The Review of Symbolic Logic
Volume 19, Number 4, December 2026

ON SELF-DISOBEDIENT LOGICS

BAS KORTENBACH
Scuola Normale Superiore

Abstract. Mixed logics have recently exploded in popularity, especially as a means of resolving semantic and other paradoxes. However, the most prominent such logics suffer from self-disobedience: they don't follow their own logical laws. Advocates of the affected logics have maintained that, contrary to first appearances, self-disobedience is not problematic. This article has two goals. First, to clarify self-disobedience by distinguishing six different versions and determining how they relate to each other and to mixed-ness. Secondly, to make the case that all three of the arguments provided by proponents of mixed logics to justify self-disobedience are unsuccessful. A more plausible fourth defense is also considered, but found not to be applicable to the most important mixed logic, $\mathbf{ST}^{\omega}$. This drastically undermines the appeal of the associated approach to semantic paradoxes.

§1. Introduction. Suppose we find ourselves in discussion with another party, attempting to convince them of some thesis A, and at a certain point we get them to agree that certain premises $P_1, \ldots, P_n$ would logically entail A. Suppose, moreover, that in fact our interlocutor maintains all of $P_1, \ldots, P_n$. Nonetheless, to our own amazement, they remain unwilling to grant A—all the while admitting that it follows logically from premises which they grant. It seems clear that there is something incoherent about their stance. To maintain that $P_1, \ldots, P_n$, and also that $P_1, \ldots, P_n$ entail A, while refusing to grant A, is to defy one's own logic. Yet the advocates of some prominent mixed logics appear to hold just such a position. How, if at all, can that be tenable?

A logic is mixed if it uses different assessment standards for the premises as opposed to the conclusions of inferences. Logics of this kind have been proposed for various purposes at least since the 1980s (brief historical overviews are in [7, 10, 36]). Some mixed logics have recently generated much interest for their applications to semantic paradoxes and paradoxes of vagueness [8, 9, 18, 26, 34]. That includes the logics in the so-called **ST**-hierarchy, introduced below. I'll refer to the project of avoiding semantic paradoxes via **ST**-hierarchy logics as the *BA-plan*, following [3].

One major worry about the BA-plan is that the **ST** logics are *self-disobedient* [28, 39]. Much like our strange interlocutor in the first paragraph, these logics declare some inferences valid, while also accepting all of their premises, yet without accepting their conclusions. In response, BA-planners have come up with a number of arguments

Received: January 14, 2026 UTC.
2020 *Mathematics Subject Classification*: Primary 03A05, Secondary 03B47, 03B50, 03B53, 03B60.
Key words and phrases: self-disobedience, mixed logic, many-valued logic, semantic paradox, metainferences, transparent truth.

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Association for Symbolic Logic. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

doi:10.1017/S1755020326101282

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

to the effect that, contrary to first appearances, self-disobedience is not problematic [1, 3, 5, 37]. The question has ramifications well beyond the BA-plan, because the self-disobedience of the **ST**-hierarchy logics is obviously closely tied to their mixed nature, so the worry threatens mixed logics across the board.

The first aim of this paper is to clarify self-disobedience on a technical level. Each of [28, 36, 39] discusses a subtly different sense in which the **ST** logics are self-disobedient, and these senses are often conflated in the literature. We will distinguish six properties which have been or should be referred to as self-obedience, and map out how they relate to each other. Moreover, I noted above that the self-disobedience of the **ST** logics clearly has something to do with them being mixed, but what exactly? We will make the connection precise.

The second purpose is to show that the responses offered by the BA-plan are unsuccessful, and to reevaluate a more promising way of assuaging the self-disobedience concern, suggested in [28]. I'll argue that the most important BA-plan response stems from a confusion, namely a failure to distinguish between claims made *about* and *by* a logical theory. Moreover, we will vindicate [28]'s assessment that the more promising line of defense is unavailable to the most important logic in the **ST**-hierarchy. This vindication involves rejecting a story more recently told in [3, 5, 27] about the inferential norms prescribed by the system. The story in question is argued to suffer from essentially the same confusion mentioned above.

The layout is as follows. §2 gives a brief primer on mixed logics and the BA-plan. §3 studies the different notions of self-obedience, and their relationship to mixed-ness. §4 critically evaluates various arguments to the effect that self-disobedience is acceptable. We conclude by assessing where this leaves the prospects of the BA-plan and of self-disobedient logics more generally. The Appendix contains some definitions and proofs left out of the main text.

§2. Mixed logic. When defining logical consequence model-theoretically, we usually stipulate that an inference is valid iff in all models where all of the premises are satisfied, at least one of the conclusions is satisfied too. Within this template, there is room for variation from logic to logic: which language we are defining validity for, what the class of models looks like, and when a model satisfies a sentence. But one common feature is that there is a single notion of satisfaction-in-a-model, which occurs twice in the template: once in relation to the premises, and again for conclusions. Mixed logics are what we get when we drop the assumption that the same satisfaction condition fulfills these two roles, and instead say that an inference is valid just in case every model which satisfies$_1$ all the premises, also satisfies$_2$ one of the conclusions. But satisfaction$_1$ and satisfaction$_2$ may be quite different.

Three-valued logic provides a notable example. Let $\mathcal{L}^{PL}$ be a propositional language, formed by closing a set of propositional atoms under negation, conjunction, disjunction and implication.[1] A *Strong Kleene valuation* is a function from the sentences of $\mathcal{L}^{PL}$ to $\{0, \frac{1}{2}, 1\}$ which respects the Strong Kleene truth tables (Figure 1). Suppose we call an inference valid just in case every Strong Kleene valuation which assigns all the premises 1 assigns one of the conclusions *at least* $\frac{1}{2}$. This is known as strict-tolerant logic, or ST

[1] For simplicity, the logics used as examples in this paper will only be defined over $\mathcal{L}^{PL}$, although some of their applications rely on first-order extensions.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

	$\neg$
1	0
$\frac{1}{2}$	$\frac{1}{2}$
0	1

$\wedge$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
0	0	0	0

$\vee$	1	$\frac{1}{2}$	0
1	1	1	1
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	0

$\supset$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	1	1

Figure 1. The Strong Kleene truth tables.

[8, 9, 33, 34]. Although the mixed definition looks odd, ST has something going for it. Ripley [33] proved that ST can be conservatively extended with a fully transparent truth-predicate, remarkably—and much unlike other three-valued approaches to truth [14, 30]—without giving up any classically valid inferences.

As usual, there is a catch. Assuming some self-referential machinery, a Liar sentence λ can be formulated in the extended language, and it must take value $\frac{1}{2}$ in every model to avoid contradiction. This means that λ is a theorem of ST, since it always has value at least $\frac{1}{2}$. At the same time, the inference from λ to any atom p is valid, because λ never has value 1. But of course, atoms are not logical theorems. So here we have our first example of a logic seemingly disobeying its own logical laws: λ is a theorem, and the inference from λ to p is ruled valid, yet p is not a theorem. As we'll see, this is but one instance of a much broader phenomenon.

Since we want to say some things about mixed logics in general, it will be useful to have language- and model-independent definitions. We only assume that we have some formal language $\mathcal{L}$, with $\mathcal{L}_0$ its set of well-formed sentences, and some class of models $\mathcal{M}$. In this abstract setting, a satisfaction standard for sentences is a relation $X \subseteq \mathcal{M} \times \mathcal{L}_0$. We will call such relations *0-standards* for short. When $\langle M, \varphi \rangle \in X$, we say that M *X*-satisfies φ, and write $M \Vdash_X \varphi$. If Γ is a set of sentences, we write $M \Vdash_X \Gamma$ to mean that $M \Vdash_X \gamma$ for every $\gamma \in \Gamma$. Similar abbreviations will be used without further comment below. If X and Y are both 0-standards for $\mathcal{M}$ and $\mathcal{L}$, then X/Y denotes a *1-standard* for $\mathcal{M}$ and $\mathcal{L}$. If Γ and Δ are sets of $\mathcal{L}$-sentences, the inference from Γ to Δ is *X/Y-valid* when: $\forall M \in \mathcal{M}$, if $M \Vdash_X \Gamma$, then also $M \Vdash_Y \delta$ for some $\delta \in \Delta$. For instance, take $\mathcal{L}^{PL}$ and the set of Boolean valuations, and define a 0-standard C so that $v \Vdash_C \varphi$ iff $v(\varphi) = 1$. Then the validities of classical propositional logic are just the validities of the 1-standard C/C. For Strong Kleene valuations, we can define two 0-standards S and T, such that $v \Vdash_S \varphi$ iff $v(\varphi) = 1$, while $v \Vdash_T \varphi$ iff $v(\varphi) \geq \frac{1}{2}$. Then Strong Kleene logic K_3 corresponds to S/S, the logic of paradox LP to T/T, and our mixed logic ST to S/T.

So much for plain inferences. We are interested not only in which sentences follow logically from which sentences, but also in which inferences follow from which inferences, which inferences between inferences follow from which inferences between inferences, and so forth. Such *metainferences* play important roles in inferential practice and logical theorizing, especially in the BA-plan. Starting with sentences, now thought of as 0-inferences, we define a hierarchy of inferential levels inductively: for any $n \in \mathbb{N}$, an $n+1$-*inference* of $\mathcal{L}$ is an expression of the form $\Gamma \Rightarrow_{n+1} \Delta$, where Γ and Δ are sets of n-inferences of $\mathcal{L}$.[2] Thus regular inferences are 1-inferences, inferences between those are 2-inferences, and so on. All of them are called inferences.

[2] Lowercase Greek letters like φ and γ will represent not only sentences, but individual n-inferences of any n. Likewise, uppercase Greek letters such as Γ and Δ represent sets

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

Naturally, we can mix premise and conclusion standards also for higher levels. We'll see below that, by doing so, we can enhance some of the desirable properties of ST. So we generalize the notion of a standard upwards. If X and Y are n-standards for $\mathcal{M}$ and $\mathcal{L}$, then X/Y is an $n+1$-standard for $\mathcal{M}$ and $\mathcal{L}$. Of course, $\Gamma \Rightarrow_{n+1} \Delta$ is X/Y-satisfied in M (written $M \Vdash_{X/Y} \Gamma \Rightarrow_{n+1} \Delta$) when, if $M \Vdash_X \Gamma$, then also $M \Vdash_Y \delta$ for some $\delta \in \Delta$. Finally, $\Gamma \Rightarrow_{n+1} \Delta$ is X/Y-valid (written $\vDash_{X/Y} \Gamma \Rightarrow_{n+1} \Delta$) if it is X/Y-satisfied in all models.

This is the so-called local definition of validity for metainferences. It contrasts with the alternative, global validity, which requires preservation of validity rather than of satisfaction. As I argue in [21], global validity is *at least* as suitable as local, if not more so. Nonetheless, we only study local validity in the current paper, because it is the notion at play in the BA-plan.

Either way, a given n-standard is only designed to tell us which inferences *of level n* are valid. But we would like logics to decide on the validity of all inferences. Therefore, from now on, by a *logic* for $\mathcal{L}$ and $\mathcal{M}$ we mean a set of standards for $\mathcal{L}$ and $\mathcal{M}$, containing precisely one n-standard for every $n \in \mathbb{N}$.[3] Equivalently, a logic can be thought of as an ω-sequence such that the nth element is an n-standard. Naturally, an n-inference is satisfied/valid according to a logic $\mathbf{L}$ just in case it is so according to the unique n-standard in $\mathbf{L}$. Call a logic *pure* if it is of the form $\{X,\ X/X,\ X/X/X/X,\ ...\}$ for some 0-standard X. Otherwise, it is *mixed*. Self-disobedience gives us some reason to worry that perhaps only pure logics are legitimate. In the sequel, we investigate under which circumstances mixed-ness produces self-disobedience (§3), and to what extent self-disobedience in turn renders a logic untenable (§4).

For now, it remains to introduce some important example logics. We may assume that the classical logician uses her own standards at every metainferential level, for premises as well as conclusions. Thus full classical propositional logic is simply the set $\mathbf{CL} = \{C,\ C/C,\ C/C/C/C,\ ...\}$. Clearly, $\mathbf{CL}$ is pure. As for mixed logics: the $\mathbf{ST}$-hierarchy consists of an infinite sequence of logics, one dubbed $\mathbf{ST}^n$ for each $n > 0$, plus a final logic $\mathbf{ST}^\omega$. The exact definitions are left for the Appendix, since they aren't needed for what's to come. We only need to know that (i) the $\mathbf{ST}$ logics are all based on Strong Kleene valuations, (ii) each is a mix of S's and T's, and (iii) each starts out with the standards T and S/T for 0- and 1-inferences, respectively.

The $\mathbf{ST}$-hierarchy has garnered a lot of attention, for the following reasons. Barrio et al. [4] showed that, for each $n > 0$, the logic $\mathbf{ST}^n$ has precisely the same validities as $\mathbf{CL}$ at all inferential levels $m \leq n$, but is weaker than $\mathbf{CL}$ at every level $m > n$. So each next logic in the sequence is a bit stronger and a bit more classical than those that came before. $\mathbf{ST}^\omega$, finally, is classical all the way: it validates precisely the same n-inferences as $\mathbf{CL}$, for all $n \in \mathbb{N}$. Moreover, Pailos [26] proved that each $\mathbf{ST}^n$, as well as $\mathbf{ST}^\omega$, can be safely extended with transparent truth. Let us call the resulting logics $\mathbf{ST}^n_+$ and $\mathbf{ST}^\omega_+$.[4]

of n-inferences. We apply the standard abbreviations, so, for example, $\Gamma, \varphi \Rightarrow_n \psi$ denotes $\Gamma \cup \{\varphi\} \Rightarrow_n \{\psi\}$.

[3] Things like LP, K_3, or ST are of course called 'logics' in the wider logical literature, which focuses chiefly on sentences and 1-inferences. But they aren't full logics in our terminology. We can think of them as partial logics. Moreover, while they strictly speaking consist only of a 1-standard, we also assume that their 0-standard is the conclusion component of their 1-standard. So T is the 0-standard of LP and ST, and S that of K_3.

[4] These are technically no longer logics, but theories based on logics. However, since nothing in the remainder depends on this distinction, I will generally just refer to them as logics.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

They are a series of fully truth-transparent, yet increasingly classical systems, topped off by a completely classical one. For this reason Barrio, Pailos, and Toranzo Calderón [3] and Pailos and Da Ré [27] among others view $\mathbf{ST}^{\omega}_{+}$ as the ideal logic of truth. This is the paradise promised by the BA-plan.[5] Self-disobedience threatens to expel us from it (other attacks have been launched [12, 17, 19–21, 31], but they lie beyond the present scope). It is time to get clear on precisely what self-disobedience is, and to demarcate the affected territory.

§3. Self-disobedience in mixed logics. A few subtly different notions of self-obedience have appeared in the literature. Each is a measure for whether the behaviour of a logic is consonant with itself across adjacent inferential levels. Specifically, self-obedience properties check whether we could actually make the inferences declared valid by a logic, without ever passing from premises which are correct by the logic's own lights, to conclusions which aren't. This section elucidates the different self-obedience properties and their relations to each other and to mixed-ness.

The first notion of self-obedience was formulated by Scambler [39] (though he referred to it as logics being 'closed under their own laws'; the term 'self-obedient' originates with [36]). I'll call it (SM), for *Scambler Multiple-conclusion*. A logic has (SM) at a level $n \geq 1$ just in case every valid n-inference with valid premises also has at least one valid conclusion. That is, $\mathbf{L}$ has (SM) at n just in case, for all n-inferences $\Gamma \Rightarrow_n \Delta$:

$$\text{If } \vDash_{\mathbf{L}} \Gamma \Rightarrow_n \Delta \text{ and } \vDash_{\mathbf{L}} \Gamma, \text{ then } \exists \delta \in \Delta \text{ s.t. } \vDash_{\mathbf{L}} \delta. \tag{SM}$$

Moreover, if $\mathbf{L}$ has (SM) at every $n \geq 1$, we'll simply say that it has (SM) (likewise for the other criteria to be introduced). As Scambler points out, none of the $\mathbf{ST}^{n}_{+}$, nor $\mathbf{ST}^{\omega}_{+}$, has (SM). For instance, all of them start with T and S/T. Therefore they validate λ, and $\lambda \Rightarrow_1 p$ for any atom p, but not p.[6]

Ripley [36] later argued that (SM) is too strong—for reasons we'll see in a moment—and proposed a weaker alternative. Let's call it (SS), for *Scambler Single-conclusion*, as it is exactly Scambler's criterion limited to inferences with at most one conclusion.[7] That is, $\mathbf{L}$ has (SS) at n just in case, for all n-inferences $\Gamma \Rightarrow_n \varphi$:

$$\text{If } \vDash_{\mathbf{L}} \Gamma \Rightarrow_n \varphi \text{ and } \vDash_{\mathbf{L}} \Gamma, \text{ then } \vDash_{\mathbf{L}} \varphi. \tag{SS}$$

[5] To be precise, by the BA-plan I mean any approach to semantic paradox based on one of the logics in the **ST**-hierarchy proper, where 'proper' means: excluding $\mathbf{ST}^{1}$, which is merely a metainferential extension of ST. In practice, the only **ST**-hierarchy proper logic used for such purposes has been $\mathbf{ST}^{\omega}$, and any other choice would seem highly arbitrary.

[6] Readers acquainted with ST may think we're just talking about failures of *Cut* or one of its variants here, but that isn't so. It's one thing to say, for example, that if $\vDash_{\mathbf{L}} \lambda$ and $\vDash_{\mathbf{L}} \lambda \Rightarrow_1 p$, also $\vDash_{\mathbf{L}} p$, and quite another to say that $\vDash_{\mathbf{L}} \{\emptyset \Rightarrow_1 \lambda, \lambda \Rightarrow_1 p\} \Rightarrow_2 (\emptyset \Rightarrow_1 p)$. Such things are closely related in pure logics, but can come far apart when mixing. All the self-obedience properties are completely independent from the validity of any type of (meta$_n$-)*Cut*. Case in point: $\mathbf{ST}^{\omega}_{+}$ lacks every self-obedience property, yet validates every form of *Cut*, at every level. Conversely, the completely empty logic has every property—vacuously, as it has no validities to obey—while invalidating every kind of *Cut*.

[7] Nonetheless, (SS) applies to multiple-conclusion logics, such as we (and Ripley) are interested in; it just ignores their multiple-conclusion inferences. The same goes for (PS) below.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

Of course, the example we gave of $\mathbf{ST}^n_+$ and $\mathbf{ST}^\omega_+$ failing (SM) was already single-conclusion, so it equally works as a counterexample to (SS).

Other notions of self-obedience can be gleaned from Porter's [28]. We've just learned that in the **ST**-hierarchy logics, the *set of validities* at a given level doesn't always obey the valid inferences of the level above. Porter points out that the same is true of *theories*. The **L**-theory of a set of inferences Γ, denoted $Th_{\mathbf{L}}(\Gamma)$, is the set of inferences which are **L**-satisfied in every model which **L**-satisfies Γ. The idea being that $Th_{\mathbf{L}}(\Gamma)$ contains everything which follows from the assumption that we are in a model of Γ. Porter asks whether reasoning according to the validities of **L**, starting from some Γ, won't ever take us beyond $Th_{\mathbf{L}}(\Gamma)$. As before, we could require this in general, or restrict to validities with at most one conclusion (Porter himself doesn't specify). So let's define two properties, (PM) and (PS), following our previous naming convention. **L** has (PM) at n just in case, for all n-inferences $\Gamma \Rightarrow_n \Delta$:

$$\text{If } \models_{\mathbf{L}} \Gamma \Rightarrow_n \Delta, \text{ then } Th_{\mathbf{L}}(\Gamma) \cap \Delta \neq \emptyset. \tag{PM}$$

Unpacking this a bit will help comparison with the other notions. So, equivalently:

$$\text{If } \models_{\mathbf{L}} \Gamma \Rightarrow_n \Delta, \text{ then } \exists \delta \in \Delta \text{ s.t. } \forall M \in \mathcal{M}, \text{ if } M \Vdash_{\mathbf{L}} \Gamma, \text{ then } M \Vdash_{\mathbf{L}} \delta. \tag{PM}$$

Restricting these conditions to single-conclusion inferences gives us, respectively:

$$\text{If } \models_{\mathbf{L}} \Gamma \Rightarrow_n \varphi, \text{ then } \varphi \in Th_{\mathbf{L}}(\Gamma). \tag{PS}$$

And equivalently:

$$\text{If } \models_{\mathbf{L}} \Gamma \Rightarrow_n \varphi, \text{ then } \forall M \in \mathcal{M}, \text{ if } M \Vdash_{\mathbf{L}} \Gamma, \text{ then } M \Vdash_{\mathbf{L}} \varphi. \tag{PS}$$

Note that whereas Scambler-style properties record whether **L**-validities preserve **L**-validity, the Porter variants check that **L**-validities preserve **L**-satisfaction. Porter observes that none of the $\mathbf{ST}^n$ nor $\mathbf{ST}^\omega$ have (PS) nor (PM). For instance, recall that all the **ST** logics start with T and S/T. Then note that $p, \neg p \Rightarrow_1 q$ is S/T-valid (as no model gives both p and $\neg p$ value 1), yet there are models which T-satisfy both p and $\neg p$ (by giving them both value $\frac{1}{2}$) without T-satisfying q.

Now why did Ripley deem (SM) too strong? The reason most relevant to our purposes is that even completely pure logics like **CL** don't generally have (SM). For example, $\emptyset \Rightarrow_1 p, \neg p$ is classically valid, even though neither p nor $\neg p$ is. I take this example to illustrate that, besides self-obedience, having (SM) also requires a type of generalized disjunction property, to the effect that whenever a set of n-inferences is collectively valid (every model satisfies at least one of them), one among them is valid on its own. Perhaps constructivists would be happy to impose such a requirement on logics. But even so, that is a question independent from self-obedience, so we better keep the two requirements separate, rather than packaging one in with the other. By contrast, it is easy to see that every pure logic has (SS).

What about (PM)? Spelling out and comparing the definitions makes it obvious that every logic which has (PM) also has (SM). It immediately follows that **CL** does not have (PM) either, and that (PM) also includes the generalized disjunction property.[8] On the

[8] Even constructivist logics won't typically have (PM). For example, intuitionistic logic doesn't, due to $p \vee q \Rightarrow_1 p, q$. Whether it has (SM) depends. Local validity is inherently model-theoretic [23, p. 258], so different intuitionistic semantics (Kripke, topological, algebraic, etc.) might produce different answers.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

other hand, it is again simple to verify that every pure logic has (PS). In the remainder, then, I will largely leave (SM) and (PM) aside as unsuitable notions of self-obedience.[9]

We may wonder, though, whether no plausible self-obedience property can take into account multiple-conclusion inferences at all. *Prima facie* there should be a sensible way to talk about validities being obeyed or not, even if they have multiple conclusions. I submit that indeed there is. Consider again the second definition of (PM). It is so strong because it requires that one and the same conclusion is satisfied in all models of the premises, rather than merely that each such model satisfies some potentially different conclusion. So I propose we flip the order of the quantifiers. Let's call the result (PD), for *Porter Distributed* (as it 'distributes' the burden over the different conclusions)[10]:

$$\text{If } \models_{\mathbf{L}} \Gamma \Rightarrow_n \Delta, \text{ then } \forall M \in \mathcal{M}, \text{ if } M \Vdash_{\mathbf{L}} \Gamma, \text{ then } \exists \delta \in \Delta \text{ s.t. } M \Vdash_{\mathbf{L}} \delta. \tag{PD}$$

An analogous move can be applied to (SM), after rewriting $\models_{\mathbf{L}} \delta$ as $\forall M \in \mathcal{M},\ M \Vdash_{\mathbf{L}} \delta$. Naturally, we call the result (SD):

$$\text{If } \models_{\mathbf{L}} \Gamma \Rightarrow_n \Delta \text{ and } \models_{\mathbf{L}} \Gamma, \text{ then } \forall M \in \mathcal{M},\ \exists \delta \in \Delta \text{ s.t. } M \Vdash_{\mathbf{L}} \delta. \tag{SD}$$

It is once more easy to check that every pure logic has (SD) and (PD). I regard them as the proper multiple-conclusion analogues of (SS) and (PS), respectively.

Altogether we have six notions of self-obedience. Figure 2 is the Hasse diagram of their ordering by strength. The Scambler- and Porter-style properties line the top left and bottom right, respectively. Within each group, the naive multiple-conclusion variant strictly entails the distributed version (because of quantifier scope), which strictly entails the single-conclusion one (as a special case). Moreover, every Porter-style criterion strictly entails the corresponding Scambler variant (because preserving satisfaction entails preserving local validity). All of the depicted entailments are immediate upon unpacking and comparing the definitions. Counterexamples to the non-entailments are left to the interested reader.

What about our example logics? We've already mentioned that **CL**, like all pure logics, has (PD), and hence all four of the properties forming the top right diamond in Figure 2. But like most logics, pure or mixed, it has neither (SM) nor (PM). We've also seen that $\mathbf{ST}^n_+$ and $\mathbf{ST}^\omega_+$ lack (SS), which means they lack all six. Lastly, while $\mathbf{ST}^n$ and $\mathbf{ST}^\omega$ themselves don't have (PS) nor hence (PD), they actually do have (SS) and (SD), as is proven in the Appendix.

So if we only care about which properties are met *fully*, it doesn't matter how high up in the **ST**-hierarchy we are—$\mathbf{ST}^1$ fares the same as $\mathbf{ST}^{500}$ and $\mathbf{ST}^\omega$. Yet it is important to also check the properties *per level*. The self-disobedience of the different **ST** logics is spread across the levels to very different degrees. Namely: for each $n \geq 1$, the corresponding logic $\mathbf{ST}^n_+$ fails (SS) and (SD) only at the levels $m \leq n$, but has them at $m > n$. $\mathbf{ST}^\omega_+$ fails these criteria at *all* levels. Similarly, $\mathbf{ST}^n$ and $\mathbf{ST}^n_+$ fail (PS) and (PD) at all the levels $m \leq n$, but have them at every $m > n$, while $\mathbf{ST}^\omega$ and $\mathbf{ST}^\omega_+$ don't have

[9] An anonymous referee asks whether we should really dismiss (SM) completely, if what we are interested in is logics being closed under their own validities, and (SM) seems to express something along these lines. The answer is that, contrary to appearances, (SM) is *not* a closure principle, as Ripley [36, p. 1254] has shown.

[10] This move requires the unpacked formulation of (PM). (PD) cannot be phrased in terms of theories.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

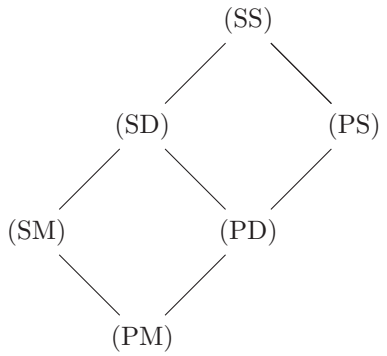


Figure 2. Self-obedience properties.

them at any level. So the further up the **ST**-hierarchy we climb, the higher the inferential level to which self-disobedience has spread, until finally we are self-disobedient all the way.[11] This observation will become crucial in §4.

One key issue for the current section remains: how does self-disobedience relate to mixed-ness? All pure logics have (PD), so mixed-ness is at least necessary for self-disobedience. But it is not sufficient, as we'll see below. This raises the question: can we characterize, in terms of the structure of the sequence of standards, precisely which mixed logics are and aren't self-obedient in each sense? It is unlikely that any elegant or informative characterization applies across the board. The reason is that we are working with an abstract and highly inclusive definition of 'standard'. Because of this, the space of all logics includes many degenerate cases, and we can say relatively little with full generality. It will be more insightful to temporarily restrict our attention to logics based on a (relevant) class of well-behaved standards.

To this end, we introduce some terminology. A logic is *compositional* if its models are valuations assigning a range of semantic values to the sentences, such that the value of any complex sentence is determined by the values of its components (see the Appendix for the formal definition). **CL** and the **ST** logics are obviously compositional.[12] Moreover, a logic is *constant complete* if the language contains a constant for each semantic value (see again the Appendix). Constants need not be logical primitives; **CL** is constant complete, though we haven't included $\bot$ and $\top$ in $\mathcal{L}^{PL}$, because $p \wedge \neg p$ and $p \vee \neg p$ suffice. The Strong Kleene tables don't produce constants automatically, but we could simply add logical constants to make the **ST** logics constant complete. The family of logics we're going to focus on is that of the compositional and constant complete ones.

Let's also agree on some notation that will come in handy throughout the remainder. If **L** is a logic, we write $\mathbf{L}(n)$ for the unique n-standard in **L**. Moreover, if X is an n-standard, $\uparrow X$ denotes the $n+1$-standard X/X. Also, if X/Y is an n-standard, $\swarrow(X/Y)$ denotes the $n-1$-standard X, and $\searrow(X/Y)$ the $n-1$-standard Y.[13] Observe

[11] The proof regarding (SS) is explained in [36, p. 1256], and the others are analogous.

[12] Though whether first-order extensions are still strictly speaking compositional depends on the semantics assigned to the quantifiers [41].

[13] The arrows bind stronger than the slice, so, for example, $\uparrow X/Y$ is read as $(\uparrow X)/Y$, not $\uparrow(X/Y)$.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

that in this notation, $\mathbf{L}$ is pure iff $\mathbf{L}(n+1) = \uparrow\mathbf{L}(n)$ for all $n \in \mathbb{N}$ iff $\swarrow\mathbf{L}(n+1) = \searrow\mathbf{L}(n+1) = \mathbf{L}(n)$ for all $n \in \mathbb{N}$. Finally, we define a relation on standards. If X and Y are both n-standards for $\mathcal{M}$ and $\mathcal{L}$, we write $X \leq Y$ when, for every M and φ, if $M \Vdash_X \varphi$ then $M \Vdash_Y \varphi$. Our characterization of self-obedience will centre on the following property.

DEFINITION 3.1 (Quasi-purity). $\mathbf{L}$ *is* quasi-pure *iff* $\mathbf{L}(n) \leq \swarrow\mathbf{L}(n+1)$ *and* $\mathbf{L}(n) \geq \searrow\mathbf{L}(n+1)$ *for all* $n \in \mathbb{N}$.

The term 'quasi-pure' indicates that if we replace both occurrences of '$\leq$' in the definition with '$=$', we obtain a definition of purity.

It turns out that, for logics in our class of interest, the four plausible self-obedience criteria collapse into one, so we can speak of self-obedience *simpliciter*. Moreover, exactly what is needed for self-obedience in such logics is not purity *per se*, but quasi-purity. This fact is proven in the Appendix.

THEOREM 3.2. *Let* $\mathbf{L}$ *be a compositional and constant complete logic. The following are equivalent*:

- $\mathbf{L}$ *is quasi-pure.*
- $\mathbf{L}$ *has* (*PD*).
- $\mathbf{L}$ *has* (*PS*).
- $\mathbf{L}$ *has* (*SD*).
- $\mathbf{L}$ *has* (*SS*).

So, for instance, the fact that logics like $\mathbf{ST}^n$ and $\mathbf{ST}^\omega$ have (SD) and (SS) without having (PD) and (PS) is only possible because we didn't include constants in $\mathcal{L}^{PL}$. $\mathbf{ST}^n$ and $\mathbf{ST}^\omega$ are not quasi-pure, so once we add constants, they lose (SD) and (SS) too. For another quick application: the so-called $\mathbf{TS}$-hierarchy is a dual of the $\mathbf{ST}$-hierarchy, obtained by swapping every S for a T and vice-versa [39]. It is easy to check that the $\mathbf{TS}$-hierarchy logics are all quasi-pure, therefore Theorem 3.2 guarantees self-obedience when constants are present. Note that this is despite those logics being thoroughly mixed, so indeed purity is not necessary for self-obedience.

To close the section, let's briefly look at the narrow gap between quasi-purity and purity proper. Call $\mathbf{L}$ *reflexive* if $\vDash_{\mathbf{L}} \varphi \Rightarrow_{n+1} \varphi$ for every n-inference φ of every $n \in \mathbb{N}$.

THEOREM 3.3. *If a logic is both quasi-pure and mixed, it is not reflexive.*

The proof is in the Appendix. Combining with Theorem 3.2 produces the following.

THEOREM 3.4. *Let* $\mathbf{L}$ *be a reflexive, compositional and constant complete logic. If* $\mathbf{L}$ *is mixed, then* $\mathbf{L}$ *is self-disobedient.*

Proof. If $\mathbf{L}$ is both reflexive and mixed, by Theorem 3.3 it is not quasi-pure. Therefore, since $\mathbf{L}$ is also compositional and constant complete, it follows by Theorem 3.2 that $\mathbf{L}$ is not self-obedient. □

Reflexivity, compositionality, and constant completeness are natural and common properties. Hence in many important cases, mixed-ness will immediately entail self-disobedience by Theorem 3.4. This shows that self-disobedience, though not universal among mixed logics, is at least far from peculiar to the $\mathbf{ST}$-hierarchy. Next, we examine whether this is problematic for the logics affected.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

§4. Is self-disobedience acceptable? To endorse a self-disobedient logic is to put oneself in the place of our interlocutor from the introduction, who refused to grant A while acknowledging that it follows from premises which they maintain.[14] More precisely, to endorse a logic which lacks (SS) is to consider something invalid while acknowledging that it follows logically from things which one considers valid. *Mutatis mutandis* for (SD), (PS), and (PD).

On the face of it, such positions appear absurd. Scambler voiced this impression when discussing *modus ponens*, in particular the instance $\lambda, \lambda \supset \bot \Rightarrow_1 \bot$, which is another example of an inference that is validated but not obeyed by the **ST**-logics.

> To say that one accepts modus ponens, also accepts λ, $\lambda \supset \bot$, but doesn't accept $\bot$, seems at first pass to border on the incoherent. [...] I'm tempted to think that anyone who really accepts λ, $\lambda \supset \bot$ and modus ponens should accept $\bot$ *by definition*, though I understand that proponents of [$\mathbf{ST}_1^+$] won't be moved by this thought. [39, p. 366]

The qualification at the end of this quote is prudent. As David Lewis [24] famously noted, incredulous stares are not exactly conclusive objections. Yet they are not dialectically irrelevant either. Lewis [25] interpreted the incredulous stares directed at his modal realism as expressions of the view that modal realism defies common sense to such an extent that it outweighs the theory's advantages. Lewis for his part recognized the steep price, but, contrary to his critics, judged it worth paying. Moreover, he considered this disagreement to be at an impasse.

> I acknowledge that my denial of common sense opinion is severe, and I think it is entirely right and proper to count that as a serious cost. How serious is serious enough to be decisive? - That is our central question, yet I don't see how anything can be said about it. [25, p. 135]

Likewise, if proponents of some self-disobedient logic were to simply insist that self-disobedience is a price worth paying in return for the virtues of their logic, I am not sure what we could say to prove them wrong. To my mind, self-disobedience strains credulity past the breaking point, hence I would remain unwilling to accept it regardless of the gains that might await, but I don't quite see how this view can be pressed.

Instead, the purpose of this section is as follows. Whereas Lewis admitted to the cost of his theory and affirmed a willingness to pay, BA-planners have maintained that self-disobedience is no substantial cost in the first place. They have put forth three arguments to the effect that, upon closer consideration, self-obedience isn't as obvious or desirable as it initially looks, and its failure is perfectly fine. This is the part of their story I want to actively dispute. In my view, each of those three arguments gets things wrong in one way or another, and self-disobedience is in fact just as bizarre as it looks at first glance.

In the remainder, then, I will go over the three arguments in turn and explain where I take issue with each. Finally, we will also explore a fourth argument for the

[14] By endorsing a logic, I mean endorsing it for a purpose along the lines of what Priest [29] calls the canonical application of logic, that is, as a model of (idealized) reasoning. Logics may be applied to all sorts of other things, a standard example being the use of classical propositional logic in modeling electrical circuits, but we leave such alternative applications aside.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

acceptability of self-disobedience, but conclude that—insofar as it can work at all—it cannot be applied to $\mathbf{ST}^{\omega}$ and $\mathbf{ST}^{\omega}_{+}$. For concreteness, I will focus chiefly on (SS), which is both the simplest and weakest self-obedience property.

4.1. Argument 1: Mixing up local and global. The first argument we'll consider is suggested by some comments of Barrio et al. [3, fn. 26] and Roffé and Pailos [37, p. 749]. It alleges that a demand for (SS) mixes up local and global validity. Recall that a metainference is locally valid when it preserves *satisfaction* from premises to conclusion, and globally valid when it preserves *validity*. We have been using local validity throughout, since it is part of the BA-plan. But by expecting (SS) to hold, we're expecting that the **L**-validities—meaning therefore the *local* **L**-validities—preserve **L**-validity. And so the idea would be that such expectations suffer from a discrepancy: the assumption in the antecedent is one of local validity, but the demands of the consequent are about preservation of validity, the hallmark of global validity.

The easiest way to see that this argument doesn't help is by noticing that it only applies to the Scambler-style criteria, not the Porter-style ones. (PS) requires precisely that **L**-validities preserve **L**-*satisfaction*. Preservation of satisfaction is the hallmark of local validity, and so there is no discrepancy at all in this case. In fact, (PS) is precisely what we end up with if we "fix" the supposed discrepancy in (SS). Now recall (Figure 2) that (PS) strictly *entails* (SS). This is to say that, when attempting to fix the perceived problem in the requirement of (SS), one only ends up making the demand *stronger*. Such a move obviously cannot rescue a logic which already lacked (SS). Furthermore, it follows that the discrepancy in (SS) was a benevolent one to begin with; the demands of the consequent are *modest* relative to the assumptions. (This also explains why every pure logic has (SS), despite its discrepancy.) Thus there is no problem with expecting (SS) itself. Indeed there cannot be, since again, that expectation is included in (PS), which plainly doesn't suffer from the alleged problem. The same argument applies to (SD) and its relation to (PD).

4.2. Argument 2: The deduction theorem analogy. The second type of argument is an analogy drawn by Barrio and Bezerra [1, p. 18]. The deduction theorem for the material conditional '$\supset$' in (say) classical logic states that any 1-inference $\Gamma, \varphi \Rightarrow_1 \psi$ is valid if and only if $\Gamma \Rightarrow_1 \varphi \supset \psi$ is valid. In both LP and K_3, this equivalence breaks down. One of the consequences is that, for example, LP doesn't generally validate *modus ponens*: $\varphi, \varphi \supset \psi \Rightarrow_1 \psi$. Or, to make the analogy fully explicit, that LP does not meet the following condition for all sentences φ and ψ:

$$\text{If } \models \varphi \supset \psi, \text{ then } \forall M \in \mathcal{M}, \text{ if } M \Vdash \varphi, \text{ then } M \Vdash \psi. \tag{$\mathrm{PS}_{\supseteq}$}$$

Similarly, in LP extended with constants or transparent truth, we don't always have that:

$$\text{If } \models \varphi \supset \psi \text{ and } \models \varphi, \text{then } \models \psi. \tag{$\mathrm{SS}_{\supseteq}$}$$

Of course, these properties are essentially (PS) and (SS), respectively, except they concern the object language connective '$\supset$' in place of the inferential arrow '$\Rightarrow_n$' of some n.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

Thus the analogy argument goes as follows. The first premise has it that (i) the failure of the deduction theorem in LP and K_3 is acceptable. And the second that (ii) failures of self-obedience are analogous to the failure of the deduction theorem in LP and K_3. From (i) and (ii), it is then concluded that (iii) failures of self-obedience are acceptable as well.

I think there are multiple reasons to doubt this argument, starting with premise (i). The material conditional of LP is typically dismissed, even by *supporters* of LP, as untenable *qua* conditional, and for reasons essentially tied to the failure of the deduction theorem. For instance, here is LP's most outspoken advocate, Priest:

> [D]ialetheism [the stance underpinning LP] disposes, once and for all, of "material implication". The reason is simple. Any conditional worth its salt, $\rightarrow$, should satisfy the *modus ponens* principle: $\{\alpha, \alpha \rightarrow \beta\} \vDash \beta$. This is, indeed, analytically part of what implication is. Yet this principle fails for material implication [in LP.] Hence material implication is not the conditional. [30, p. 83]

Much work on LP is devoted precisely to the study of alternative conditionals, which may be added to the language in order to take over the roles that we want a proper conditional to play (see [6] and the references therein).

Premise (i) may be more plausible in the case of K_3. Yet this is only because K_3's failure of the deduction theorem is very different from LP's. The deduction theorem requires a general equivalence between $\vDash \Gamma, \varphi \Rightarrow_1 \psi$ and $\vDash \Gamma \Rightarrow_1 \varphi \supset \psi$. LP delivers the left-to-right direction but not the right-to-left one, while for K_3 it is exactly the other way around. One consequence is that unlike LP, K_3 does validate *modus ponens* for the material conditional. Also unlike LP, K_3 satisfies (PS$\supset$) and (SS$\supset$).[15] But then K_3's failure of the deduction theorem is nothing like self-disobedience. The upshot is that the analogy argument is really two separate arguments: one about LP and one about K_3. In the former, premise (i) fails, and in the latter, premise (ii) does.

Lastly, even if premises (i) and (ii) both held up, this wouldn't lend support to conclusion (iii). When two problems are "analogous," one of them might still be far more serious than the other, for example, by affecting a more critical part of the system. A fractured toe is analogous to a fractured spine. Likewise, a given issue might be tolerable when merely affecting some connective definable in the object language, yet disastrous when affecting the inferential arrow '$\Rightarrow_n$' itself—the logic's very notion of validity.

Concretely, for one, the difference means that while LP and K_3 are often extended with alternative conditionals, so as to avoid using '$\supset$', the analogous maneuver is not available to advocates of self-disobedient logics. To adopt a different inferential arrow (for any level) would be to abandon ship and adopt a different logic altogether. If one merely modifies or extends the object language of a system, there is a meaningful sense in which we can still speak of the result as (another version of) the same logic. But the same cannot be said once one has altered the underlying notion of validity of some level by replacing one of the inferential arrows.

Summing up: the failure of the deduction theorem for '$\supset$' in LP is analogous to self-disobedience, and unacceptable, while that in K_3 is not analogous to start

[15] These facts are immediate by inspection of the truth tables in Figure 1.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

with. Moreover, even if some other object-language conditional turns out to fail the deduction theorem in a manner that's both acceptable and analogous to self-disobedience, this still shouldn't lead us to accept self-disobedience.[16]

4.3. Argument 3: Mixed obedience. The third argument is the one most emphasized by BA-planners. It is developed in [3, pp. 12579–12581] and [37, pp. 748–750], and appealed to by [1, p. 19]. The main claim is that, while we should demand some kind of self-obedience from logics, the specifics of (SS) and the like are only appropriate with respect to pure logics. The properties must be adapted before being demanded of mixed logics, in order not to distort the spirit of the original demand—or so the story goes. The resulting modified property, *mixed obedience* or *mixed closure*, is then shown to be satisfied by the **ST** logics.

The reason why requiring (SS) is supposedly inappropriate for mixed logics is as follows. Let us recall the definition of (SS), but this time, we replace each occurrence of '$\models_{\mathbf{L}}$' with a description of the specific standard at play in each case:

$$\text{If } \models_{\swarrow\mathbf{L}(n)/\searrow\mathbf{L}(n)} \Gamma \Rightarrow_n \varphi \text{ and } \models_{\mathbf{L}(n-1)} \Gamma, \text{ then } \models_{\mathbf{L}(n-1)} \varphi. \tag{SS}$$

This is just (SS), rewritten to be explicit about the fact that an inference of a given level is valid in a logic just in case it is valid in that logic's standard for the relevant level. So '$\models_{\mathbf{L}} \Gamma \Rightarrow_n \varphi$' was short for '$\models_{\mathbf{L}(n)} \Gamma \Rightarrow_n \varphi$', and $\mathbf{L}(n) = \swarrow\mathbf{L}(n)/\searrow\mathbf{L}(n)$, while '$\models_{\mathbf{L}} \Gamma$' and '$\models_{\mathbf{L}} \varphi$' were short for '$\models_{\mathbf{L}(n-1)} \Gamma$' and '$\models_{\mathbf{L}(n-1)} \varphi$', respectively.

The observation meant to motivate mixed obedience is this: when Γ first appears in (SS), the standard to which it ends up being measured is $\swarrow\mathbf{L}(n)$, the premise component of $\mathbf{L}(n)$. Yet the second appearance of Γ is measured according to $\mathbf{L}(n-1)$. Similarly, the first occurrence of φ is held to $\searrow\mathbf{L}(n)$, while the second occurrence is assessed with respect to $\mathbf{L}(n-1)$. Of course, if $\mathbf{L}$ is pure, then $\swarrow\mathbf{L}(n) = \mathbf{L}(n-1) = \searrow\mathbf{L}(n)$. But in mixed logics these equalities might not hold, in which case there appears to be a *mismatch* in (SS): we first assume something about the way Γ and φ relate to $\swarrow\mathbf{L}(n)$ and $\searrow\mathbf{L}(n)$, and from this we expect to draw conclusions about how they relate to $\mathbf{L}(n-1)$. It's because of this mismatch that [1, 3, 37] deem the original formulation unsuitable for mixed logics. As Roffé and Pailos put it:

> Scambler's Closure demands that if (i) if Γ is satisfied by a valuation according to $[\swarrow\mathbf{ST}^{\omega}_{+}(2)]$, then Δ is satisfied by the valuation in $[\searrow\mathbf{ST}^{\omega}_{+}(2)]$, and (ii) Γ is valid according to $[\mathbf{ST}^{\omega}_{+}(1)]$, then (iii) Δ is valid in $[\mathbf{ST}^{\omega}_{+}(1)]$. But there is no reason why (iii) should follow from (i) and (ii). [37, p. 749]

Thus the supposed improvement, mixed obedience, is obtained by correcting the mismatch.

> [F]or spirit of the critics' demand to be adequately preserved in a mixed metainferential setting [...] one should read the premises in the same way that one understands the premise part of $[\models_{\mathbf{L}} \Gamma \Rightarrow_n \varphi]$. [37, p. 749]

[16] A closely related analogy is used for different purposes in [2, p. 1163]. It is susceptible to the first and third of our criticisms, if not the second.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

In case of (SS) this would result in the following:

$$\text{If } \vDash_{\swarrow\mathbf{L}(n)/\searrow\mathbf{L}(n)} \Gamma \Rightarrow_n \varphi \text{ and } \vDash_{\swarrow\mathbf{L}(n)} \Gamma, \text{ then } \vDash_{\searrow\mathbf{L}(n)} \varphi. \qquad \text{(MSS)}$$

It is easy to see that every logic in the **ST**-hierarchy has this property, and so it is concluded that they are self-obedient after all. In fact, every logic whatsoever has (MSS), so it might be argued by the same token that all mixed logics are self-obedient.[17]

I take it this argument too is flawed. To understand why, it is essential to mind the distinction between the relativized claim that some inference is valid *in a given standard/logic*, and the absolute claim that it is *valid*, that is, valid *simpliciter*. A claim of the former kind is *about* a theory of logic, whereas the latter are the kinds of claims made *by* such a theory. Moreover, as Field has noted in a different context:

> someone who thinks that "disjunctive syllogism" [...] is not a valid form of inference will, if she accepts a bare minimum of mathematics, agree that the inference is *classically* valid, and will say that that just shows that classical validity outruns genuine *validity*. Those who accept disjunctive syllogism don't just believe it *classically* valid, which is beyond serious contention; they believe it *valid*. [...] Whatever logic L one advocates, one should recognize a distinction between the concept 'valid-in-L' and the concept 'valid'. [15, pp. 33–34]

The mixed obedience argument goes wrong by overlooking this distinction.

For suppose that the concern about self-disobedience was as follows. "Proponents of (say) $\mathbf{ST}^{\omega}_{+}$ are committed to the claim that $\lambda \Rightarrow_1 p$ is valid *in* $\mathbf{ST}^{\omega}_{+}$. Moreover, they are committed to the claim that λ is valid *in* $\mathbf{ST}^{\omega}_{+}$. Therefore, they should be committed to the claim that p is valid *in* $\mathbf{ST}^{\omega}_{+}$." Clearly, the conclusion of this objection doesn't follow from its premises. In general, this reasoning pattern doesn't go through for mixed logics, precisely because of the aforementioned mismatch. So if this had been the concern, then pointing out the mismatch and proposing a corrected pattern which accounts for mixed standards would have hit the nail on the head. But that mischaracterizes the concern. To see why, simply note that relativized validity claims, about what is and isn't valid *in* $\mathbf{ST}^{\omega}_{+}$, aren't up for debate in the first place. That $\lambda \Rightarrow_1 p$ and λ are valid in $\mathbf{ST}^{\omega}_{+}$, and p is not, are simple mathematical facts. There is no room for disagreement here.

Instead, there is disagreement about what is valid *simpliciter*, and it's there that the worry arises. Proponents of $\mathbf{ST}^{\omega}_{+}$ are not only committed to the relativized claims that $\lambda \Rightarrow_1 p$ and λ are valid *in* $\mathbf{ST}^{\omega}_{+}$ (as everyone is). By endorsing $\mathbf{ST}^{\omega}_{+}$, they (unlike their critics) are also committed to the absolute claims that $\lambda \Rightarrow_1 p$ and λ are valid. And from *those* commitments, it does seem to follow that they are committed to the absolute claim that p is valid.

[17] Here I've covered (SS), but the same mismatch observations, and analogous moves to mixed counterparts, could be applied to (PS), (SD), etc. While [3] concentrated on (SS), [1] doesn't specify, and [37] treats what seems to be (SM). Moreover, the BA-planners only discuss **ST** logics, whereas I've generalized the central point. In particular: they focus on premise-side mismatches (cf. the last quote from [37]), since the **ST** logics happen to be such that $\mathbf{L}(n-1) = \searrow\mathbf{L}(n)$ anyway. But presumably the same argument would apply to the conclusion side, if needed, hence my (MSS) adjusts both.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

The mismatch observation doesn't apply, because we are not talking about validity *in certain standards*. Invoking the mismatch observation would require first substituting out absolute validity claims for relativized ones. Such substitutions are made extensively and deliberately in [3, 37], with the idea that this merely makes things more explicit. But in fact, due to the distinction emphasized above, such substitutions don't preserve meaning. As a consequence, applying them here leads to a major misrepresentation of the point.

Note finally that, contrary to (SS), (MSS) can't express any requirement on absolute validity claims. Each logic **L** represents a body of absolute validity claims, namely to the effect that the validities *simpliciter* are precisely all and only the validities of **L**. But at each n, the n-inferences which **L** validates, and hence claims are valid *simpliciter*, are all and only those that are valid in $\mathbf{L}(n)$. Therefore (SS) gets things exactly right by only considering validity in $\mathbf{L}(n)$ and $\mathbf{L}(n-1)$, while (MSS) blunders by looking at $\searrow\mathbf{L}(n)$ and $\swarrow\mathbf{L}(n)$, which simply do not correspond to the absolute validity claims actually made by **L**.[18]

Mixed obedience was introduced with the aim of doing better justice to the idea underlying the demand for self-obedience. Unfortunately, it rests on a misunderstanding of said idea.

4.4. Argument 4: The bounds reading. The fourth and final argument to be discussed is a bit different from the others. For one, it hasn't been pursued by BA-planners, but was suggested by Porter [28]. Secondly, the key point of contention won't be so much whether the strategy works, but for which logics it is available.

We start with the bounds reading of validity. In brief, Restall [32] and Ripley [34] interpret validity for 1-inferences along the following lines: $\Gamma \Rightarrow_1 \Delta$ is valid just in case it is incoherent, or *out of bounds*, to simultaneously accept every sentence in Γ and reject every sentence in Δ. There's much to be said about such interpretations, but the important point for us is that they can potentially justify failures of self-obedience. Ripley maintains (contra Restall) that there are sentences, such as λ, which it is always out of bounds to assert or deny. In that case, $\lambda \Rightarrow_1 p$ is valid because it is out of bounds to assert λ, and $\emptyset \Rightarrow_1 \lambda$ is valid because it is out of bounds to deny λ, yet it might be in bounds to deny p and hence $\emptyset \Rightarrow_1 p$ need not be valid. This is essentially how Ripley [34] explains the self-disobedience of ST. Similarly, Porter [28] proposes that if the bounds reading is extended to higher levels, failures of self-obedience at those levels may follow as well. In fact, extended bounds interpretations have already been developed for the **ST**-hierarchy [3, 13], though [3] doesn't pick up Porter's suggestion of employing such an interpretation in response to the self-disobedience concern, appealing instead to mixed obedience, as we saw. Either way, insofar as the bounds interpretation holds up for a given logic, the concern might be assuaged.

However, Porter adds an important qualification. When the bounds reading is applied to a given level, the validities of that level aren't interpreted as genuine inferences

[18] An anonymous referee objects that one of the lessons from recent work surrounding the **ST**-hierarchy seems to be that logics cannot be uniquely identified purely by their validities. However, we should distinguish how logics are *identified* from how they are *evaluated*. So I agree that logics cannot be uniquely identified solely by their validities (see [16, 22] for more on this), and that we instead should characterize them by their pairs of standards at each level (as we have done throughout). Yet I reckon we should evaluate them by their own validities, rather than the validities of their premise and conclusion standards.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

anymore. Knowing that such an "inference" is valid now only tells us that a certain position (meaning a combination of assertions and denials) is out of bounds. It no longer prescribes an inferential step we can safely take to expand our knowledge or commitments. Of course, on a bounds reading, the validity of $\Gamma \Rightarrow_n \Delta$ tells us that if we are committed to Γ, we *can't* sensibly commit to denying all of Δ. But it doesn't tell us anything about what we *can* do, in particular, about any consequences or commitments we can deduce given our commitment to Γ. In this way, the logic no longer functions *as a logic* at the level in question.[19] Priest [31, p. 176], Rumfitt [38, pp. 78–81] and Steinberger [40, pp. 352–353] have made similar points about the bounds interpretation.

In case of something like $\mathbf{ST}^1$ it needn't be inherently problematic, Porter maintains, because we can say that what makes the logic a logic is the level of 2-inferences.

> [$\mathbf{ST}^1$] is then not a tool for reasoning about *sentences*, but a tool for reasoning about *positions*. On this interpretation, the 1-inferences are the claims *about which* we are making inferences; they do not themselves correspond to rules that we use to make inference. [28, p. 1383]

Of course, this assumes that the bounds reading covers only the 1-inferences, not the 2-inferences. Crucially, this can be assumed because we don't *need* the bounds interpretation for 2-inferences, because $\mathbf{ST}^1$ is self-obedient at levels $n \geq 2$, so we can safely apply a more conventional interpretation there.

Now recall an observation stressed in §3: for each n, the logics $\mathbf{ST}^n$ and $\mathbf{ST}^n_+$ lack some of the self-obedience criteria at the levels $m \leq n$, but have them at $m > n$. $\mathbf{ST}^\omega$ and $\mathbf{ST}^\omega_+$, on the other hand, fail (PS) and (PD) at *all* levels, as does $\mathbf{ST}^\omega_+$ with (SS) and (SD). These seemingly minor details now become vital. They suggest that, for each n, the logics $\mathbf{ST}^n$ and $\mathbf{ST}^n_+$ can be given a bounds reading up to and including their n-th level, to justify all failures of self-obedience, combined with a conventional interpretation of the remaining levels, to allow the logic to function as a tool for reasoning about (lower level) positions. Making such a combined interpretation plausible would clearly require a lot more explaining, but there seems to at least be some room to work in here. Unfortunately, the same can't be said for $\mathbf{ST}^\omega$ or $\mathbf{ST}^\omega_+$—the crown jewels of the BA-plan. These logics are self-disobedient at every level, and so we'd need to apply the bounds reading to every level to justify this. But that doesn't leave any level for them to be understood or used as genuine logics. For this reason Porter concludes that self-disobedience remains a problem at the absolute top of the **ST**-hierarchy, if not below there.[20]

---

[19] It might even be argued that the bounds reading succeeds in making self-disobedience plausible only precisely by retreating from the canonical application (fn. 14 above) at the relevant level. See, for example, Ripley's own comments in [35, p. 239] about the tenuous connection between bounds and 'logic'.

[20] Though the top isn't $\mathbf{ST}^\omega$ for Porter, but something he calls ST_∞. This is because he includes transfinite metainferential levels (see [22] for an overview). In a more standard setting like ours, with finite levels only, the argument applies to $\mathbf{ST}^\omega$ and $\mathbf{ST}^\omega_+$. Also, Porter sometimes suggests that the problem is that those logics are "nontransitive" at all levels. But his definition of "transitivity" [28, p. 1373] is highly non-standard, concerning neither transitivity of consequence, nor any kind of *Cut* (cf. our fn. 6 above). Porter [28, fn. 6] instead defends its significance via its close relation to (PS)/(PM).

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

On the other hand, BA-planners have repeatedly insisted [3, 5, 27] that $\mathbf{ST}^{\omega}$ and $\mathbf{ST}^{\omega}_{+}$ do prescribe applicable rules of inference, despite criticisms like Porter's (or Golan's [19]). Let's assess the planners' case for this claim. I submit that their arguments are mistaken for essentially the same reason as the mixed obedience argument. We start with Barrio et al.'s direct response to Porter.

> [W]hen interpreted in a mixed way, the **ST**-theories can provide the norms for inferential transitions, as Porter seems to demand of a logic. In [the section on mixed obedience], we have shown that we only need to check [...] whether the tautologies/theorems (the n-validities) also satisfy the standard for premises of the 1-inferences (n + 1-inferences). [3, p. 12582]

The idea is that, while the n-validities of, for example, $\mathbf{ST}^{\omega}$ can't be understood as deduction rules of the form "if the premises are satisfied/valid, we can infer that the conclusion is too," they *can* be understood as rules of the form "if the premises are satisfied/valid *in* $\swarrow\mathbf{ST}^{\omega}(n)$, we can infer that the conclusion is so *in* $\searrow\mathbf{ST}^{\omega}(n)$." The problem with this idea is a familiar one. The mixed obedience response confused *claims about* a logic with the *claims of* that logic. The current argument extends this by confusing *reasoning about* a logic for *reasoning with/according to* it.

Specifically: to infer something like $\vDash_{\searrow\mathbf{ST}^{\omega}(n)}\varphi$ from the assumption that $\vDash_{\swarrow\mathbf{ST}^{\omega}(n)}\Gamma$ does not qualify as reasoning *according to* the validities of $\mathbf{ST}^{\omega}$ (even if $\vDash_{\mathbf{ST}^{\omega}}\Gamma\Rightarrow_{n}\varphi$). For starters, neither $\vDash_{\searrow\mathbf{ST}^{\omega}(n)}\varphi$ nor $\vDash_{\swarrow\mathbf{ST}^{\omega}(n)}\Gamma$ can even be expressed in $\mathbf{ST}^{\omega}$'s object language, hence there is certainly no validity of $\mathbf{ST}^{\omega}$ which says that one follows from the other. Secondly, as pointed out earlier, claims about what is and isn't valid *in* $\swarrow\mathbf{ST}^{\omega}(n)$ or $\searrow\mathbf{ST}^{\omega}(n)$ aren't up for debate. We can all agree about whether one such claim follows from another or not, even if we don't endorse $\mathbf{ST}^{\omega}$ and its validities, hence it can't be $\mathbf{ST}^{\omega}$ which licenses this inferential step. The reason we agree is that we agree on which metatheory is being used, that is, classical set theory.[21] So when we infer $\vDash_{\searrow\mathbf{ST}^{\omega}(n)}\varphi$ from $\vDash_{\swarrow\mathbf{ST}^{\omega}(n)}\Gamma$ given $\vDash_{\mathbf{ST}^{\omega}}\Gamma\Rightarrow_{n}\varphi$, we are in fact using classical logic, not $\mathbf{ST}^{\omega}$.

At this point, the BA-planner may retort that Barrio et al. have nonetheless identified a way in which $\mathbf{ST}^{\omega}$ can be connected to reasoning. But being "connected to" reasoning in this sense—that other logics allow for reasoning about the logic in question—isn't nearly enough. The empty logic is also connected to reasoning in the sense that classical logic can prove all sorts of things about it. Yet none of this reasoning, nor any other reasoning, can be carried out according to the empty logic itself, which is why the latter isn't a viable theory of logic.

The discussion of self-obedience in [3] actually begins with the introduction of another language, consisting of labeled expressions of the form '$X:\varphi$', where X is an n-standard and φ an n-inference. $X:\varphi$ is satisfied in just those models which X-satisfy φ. Notice that satisfaction and hence validity of a labeled expression is not itself relative to a standard. We've thus moved to a completely different system, whose object language is part of what was previously the metalanguage, and whose validities belong to what was previously the metatheory. Yet in the subsequent discussion, the

[21] Metatheory here not in the sense of metainferential levels, but of semantic assent. The literature on the **ST**-hierarchy follows the common practice of developing non-classical logics against the backdrop of a classical metatheory.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

new system's validities *about* a given standard are repeatedly mistaken for validities *of* said standard. For instance, it is claimed that '$S/T : \emptyset \Rightarrow_1 \lambda$' is valid in $\mathbf{ST}^{\omega}_{+}$ [3, p. 12578], or that "**ST** *does not validate Modus Ponens—that is,* $A, A \to B \vDash B$*—, but, actually, this labeled version of it:* $\mathbf{S} : A, \mathbf{S} : A \to B \vDash \mathbf{T} : B$" [3, fn. 24]. It is this misattribution of labeled validities which also leads the BA-planners to maintain that $\mathbf{ST}^{\omega}$ and $\mathbf{ST}^{\omega}_{+}$ provide us with rules for inferential transitions. Namely, [3, 5, 27] appeal to a labeled proof system for $\mathbf{ST}^{\omega}$ called $\mathbf{G3SK}^{\omega}$ [11], which they take to demonstrate how one can/should reason according to $\mathbf{ST}^{\omega}$ or $\mathbf{ST}^{\omega}_{+}$. By now the reader can probably anticipate what I have to say about this appeal: the proof rules of $\mathbf{G3SK}^{\omega}$ govern labeled expressions, hence they represent inferential transitions *about* (the standards of) $\mathbf{ST}^{\omega}$, none of which are by rules *of* $\mathbf{ST}^{\omega}$.[22]

Thus the situation of $\mathbf{ST}^{\omega}$ and $\mathbf{ST}^{\omega}_{+}$ remains problematic. We can perhaps justify their self-disobedience by assigning them a bounds reading, but only at the cost of no longer interpreting their validities as genuine inferences. All attempts to show that the logics nonetheless prescribe rules for inferential transitions amount to misidentifications of what is actually classical metareasoning. Thus a bounds reading would drastically undermine the appeal of $\mathbf{ST}^{\omega}$ and $\mathbf{ST}^{\omega}_{+}$, which were primarily motivated by their supposed solution of the semantic paradoxes. For as Golan observed:

> If there is no genuine reasoning, there is no way to derive the paradox and hence, one could argue, no paradox at all. Ultimately, that is why all attempts to solve the paradox aim at accommodating truth—preferably, transparent truth—within a [...] system in which genuine reasoning can in principle be carried out. [19, p. 817]

In sum, though here was not the place to engage in a full-fledged assessment of the bounds reading, we have vindicated the limitation suggested by Porter. Insofar as the bounds interpretation can justify self-disobedience, this escape route is only open to those mixed logics which aren't self-disobedient at *every* level. At least, that is, if one intends for their logic to be used and understood as a genuine logic—which is certainly a prerequisite for, say, resolving a paradox.

§5. Conclusion. The purpose of this paper has been to illuminate the versions, causes, and consequences of the self-disobedience phenomenon. We have seen that self-obedience can refer to a variety of in general non-equivalent notions. Moreover, we observed that far from every mixed logic is self-disobedient, though we also showed that fairly weak conditions suffice (Theorem 3.4). Finally, I've argued that all attempts to downplay the problematic nature of self-disobedience fail, at least in those cases where self-disobedience affects all inferential levels, due especially to a recurring confusion between claims of and about a logic. Nonetheless, for all that was said, a persistent advocate of a self-disobedient logic has the option of flatly maintaining that the benefits outweigh the costs. For there is still no obvious method of refuting such a claim. However, we should at least demand that they advertise honestly, and acknowledge that self-disobedience (at all levels) is a hefty price. Hefty enough to dissuade most buyers on the market, I would wager, though that remains to be seen.

[22] To be clear: I'm in no way objecting to $\mathbf{G3SK}^{\omega}$. I'm merely making an observation about which kind of reasoning its rules codify.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

To close, let me highlight the fact that the material of §3 depended heavily on the assumption of local validity. To illustrate: if we instead adopt global-global [21], then (SM) becomes a perfectly plausible self-obedience notion at $n \geq 2$, satisfied by all pure logics, while the Porter-style criteria and (SD) become highly unlikely. I leave it for future work to map out the space of self-obedience properties, and their relation to the structure of logics, given global validity. On the other hand, note also that at least regarding (SS), but with the obvious exception of §4.1, the arguments and conclusions of §4 carry over essentially unchanged to a global setting.

§Appendix A. We start by defining the **ST**-hierarchy in full. Recall the arrow notation introduced in §3: if X is a standard, $\uparrow X$ denotes X/X, and if X/Y is a standard, $\swarrow(X/Y)$ denotes X and $\searrow(X/Y)$ denotes Y. We add that $\overline{X/Y}$ denotes Y/X. Moreover, if X is an n-standard, $\widehat{X}$ denotes the full logic generated by X as follows: $\{\searrow^m X \mid 0 < m \leq n\} \cup \{X\} \cup \{\uparrow^m X \mid 0 < m\}$, where $\searrow^m$ signifies m-many successive applications of $\searrow$, and likewise for $\uparrow^m$. Next we define the following sequence of standards for $\mathcal{L}^{PL}$ and the class of Strong Kleene valuations: $T_1 = S/T$, and $T_{n+1} = \overline{T_n}/T_n$. Note that this sequence contains one n-standard, namely T_n, for each $n \geq 1$.

Definition A.1 (The **ST**-hierarchy). *For every* $n \geq 1$, $\mathbf{ST}^n := \widehat{T_n}$. *Moreover,* $\mathbf{ST}^\omega := \{T\} \cup \{T_n \mid n \geq 1\}$.

Next we prove the claim of §3 that these logics meet (SD) and hence (SS). We will make use of a well-known lemma about sharpenings. A sharpening of a Strong Kleene valuation v is any Boolean valuation which agrees with v wherever v already assigned a formula a value in $\{0, 1\}$.

Lemma A.2 (Monotonicity). *For any* $n \geq 1$ *and* n*-inference* φ*, if* v *is a Strong Kleene valuation, and* v' *a sharpening of* v*, then* $v' \Vdash_{T_n} \varphi$ *entails* $v \Vdash_{T_n} \varphi$.

Proof. See [39, Lemma 20]. □

Note also that, if v' is Boolean, then T and S both come down to C as far v' is concerned. Therefore $v' \Vdash_{T_n} \varphi$ iff $v' \Vdash_{\mathbf{CL}} \varphi$.

Theorem A.3. *For all* $n \geq 1$, $\mathbf{ST}^n$ *has* (*SD*).

Proof. Suppose $\vDash_{\mathbf{ST}^n} \Gamma \Rightarrow_m \Delta$, and that $\vDash_{\mathbf{ST}^n} \Gamma$. Take an arbitrary Strong Kleene valuation v. If $m > n$, then $\mathbf{ST}^n(m) = \uparrow\mathbf{ST}^n(m-1)$ and the result is immediate. So suppose instead that $m \leq n$. Then $\mathbf{ST}^n(m-1) = T_{m-1}$. Since $\mathbf{ST}^n$ has the same validities as **CL** at all levels up to n, we also have $\vDash_{\mathbf{CL}} \Gamma \Rightarrow_m \Delta$ and $\vDash_{\mathbf{CL}} \Gamma$. Since **CL** is pure, it has (SD), hence for every Boolean valuation v', we have $v' \Vdash_{\mathbf{CL}} \delta$ for some $\delta \in \Delta$. Our Strong Kleene valuation v sharpens to some Boolean valuation, hence $v \Vdash_{T_{m-1}} \delta$ for some $\delta \in \Delta$ follows by the monotonicity lemma. Therefore, as $\mathbf{ST}^n(m-1) = T_{m-1}$, we have $v \Vdash_{\mathbf{ST}^n} \delta$ for some $\delta \in \Delta$, as needed. □

Consequently, every $\mathbf{ST}^n$ meets (SS). Moreover, (SS) and (SD) of $\mathbf{ST}^\omega$ follow as corollaries, because $\mathbf{ST}^\omega(m) = \mathbf{ST}^n(m)$ whenever $m \leq n$.

It remains to formally define compositionality and constant completeness, and to prove Theorems 3.2 and 3.3.

Definition A.4 (Compositionality). *A* 0*-standard* X *for* $\mathcal{M}$ *and* $\mathcal{L}$ *is* compositional *when there exist a set* τ *of values and a designated subset* $D \subseteq \tau$*, such that each*

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

$M \in \mathcal{M}$ is a function $M : \mathcal{L}_0 \to \tau$ so that $M \Vdash_X \varphi$ iff $M(\varphi) \in D$, and for every n-ary operator $\circ$ of $\mathcal{L}$, there exists a function $f : \tau^n \to \tau$ such that $M(\circ\langle\varphi_1, \ldots, \varphi_n\rangle) = f\langle M(\varphi_1), \ldots, M(\varphi_n)\rangle$ for all $M \in \mathcal{M}$ and $\varphi_1, \ldots, \varphi_n \in \mathcal{L}_0$.

Definition A.5 (Constant completeness). *A compositional 0-standard for $\mathcal{M}$ and $\mathcal{L}$ is* constant complete *when for every value $t \in \tau$, there exists sentence $\theta_t \in \mathcal{L}_0$ such that $M(\theta_t) = t$ for all $M \in \mathcal{M}$.*

An n-standard X/Y for $n > 0$ is compositional (constant complete) if both X and Y are, and a logic is compositional (constant complete) if all its standards are.

We will make use of compositionality and constant completeness as follows. Given any model M and inference φ, let φ^M be the inference obtained from φ by simultaneously replacing every atomic subformula a of every formula appearing in φ with $\theta_{M(a)}$, that is, the constant for the value that M assigned to a. We borrow a lemma about this construction, and use it to prove Theorem 3.2.

Lemma A.6. *Let X be compositional and constant complete. Take any two models M, M' and any inference φ. Then $M \Vdash_X \varphi$ iff $M' \Vdash_X \varphi^M$.*

Proof. See [21, Lemma 4.8]. □

Theorem 3.2. *Let* **L** *be a compositional and constant complete logic. The following are equivalent*:

- **L** *is quasi-pure.*
- **L** *has* (*PD*).
- **L** *has* (*PS*).
- **L** *has* (*SD*).
- **L** *has* (*SS*).

Proof. It suffices to show that quasi-purity entails (PD), and that (SS) entails quasi-purity. Then the rest follows by the entailments depicted in Figure 2.

So suppose first that **L** is quasi-pure, and that $\vDash_{\mathbf{L}} \Gamma \Rightarrow_n \Delta$ and $M \Vdash_{\mathbf{L}} \Gamma$. This means that (i) $\vDash_{\mathbf{L}(n)} \Gamma \Rightarrow_n \Delta$ and (ii) $M \Vdash_{\mathbf{L}(n-1)} \Gamma$. By quasi-purity, (ii) entails $M \Vdash_{\swarrow\mathbf{L}(n)} \Gamma$. This, combined with (i), means that there is some $\delta \in \Delta$ such that $M \Vdash_{\searrow\mathbf{L}(n)} \delta$. By again quasi-purity, the latter entails $M \Vdash_{\mathbf{L}(n-1)} \delta$, which means $M \Vdash_{\mathbf{L}} \delta$. Thus **L** has (PD) as desired.

Finally, we suppose for contraposition that **L** is not quasi-pure, and prove that it does not have (SS). Non-quasi-purity means that there is some n such that either $\mathbf{L}(n) \not\leq \swarrow\mathbf{L}(n+1)$ or $\mathbf{L}(n) \not\geq \searrow\mathbf{L}(n+1)$. In the former case, there exist model M and n-inference φ such that $M \Vdash_{\mathbf{L}(n)} \varphi$ but $M \nVdash_{\swarrow\mathbf{L}(n+1)} \varphi$. Since **L** is compositional and constant complete, it then follows by the previous lemma that $M' \Vdash_{\mathbf{L}(n)} \varphi^M$ and $M' \nVdash_{\swarrow\mathbf{L}(n+1)} \varphi^M$ for every model M'. But this means that both $\vDash_{\mathbf{L}} \varphi^M$ and $\vDash_{\mathbf{L}} \varphi^M \Rightarrow_{n+1} \emptyset$, hence **L** does not have (SS).[23] The case where $\mathbf{L}(n) \not\geq \searrow\mathbf{L}(n+1)$ is analogous. □

Lastly, for Theorem 3.3, we first prove a general lemma about the $\leq$ relation which we defined on standards.

[23] (SS) (like (PS)) also covers inferences without any conclusion, in which case the condition reduces to: if $\vDash_{\mathbf{L}} \Gamma \Rightarrow_n \emptyset$, then $\nvDash_{\mathbf{L}} \Gamma$ (resp. $M \nVdash_{\mathbf{L}} \Gamma$). If we instead restrict (SS) and (PS) to inferences with *precisely one* conclusion, Theorem 3.2 still holds for logics with at least one invalidity at each level.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

LEMMA A.7. *The relation $\leq$ is a partial order.*

Proof. Reflexivity and transitivity are obvious, so we only need to prove antisymmetry: if $X \leq X'$ and $X \geq X'$, then $X = X'$. We show this by induction on the levels. Level 0 is immediate by extensionality of sets. For the induction step, assume antisymmetry holds at n, take $n+1$-standards X/Y and X'/Y', and assume for contraposition that $X/Y \neq X'/Y'$. This means either $X \neq X'$ or $Y \neq Y'$. In the former case, by induction hypothesis either $X \not\leq X'$ or $X \not\geq X'$. Say $X \not\leq X'$ without loss of generality. Then there exist model M and n-inference φ such that $M \Vdash_X \varphi$ but $M \nVdash_{X'} \varphi$. This means $M \nVdash_{X/Y} \varphi \Rightarrow_{n+1} \emptyset$ but $M \Vdash_{X'/Y'} \varphi \Rightarrow_{n+1} \emptyset$. Hence $X/Y \not\geq X'/Y'$. The case where $Y \neq Y'$ is analogous. So $\leq$ is a partial order. □

THEOREM 3.3. *If a logic is both quasi-pure and mixed, it is not reflexive.*

Proof. If $\mathbf{L}$ is mixed, this means that at some n, we *don't* have $\swarrow\mathbf{L}(n+1) = \mathbf{L}(n) = \searrow\mathbf{L}(n+1)$. However, quasi-purity means that $\swarrow\mathbf{L}(n+1) \geq \mathbf{L}(n) \geq \searrow\mathbf{L}(n+1)$. Hence, since $\leq$ is a partial order, it must be that $\swarrow\mathbf{L}(n+1) \not\leq \searrow\mathbf{L}(n+1)$. Thus there are M and φ such that $M \Vdash_{\swarrow\mathbf{L}(n+1)} \varphi$ but $M \nVdash_{\searrow\mathbf{L}(n+1)} \varphi$. Hence $M \nVdash_{\mathbf{L}(n+1)} \varphi \Rightarrow_{n+1} \varphi$, which means $\nvdash_{\mathbf{L}} \varphi \Rightarrow_{n+1} \varphi$, so $\mathbf{L}$ is not reflexive. □

Acknowledgments. I would like to express my gratitude to Pedro del Valle-Inclán, Jonathan Erenfryd, Camillo Fiore, Guillaume Massas, Federico Pailos, and two anonymous referees of *The Review of Symbolic Logic*, for comments on earlier drafts, which lead to substantial improvements.

Funding. No funding to declare.

BIBLIOGRAPHY

[1] Barrio, E., & Bezerra, E. (2024). The logics of a universal language. ***Asian Journal of Philosophy***, **3**(1), 12.

[2] Barrio, E., Fiore, C., & Pailos, F. (2024). Meta-classical non-classical logics. ***The Review of Symbolic Logic***, **17**(4), 1146–1171.

[3] Barrio, E., Pailos, F., & Calderón, J. T. (2021). Anti-exceptionalism, truth and the BA-plan. ***Synthese***, **199**(5), 12561–12586.

[4] Barrio, E., Pailos, F., & Szmuc, D. (2020). A hierarchy of classical and paraconsistent logics. ***Journal of Philosophical Logic***, **49**(1), 93–120.

[5] Bezerra, E., & Erenfryd, J. (2026). *Philosophical applications*. In Teijeiro, P., and Barrio, E., editors. ***Metainferences in Substructural Logics***. Cham: Springer, pp. 195–222.

[6] Carrara, M., & Martino, E. (2014). Logical consequence and conditionals from a dialetheic perspective. ***Logique et Analyse***, **227**, 359–378.

[7] Chemla, E., Egré, P., & Spector, B. (2017). Characterizing logical consequence in many-valued logic. ***Journal of Logic and Computation***, **27**(7), 2193–2226.

[8] Cobreros, P., Egré, P., Ripley, E., & van Rooij, R. (2012). Tolerant, classical, strict. ***Journal of Philosophical Logic***, **41**(2), 347–385.

[9] ———. (2013). Reaching transparent truth. ***Mind***, **122**(488), 841–866.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

[10] ———. (2015). *Vagueness, truth and permissive consequence*. In Achourioti, T., Galinon, H., Fernández, J. M., and Fujimoto, K., editors. ***Unifying the Philosophy of Truth***. Netherlands: Springer, pp. 409–430.

[11] Cobreros, P., La Rosa, E., & Tranchini, L. (2022). Higher-level inferences in the strong-Kleene setting: A proof-theoretic approach. ***Journal of Philosophical Logic***, **51**, 1417–1452.

[12] Dicher, B., & Paoli, F. (2023). Metainferential levels and inferential networks. ***Logique et Analyse***, **264**, 357–384.

[13] Ferguson, T. M., & Ramírez-Cámara, E. (2022). Deep ST. ***Journal of Philosophical Logic***, **51**, 1261–1293.

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

[15] ———. (2015). *What is logical validity?* In Caret, C. R., and Hjortland, O. T., editors. ***Foundations of Logical Consequence***. Oxford: Oxford University Press, pp. 33–70.

[16] Fiore, C. (2026). Identity criteria for metainferential logics. ***Synthese***, **207**(5), Article no. 189.

[17] Fjellstad, A. (2022). Metainferential reasoning on strong Kleene models. ***Journal of Philosophical Logic***, **51**(6), 1327–1344.

[18] Golan, R. (2022). Metainferences from a proof-theoretic perspective, and a hierarchy of validity predicates. ***Journal of Philosophical Logic***, **51**, 1295–1325.

[19] ———. (2023). On the metainferential solution to the semantic paradoxes. ***Journal of Philosophical Logic***, **52**(3), 797–820.

[20] Hlobil, U. (2022). A truth-maker semantics for ST: Refusing to climb the strict/tolerant hierarchy. ***Synthese***, **200**(5), 368.

[21] Kortenbach, B. (2025). Appreciating global validity. ***Synthese***, **206**(2), Article no. 67.

[22] ———. (2025). Every logic is unique: Transfinite metainferences in classical set theory. ***Erkenntnis***, **91**, 3149–3167.

[23] Kortenbach, B., Incurvati, L., & Schloeder, J. J. (2025). Multilateral supervaluationism and classicality. ***Journal of Philosophical Logic***, **54**, 247–290.

[24] Lewis, D. (1973). ***Counterfactuals***. Cambridge: Harvard University Press.

[25] ———. (1986). ***On the Plurality of Worlds***, Vol. 322. Oxford: Blackwell.

[26] Pailos, F. (2020). A fully classical truth theory characterized by substructural means. ***The Review of Symbolic Logic***, **13**(2), 249–268.

[27] Pailos, F., & Da Ré, B. (2023). ***Metainferential Logics***, Vol. 61. Springer Nature.

[28] Porter, B. (2022). Supervaluations and the strict-tolerant hierarchy. ***Journal of Philosophical Logic***, **51**(6), 1367–1386.

[29] Priest, G. (2003). On alternative geometries, arithmetics and logics; a tribute to Łukasiewicz. ***Studia Logica***, **74**(3), 441–468.

[30] ———. (2006). ***In Contradiction: A Study of the Transconsistent*** (second edition). Oxford: Oxford University Press.

[31] ———. (2024). *Substructural solutions to the semantic paradoxes: A dialetheic perspective*. In Petrolo, M., and Venturi, G., editors. ***Paradoxes between Truth and Proof***. Springer, pp. 165–192.

[32] Restall, G. (2005). *Multiple conclusions*. In Hájek, P., Valdés-Villanueva, L., and Westerståhl, D., editors. ***Logic, Methodology and Philosophy of Science: Proceedings of the Twelfth International Congress***. London: Kings College Publications, pp. 189–205.

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press

[33] Ripley, E. (2012). Conservatively extending classical logic with transparent truth. ***The Review of Symbolic Logic***, **5**(2), 354–378.

[34] ———. (2013). Paradoxes and failures of cut. ***Australasian Journal of Philosophy***, **91**(1), 139–164.

[35] ———. (2021). Précis de uncut. ***Análisis filosófico***, **41**(2), 235–260.

[36] ———. (2022). One step is enough. ***Journal of Philosophical Logic***, **51**, 1233–1259.

[37] Roffé, A. J., & Pailos, F. (2021). Translating metainferences into formulae: Satisfaction operators and sequent calculi. ***The Australasian Journal of Logic***, **18**(7), 724–743.

[38] Rumfitt, I. (2008). Knowledge by deduction. ***Grazer Philosophische Studien***, **77**(1), 61–84.

[39] Scambler, C. (2020). Classical logic and the strict tolerant hierarchy. ***Journal of Philosophical Logic***, **49**(2), 351–370.

[40] Steinberger, F. (2011). Why conclusions should remain single. ***Journal of Philosophical Logic***, **40**(3), 333–355.

[41] del Valle-Inclán, P. (2024). Carnap's problem, definability and compositionality. ***Journal of Philosophical Logic***, **53**(5), 1321–1346.

SCUOLA NORMALE SUPERIORE, PISA
ITALY
E-mail: bas.kortenbach@gmail.com

https://doi.org/10.1017/S1755020326101282 Published online by Cambridge University Press