$. We may then take this to be (at least extensionally) the needed independent account of paradoxicality. Indeed, I suspect this to be the notion of paradoxicality M&R have in mind. However, securing that $\vdash Par\ulcorner\phi\urcorner$ whenever $Par\ulcorner\phi\urcorner$ requires S to have certain properties that must be considered carefully. I'll return to that option in §4.2.

For now, let us suppose it not to be the case that $\vdash Par\ulcorner\phi\urcorner$ whenever $Par\ulcorner\phi\urcorner$. Then, since $\vdash Par\ulcorner\phi\urcorner$ iff $\phi\vee\neg\phi\vdash\bot$, it follows that there are ϕ such that $Par\ulcorner\phi\urcorner$ but $\phi\vee\neg\phi\nvdash\bot$. What, then, might the paradoxicality of ϕ amount to?

Here's perhaps the most plausible account that meets these conditions: $Par\ulcorner\phi\urcorner$ iff $\phi\vee\neg\phi\vdash\psi$ for some ψ that is false (or indeterminate) in S. Given such an account, there are guaranteed to be ϕ that are paradoxical though $\phi\vee\neg\phi\nvdash\bot$. For example: suppose $\phi\vee\neg\phi\vdash\psi$ for a ψ that is false in S but not derivably so (perhaps because ψ is a mathematical claim that is undecidable in S, or is otherwise contingent relative to the proof procedure of S). Then $\psi\nvdash\bot$, so there is no guarantee that $\phi\vee\neg\phi\vdash\bot$. In fact, assuming we have such false-but-unprovably-so ψ in S, we are guaranteed cases where $\phi\vee\neg\phi\nvdash\bot$. One such is the Curry sentence κ_ψ, provably equivalent to $T\ulcorner\kappa_\psi\urcorner\to\psi$. In extensions of K3T, $\kappa_\psi\vee\neg\kappa_\psi\vdash\psi$,[17] though $\kappa_\psi\vee\neg\kappa_\psi\nvdash\bot$.[18]

Note that on this account of paradoxicality, it will never both be the case that $\vdash Par\ulcorner\phi\urcorner$ and $\vdash\phi\vee\neg\phi$. If it were so, then one could derive some ψ that is false (or indeterminate) in S. But if S is sound, this cannot happen. As a consequence, any sequent rule with $\Rightarrow Par\ulcorner\phi\urcorner$ and $\Rightarrow\phi\vee\neg\phi$ as sole premises will be vacuously valid, including Par-E$_w$.

Furthermore, Par-I$_w$ is extremely plausible. For, if $\phi\vee\neg\phi$ derives $\bot$ in S, then it derives something that is *provably* false in S. So, given the account of paradoxicality at play, one ought to be able to *derive* $Par\ulcorner\phi\urcorner$ *in S.* It follows that this account of paradoxicality supports both of Par-I$_w/E_w$, and so the material equivalence between $\phi\vee\neg\phi\vdash\bot$ and $\vdash Par\ulcorner\phi\urcorner$.

However, the account does not support the full strength Par-E, nor Clark's reasoning in 'Misguided reasoning (generalized)'. Supposing $\Gamma\vdash Par\ulcorner\phi\urcorner$ and $\Delta\vdash\phi\vee\neg\phi$, it follows only that $\Gamma,\Delta\vdash\psi$ for some false or indeterminate ψ. There is no way to argue for the stronger conclusion that

[16] For brevity, let '$Par\ulcorner\phi\urcorner$', when not preceded by '$\vdash$' or '$\Rightarrow$', hereafter abbreviate '$Par\ulcorner\phi\urcorner$ is true in S'.

[17] Sequent calculus proof: From $\kappa_\psi\Rightarrow\kappa_\psi$ we can derive $\kappa_\psi\Rightarrow T\ulcorner\kappa_\psi\urcorner$ by T-I. $\kappa_\psi\Rightarrow\kappa_\psi$ is itself equivalent to $\kappa_\psi\Rightarrow T\ulcorner\kappa_\psi\urcorner\to\psi$, so $\to$-E and Cut deliver $\kappa_\psi\Rightarrow\psi$. Since κ_ψ is equivalent to $\neg T\ulcorner\kappa_\psi\urcorner\vee\psi$, we can derive $\neg\kappa_\psi\Rightarrow\kappa_\psi$ from $\neg\kappa_\psi\Rightarrow\neg\kappa_\psi$ by $\neg\neg$-E, T-E, and (derivable) DeMorgan equivalence. $\neg\kappa_\psi\Rightarrow\psi$ then follows by Cut, and $\kappa_\psi\vee\neg\kappa_\psi\Rightarrow\psi$ follows by $\vee$-E from $\kappa_\psi\vee\neg\kappa_\psi\Rightarrow\kappa_\psi\vee\neg\kappa_\psi$.

[18] In the case where ψ is an undecidable mathematical claim, this can also be proved: Since κ_ψ is equivalent to $\neg T\ulcorner\kappa_\psi\urcorner\vee\psi$, we can derive $\psi\Rightarrow\kappa_\psi\vee\neg\kappa_\psi$ from $\psi\Rightarrow\psi$ by two instances of $\vee$-I. If $\kappa_\psi\vee\neg\kappa_\psi\Rightarrow\bot$ were derivable, it would then follow by Cut that $\psi\Rightarrow\bot$ is as well. But since ψ is undecidable, $\psi\Rightarrow\bot$ is not derivable. So $\kappa_\psi\vee\neg\kappa_\psi\Rightarrow\bot$ is not either.

$\Gamma, \Delta \vdash \bot$. On the contrary, as we've seen, so long as S contains undecidable falsehoods, there will be counterexamples.

More generally: take any account of paradoxicality for which (i) $\vdash Par\ulcorner\phi\urcorner$ iff $\phi \vee \neg\phi \vdash \bot$; and (ii) it is not the case that $\vdash Par\ulcorner\phi\urcorner$ whenever $Par\ulcorner\phi\urcorner$. It follows that there are cases of $Par\ulcorner\phi\urcorner$ such that $\phi \vee \neg\phi \nvdash \bot$. If there are such cases, then we have no compelling reason to think Par-E valid. We can put this in terms of Clark's reasoning in 'Misguided reasoning (generalized)'. Supposing Γ derives $Par\ulcorner\phi\urcorner$ and Δ derives $\phi \vee \neg\phi$, Clark needs some warrant to go on to infer that Γ, Δ derives $\bot$. In 'Misguided reasoning (generalized)', that warrant is provided by the parenthetical: "(... Γ entails that $\phi \vee \neg\phi$ entails $\bot$)". However, we are by stipulation considering accounts on which $Par\ulcorner\phi\urcorner$ *does not* entail that $\phi \vee \neg\phi$ entails $\bot$. So without some alternative motivation, Clark's reasoning is simply unwarranted.

To be clear: I have not proved that there *can* be no account of paradoxicality that both satisfies conditions (i) and (ii) and fills in the gap in Clark's reasoning. It is, however, utterly unclear what such an account might be. As we've seen, on perhaps the most plausible account of paradoxicality that satisfies conditions (i) and (ii), there are counterexamples to Clark's reasoning. In any case, until a reason to think there is some such account is offered, K3T theorists can and should maintain that M&R's triviality proof shows that there simply isn't any.

4.2 If $Par\ulcorner\phi\urcorner$ is true iff it is derivable

The argument above rests on the supposition that it is not the case that $\vdash Par\ulcorner\phi\urcorner$ whenever $Par\ulcorner\phi\urcorner$. If we drop that assumption, we can consider the account of paradoxicality that seems most in line with M&R's argument, viz. one where $Par\ulcorner\phi\urcorner$ is (at least extensionally equivalent to) $\phi \vee \neg\phi \vdash \bot$. On this account, Par-I_w is valid iff $\vdash Par\ulcorner\phi\urcorner$ whenever $Par\ulcorner\phi\urcorner$. But why should we take this to hold? The only natural answer is that it should hold whenever we are reasoning in an S that can derive the facts about its own derivability relation.[19] Let us grant, then, that M&R intend their rules to apply just for systems S that satisfy this condition.

There is now a problem. Let $Der(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner)$ be the object-language rendering of $\phi \vdash \psi$ (there must be some such rendering if S is to derive the facts about its own derivability relation). Then $Par\ulcorner\phi\urcorner$ iff $Der(\ulcorner\phi \vee \neg\phi\urcorner, \ulcorner\bot\urcorner)$.[20] This follows from how we've stipulatively introduced Der and the fact that $Par\ulcorner\phi\urcorner$ iff $\phi \vee \neg\phi \vdash \bot$. So the validity of Par-I_w/E implies the validity of:

$$\frac{\phi \vee \neg\phi \Rightarrow \bot}{\Rightarrow Der(\ulcorner\phi \vee \neg\phi\urcorner, \ulcorner\bot\urcorner)} \qquad \frac{\Gamma \Rightarrow Der(\ulcorner\phi \vee \neg\phi\urcorner, \ulcorner\bot\urcorner) \qquad \Delta \Rightarrow \phi \vee \neg\phi}{\Gamma, \Delta \Rightarrow \bot}$$

These are, of course, just special cases of the more general rules:

$$\frac{\phi \Rightarrow \psi}{\Rightarrow Der(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner)}\ Der\text{-}\mathrm{I}_w \qquad \frac{\Gamma \Rightarrow Der(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner) \qquad \Delta \Rightarrow \phi}{\Gamma, \Delta \Rightarrow \psi}\ Der\text{-E}$$

Now, if Der is arithmetically defined, then it is susceptible to Gödel's theorems and their consequences. The latter include Löb's Theorem: the result that $\vdash \phi$ whenever $\vdash Prov\ulcorner\phi\urcorner \to \phi$ (for all

[19] What if S is just stipulated to be the correct system for reasoning about paradoxicality claims? This won't work because we cannot assume that even that system derives every true paradoxicality claim. The true theory of paradoxicality may be unaxiomatizable.

[20] Again, read '$Der(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner)$', when not preceded by '$\vdash$' or '$\Rightarrow$', as '$Der(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner)$ is true in S'.

ϕ, where *Prov* is the defined provability predicate of the system in question). Supposing that (i) *Der* claims satisfy LEM (which they presumably do, since they are claims of arithmetic); and (ii) $Prov\ulcorner\phi\urcorner := Der(\ulcorner\top\urcorner, \ulcorner\phi\urcorner)$; it follows that, given *Der*-E, $\vdash Prov\ulcorner\phi\urcorner \rightarrow \phi$ for all ϕ.[21] Given Löb's Theorem, this implies the triviality of S. So *Der*-E cannot hold unrestrictedly for S.

M&R consider this objection and deny that *Par* is to be interpreted via a defined derivability predicate. But this does not avoid the problem. If $Par\ulcorner\phi\urcorner$ is extensionally equivalent to $\phi \vee \neg\phi \vdash \bot$, and we are considering a system that derives its own derivability facts (expressed via the predicate *Der*), then *Der* claims have to satisfy rules corresponding to the rules governing *Par*, even if *Par* is a primitive. This fact cannot be stipulated away.

It does not help to insist that *Der* be a primitive as well, so as to avoid the problematic consequence of Löb's Theorem. If the derivability relation of S is a derivability relation in the usual sense—i.e. the closure of a decidable set of axioms under a decidable set of rules—then it *is* arithmetically definable, in the standard way. So, if the primitive *Der* really does express the derivability relation of S, then the validity of *Der*-I_w/E will imply the validity of the same rules for the defined derivability predicate.

One could try to defend M&R's position as follows. Suppose that we're reasoning in an S whose derivability relation is the relation of *intuitive deductive reasoning*. This, the argument goes, *is* a derivability relation, albeit an informal one. However, it is not reducible to closure of a decidable set of axioms under a decidable set of rules. So if we restrict M&R's rules to systems S closed under this derivability relation—these are presumably the systems of most interest for reasoning anyway—the argument for the definability of derivability fails.

The problem with this suggestion is that it seems to be part of the concept of deductive demonstration that such demonstrations are effectively recognizable as such. This is part of what makes (informal) mathematical proofs compelling: whether they really are proofs is something we can systematically check line-by-line via an effective procedure. Plausibly, it follows that the set of intuitive derivations is decidable;[22] and so that the intuitive derivability relation is definable.[23]

But suppose that this reply is wrong and that the very fact that *Der*-E fails if the set of intuitive derivations is decidable shows that it is not decidable. Now that we have settled on a specific derivability relation, we must ask whether it is among those that derives its own derivability facts. It is tempting to expect that it does, based on the following argument. Suppose that you (intuitively) derive ψ from ϕ. Then, from the very fact that you've made this derivation, you can infer that ψ is intuitively derivable from ϕ. In other words: you can (intuitively) derive the derivability fact.

Note, however, that this argument makes implicit but crucial use of the recognizability of derivations. To infer from an intuitive derivation of ψ from ϕ that ψ is intuitively derivable from ϕ, one must be able to recognize *that* the derivation is a derivation. But if the set of intuitive derivations isn't decidable, then derivations aren't effectively recognizable as such. This makes *Der*-I_w extremely dubious. Suppose, for example, that I intuitively derive $\bot$ from $\lambda \vee \neg\lambda$. Under current

[21]Proof: The SRef instance of $Der(\ulcorner\top\urcorner, \ulcorner\phi\urcorner)$ and $\Rightarrow \top$ derive $Der(\ulcorner\top\urcorner, \ulcorner\phi\urcorner) \Rightarrow \phi$ by *Der*-E. From this we can derive $Der(\ulcorner\top\urcorner, \ulcorner\phi\urcorner) \vee \neg Der(\ulcorner\top\urcorner, \ulcorner\phi\urcorner) \Rightarrow Der(\ulcorner\top\urcorner, \ulcorner\phi\urcorner) \rightarrow \phi$ (by $\rightarrow$-I_w). This together with $\Rightarrow Der(\ulcorner\top\urcorner, \ulcorner\phi\urcorner) \vee \neg Der(\ulcorner\top\urcorner, \ulcorner\phi\urcorner)$ derives $\Rightarrow Der(\ulcorner\top\urcorner, \ulcorner\phi\urcorner) \rightarrow \phi$ via Cut.

[22]See Priest [2006, §3.2] for an elaboration of this argument.

[23]One might worry that if intuitive derivability were arithmetically definable, we would be able to prove the Gödel sentence for it (by the same method that we intuitively infer the truth of Gödel sentences in general). It would follow that the sentence is both provable and not provable, and so that systems closed under intuitive deductive reasoning are inconsistent. However, it is not clear whether one really could prove the Gödel sentence under these conditions. I cannot go into this matter here, but see Field [2019, §6.7] and Priest [2019, §14.4] for discussion.

suppositions, I cannot effectively recognize whether this derivation really is a derivation. As a result, I should not be able to derive $Der(\ulcorner\lambda \vee \neg\lambda\urcorner, \ulcorner\bot\urcorner)$ merely from that derivation.

To summarize the overall argument: if it is not the case that $\vdash Par\ulcorner\phi\urcorner$ whenever $Par\ulcorner\phi\urcorner$, then Par-E is entirely unmotivated (this was shown in §4.1). But if it is the case, then, given the required equivalence between $\phi \vee \neg\phi \vdash \bot$ and $\vdash Par\ulcorner\phi\urcorner$, it follows that $Par\ulcorner\phi\urcorner$ iff $\phi \vee \neg\phi \vdash \bot$. So the validity of Par rules requires the validity of corresponding rules for a derivability predicate, Der. If we are working with a system S whose derivability relation is definable, then, as a result of Löb's Theorem, the elimination rule for Der (of which Par-E is just a special case) must fail. If it is not definable, it follows that we are working with an S whose set of derivations is not decidable—and so whose derivations are not effectively recognizable as such. In this case, the introduction rule for Der (of which Par-I_w is just a special case) is, at best, unmotivated and highly suspect.

4.3 Validity instead of derivability?

The preceding argument rests on reading the words 'derives', 'infers', 'proves', and 'entails' in M&R's cases uniformly as referring to derivation (in a sentential proof system corresponding to the sequent calculus axiomatization M&R work with). I believe this is the intended reading. It would not do to read them in a non-uniform way, for then the move from the cases to M&R's sequent rules would rest on equivocating on the interpretation of '$\Rightarrow$'. And since the cases are supposed to exemplify *reasoning* that must be accommodated, derivation is by far the most natural option.

Nevertheless, it is worth noting that interpreting the derivability of $\Gamma \Rightarrow \psi$ (in S) as the *validity* of the move from Γ to ψ (in S) does no better at motivating M&R's rules. Let '$\Gamma \twoheadrightarrow \psi$' abbreviate 'the inference from Γ to ψ is valid in S'.[24] We may now ask whether $\twoheadrightarrow Par\ulcorner\phi\urcorner$ whenever $Par\ulcorner\phi\urcorner$. If not, the same argument from §4.1 goes through, mutatis mutandis.

Matters are not quite as straightforward if it is the case, because the argument of §4.2 makes crucial appeal to facts that are distinctively about derivability. However, a similar argument will work. The supposition that $\twoheadrightarrow Par\ulcorner\phi\urcorner$ whenever $Par\ulcorner\phi\urcorner$ is only motivated if S's validities are valid *in* S. So S must contain a validity predicate, Val; and, since M&R's rules now require that $\twoheadrightarrow Par\ulcorner\phi\urcorner$ iff $\phi \vee \neg\phi \twoheadrightarrow \bot$, it follows that $Par\ulcorner\phi\urcorner$ iff $Val(\ulcorner\phi \vee \neg\phi\urcorner, \ulcorner\bot\urcorner)$. So the validity of Par-I_w/E requires the validity of rules which are just special cases of rules, Val-I_w/E, exactly analogous to the Der rules.[25]

However, if we're working with an S whose validity relation is set-theoretically definable, one of these rules must fail. In particular, if S's validities are themselves valid in S, Val-E must fail. This follows from the consistency of set theory. But if Val isn't set-theoretically definable, there's no reason to suppose Val-I to hold. Recall that §4.2 featured an argument for Der-I, based on the deductive closure of systems governed by informal deductive reasoning. But the only way to make a similar argument for validity is indirectly, via the fact that intuitive derivability presumably implies intuitive validity. This will of course just run into the same problem as the original argument for Der-I. And while we can simply stipulate (as suggested in n. 19) interest in systems which contain all truths about validity (or paradoxicality), we cannot similarly stipulate a system in which all truths about validity (or paradoxicality) are *valid*. There may be no such system.

[24] To make sense of 'valid in S', we must expand the notion of *system* introduced in n. 11. We now work with systems which are triples of a derivation procedure, a theory, *and a validity relation*.

[25] Note that on this reading, the paradoxicality paradox is just a special case of the validity Curry paradox. See Beall and Murzi [2013].

5 A concluding thought on Der-E

The advocate of revenge has one more available move. Let us consider again the case where sequents are interpreted via '$\vdash$' claims; where '$\vdash$' and its object-language counterpart 'Der' refer to the relation of derivability by intuitive deductive reasoning; and where we accept the arguments in favor of the definability of the latter. In this case, the argument against Der-E (and so Par-E) was via Löb's Theorem. Importantly, it was *not* against the intuitive appeal of Der-E.

One can object here that if we have an expressive need for Der-E, it is ad hoc to drop the rule instead of weakening the logic yet do the opposite for the truth rules.[26] But, the objection might continue, *of course* we have such an expressive need: it's the need to detach on (intuitive) derivability claims! So one ought to treat the paradoxes on a par: either keep classical logic and restrict both the Der and truth rules, or else weaken the logic enough to avoid the consequences of Löb's Theorem. Call this the parity argument.

As stated, it is too quick. K3T theorists acknowledge and claim to fulfill the burden of demonstrating an expressive need for the *unrestricted* truth rules (see Field [2008, §13]). Otherwise, they would have no case to offer against classical logicians who restrict the truth rules but keep them in ordinary settings. Analogously, what is needed to successfully run the parity argument is a demonstrated expressive need for *unrestricted* Der-E. For this purpose, paradigm cases like 'Misguided reasoning (generalized)' do not suffice.

In fact, it's doubtful whether there is any such expressive need. The expressive need to detach on intuitive derivability claims would seem to be covered by a restriction of Der-E to instances where the Der claim in question is *true*. After all, it is only in such cases that one may actually detach. What's more, there's good reason to think Der-E *should* fail otherwise. This is because derivability relations should not reflect false claims about them.

For example: let ϕ be some arbitrary contingent sentence and χ be the claim 'Adjunction is not valid according to intuitive deductive reasoning'. Presumably, $\{\phi, \chi\}$ intuitively derives $\phi \wedge \chi$, despite the content of χ—precisely because χ is false. Instances of Der-E where the Der claim is false should fail by analogous reasoning. This is perhaps clearer when we consider the rule $\phi, Der(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner) \Rightarrow \psi$, equivalent to Der-E given Cut and SRef. If $Der(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner)$ is false, then it together with ϕ *should not* suffice to derive ψ—that it precisely what its falsity comes to.[27]

One may object that what is now missing is the ability to reason (non-trivially) with counterlogicals. But if having this ability requires that the turnstile reflect whatever object-language claims are made about it in the premise-set, even when those claims are false, then no familiar rule of logic can hold unrestrictedly. Those who reject the unrestricted Der-E can surely insist that if this is what non-trivial reasoning with counterlogicals comes to, such reasoning is not coherent.

With this in mind, let me return to the parity argument. If it is sound, then K3T theorists bear the burden of showing that a restriction of Der-E exists that is (i) weak enough to avoid paradox in K3T augmented with Der-I_w; and (ii) strong enough to accommodate whatever expressive needs

[26] M&R briefly argue something in the neighborhood of this in their §6.1.

[27] An argument like this is briefly pushed in Zardini [2013, p. 636–7]. A similar point is made by Russell even as far back as 1903: "It should be observed that the method of supposing an axiom false, and deducing the consequences of this assumption... is not here universally available. For all our axioms are principles of deduction; and if they are true, the consequences which appear to follow from the employment of an opposite principle will not really follow, so that arguments from the supposition of the falsity of an axiom are here subject to special fallacies." Russell [2009, p. 15–16].

we do in fact have to reason with intuitive derivability claims. Though the suggestion still needs to be worked out in detail, the considerations above strongly suggest that a restriction of *Der*-E to instances featuring true *Der* claims can successfully discharge this burden.[28]

References

Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72(19):690–716, 1975. URL `https://doi.org/10.2307/2024634`.

Hartry Field. *Saving Truth from Paradox*. Oxford University Press, Oxford, 2008.

Julien Murzi and Lorenzo Rossi. Generalized revenge. *Australasian Journal of Philosophy*, 98(1): 153–177, 2020. URL `https://doi.org/10.1080/00048402.2019.1640323`.

Lucas Rosenblatt. Paradoxicality without paradox. *Erkenntnis*, pages 1–20, forthcoming. URL `https://doi.org/10.1007/s10670-021-00405-w`.

Hartry Field. Disarming a paradox of validity. *Notre Dame Journal of Formal Logic*, 58(1):1–19, 2017. URL `https://doi.org/10.1215/00294527-3699865`.

Graham Priest. *In Contradiction*. Oxford University Press, Oxford, 2nd edition, 2006.

Hartry Field. Paraconsistent or paracomplete? In Can Başkent and Thomas Macaulay Ferguson, editors, *Graham Priest on Dialetheism and Paraconsistency*, pages 73–125. Springer Verlag, 2019.

Graham Priest. Some comments and replies. In Can Başkent and Thomas Macaulay Ferguson, editors, *Graham Priest on Dialetheism and Paraconsistency*, pages 575–675. Springer Verlag, 2019.

Jc Beall and Julien Murzi. Two flavors of Curry's paradox. *Journal of Philosophy*, 110(3):143–165, 2013. URL `https://doi.org/10.5840/jphil2013110336`.

Elia Zardini. Naïve logical properties and structural properties. *Journal of Philosophy*, 110(11): 633–644, 2013. URL `https://doi.org/10.5840/jphil2013110118`.

Bertrand Russell. *Principles of Mathematics*. Routledge, 2009.

[28]I extend a warm thanks to Graham Priest, Lorenzo Rossi, Julien Murzi, Marko Malink, Kit Fine, an anonymous referee, and especially Hartry Field, for helpful comments and discussion on earlier versions of this paper.