A Truth-Maker Semantics for ST: Refusing to Climb the Strict/Tolerant Hierarchy

Ulf Hlobil

1. Department of Philosophy, Concordia University, Montreal, Canada

2. Department of Philosophy, Univerzita Hradec Králové, Hradec Králové, Czechia

email: ulf.hlobil@concordia.ca

Abstract

The paper presents a truth-maker semantics for Strict/Tolerant Logic (ST), which is the currently most popular logic among advocates of the non-transitive approach to paradoxes. Besides being interesting in itself, the truth-maker presentation of ST offers a new perspective on the recently discovered hierarchy of meta-inferences that, according to some, generalizes the idea behind ST. While fascinating from a mathematical perspective, there is no agreement on the philosophical significance of this hierarchy. I aim to show that there is no clear philosophical significance of meta-inferences above the first level.[1]

1 Introduction

My aims in this paper are twofold: First, I want to present an exact truth-maker semantics, in the spirit of Kit Fine (2017a), for the non-transitive

[1]This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this article is published in *Synthese*, and is available online at https://doi.org/10.1007/s11229-022-03820-w

logic called ST (for "strict/tolerant" logic) (Cobreros et al., 2012). Second, on the basis of this truth-maker semantics, I want to raise some doubts about generalizing the strategy behind ST by moving up a hierarchy of meta-inferences, i.e., by climbing the strict/tolerant hierarchy, as I shall call it (Scambler, 2020a; Barrio et al., 2019b). Thus, my first aim is a clear-cut technical project, while it will become clear in due course that my second aim turns on controversial foundational issues in the philosophy of logic and will, hence, be of a more tentative nature.

Let me clarify my last remark by foreshadowing some aspects of my second project: There is no doubt that the discovery of the strict/tolerant hierarchy is a great piece of logico-mathematical work. However, it raises an issue that, to my mind, hasn't been appreciated enough: What do we model or theorize by defining consequence relations over meta-inferences? How should we understand meta-inferences and their validity at a conceptual, philosophical level? Parallel questions can be raised about inferences, or as some prefer to say—rightly,[2] I think—implications or entailments. And the questions regarding inferences and meta-inferences are not independent of each other. We may perhaps tolerate some lack of clarity regarding what we are modeling or theorizing by consequence relations at the inferential level because we agree enough on particular verdicts about what follows from what. I aim to bring out, however, that this is not the case for the meta-inferences.

In particular, there is no reason to assume that staying close to classical logic is a virtue in meta-inferential consequence relations unless this can be justified on the basis of what we are modeling or theorizing by such meta-inferential consequence relations. Indeed, as I aim to show, in the absence of any philosophical understanding of meta-inferences, it is not even clear that it is a virtue in a logical theory to include claims about

[2]Nevertheless, I will stick to the terminology of inferences as abstract entities and not mental or dialogical acts, as this fits better with the talk of meta-inferences, which has come to be dominant in this area of the literature.

meta-inferential validity. This is crucial because my opponents' argument for climbing the strict/tolerant hierarchy and for not stopping at ST is, in effect, that doing so allows us to accept classical principles (in particular the Cut rule) at all meta-inferential levels.

The truth-maker semantics for ST below suggests a philosophical interpretation of consequence at different levels that casts doubt on the idea that staying classical at meta-inferential levels should count as a virtue. While we can formulate the strict/tolerant hierarchy on the basis of this truth-maker semantics, I cannot find any clear philosophical motivation for taking this hierarchy of relations to be of any philosophical significance, e.g., as a response to the truth-theoretic paradoxes. I will spell out and clarify this argument throughout the paper. However, the development of the truth-maker semantics for ST will take center stage.

Here is the plan: In Section 2, I introduce the strict/tolerant logic ST and the strict/tolerant hierarchy. Next, in Section 3, I present a truth-maker semantics for ST. In Section 4, I use this truth-maker semantics to raise some doubts about moving up the strict/tolerant hierarchy, given the reasons presented in the literature so far. Section 5 concludes.

2 Strict/Tolerant Logic and Its Hierarchy

Probably two of the most remarkable recent developments in the research on truth-theoretic (and some other) paradoxes are (i) the development of the non-transitive approach to paradox, which underwrites every classically valid inference (Cobreros et al., 2012, 2013; Ripley, 2012, 2013), and (ii) the generalization of this strategy to meta-inferences of arbitrarily high levels (Pailos, 2020; Barrio et al., 2019a,b). These two developments provide fruitful soil for much current research (e.g. Scambler, 2020a,b; Cobreros et al., 2021, 2020a,b; Dicher and Paoli, 2019; Fitting, 2021a,b; Da Ré et al., 2020). I want to endorse the first development but reject the second.

Hence, I must provide some background, which is the aim of this section. I'll take the two developments in turn.

2.1 Strict/Tolerant Logic

The basic idea behind the non-transitive approach to paradox is that paradoxical sentences are entailed by the empty set (or something that we accept and cannot rationally give up) and that they entail absurd or clearly false sentences but that this doesn't mean that the empty set (or something we accept) entails absurd or clearly false sentences.

The logic ST is a way to implement this idea. ST is usually presented as a three-valued logic that uses the strong Kleene truth-tables, with truth-values 0, $\frac{1}{2}$, and 1. It differs from Priest's Logic of Paradox (LP) and Strong-Kleene Logic (K3) not in its models but merely in the definition of consequence. In LP, e.g., an inference is valid iff, in all models, the inference preserves value $\geq \frac{1}{2}$ from all the premises to at least one conclusion. By contrast, an inference is ST-valid iff, in every model, either one of the premises doesn't receive value 1 or one of the conclusions doesn't receive value 0. The idea is that whenever the premises all meet the strict standard of having values of 1, at least one conclusion meets the tolerant standard of having a value of at least $\frac{1}{2}$. Hence the label "strict/tolerant" logic. Let's make this precise for propositional logic (to which I restrict myself throughout for simplicity).

Definition 1. Let $\mathfrak{L}$ be the language with a finite stock of atomic sentences, `At`, and the usual syntax, with $\psi \in \texttt{At}$: $\phi ::= \psi \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi$.

Valuations (or models) are functions from sentences of $\mathfrak{L}$ to truth-values in $\{0, \frac{1}{2}, 1\}$ that obey the strong Kleene truth-tables, i.e.:

- $v(\neg\phi) = 1 - v(\phi)$
- $v(\phi \wedge \psi) = min(v(\phi), v(\psi))$

- $v(\phi \vee \psi) = max(v(\phi), v(\psi))$

Definition 2. ST-*validity*: $\models_{ST} \subseteq \mathcal{P}(\mathfrak{L}) \times \mathcal{P}(\mathfrak{L})$ such that $\Gamma \models_{ST} \Delta$ iff for every valuation, v, such that $\forall \gamma \in \Gamma\ (v(\gamma) = 1)$ there is a $\delta \in \Delta$ such that $v(\delta) \in \{\frac{1}{2}, 1\}$.

To see how this logic handles paradoxical sentences (without introducing a first-order language and model-theory), let's see what happens when we add a sentence λ that is equivalent to its negation, i.e., we require that $v(\lambda) = v(\neg\lambda)$ in all valuations. Given the truth-table for negation, the only way to fulfill this requirement is that $v(\lambda) = \frac{1}{2}$, as is familiar from Kripke's (1975) fixed point construction of Kleene-Kripke models. Hence, $\emptyset \models_{ST} \lambda$ and $\lambda \models_{ST} \emptyset$. But $\emptyset \not\models_{ST} \emptyset$. Thus, ST-validity is not transitive, and it is easy to see that the argument extends to all the usual versions of Cut.

One major selling point of ST is that it validates all classically valid inferences. We can add a transparent truth-predicate to ST and the result will be a conservative extension of classical logic (Ripley, 2012). To see this, it suffices to note that any ST counterexample to an inference is a valuation in which none of the premises or conclusions receives the value $\frac{1}{2}$; and if such a counterexample exists, there exists a counterexample in which no sentence receives value $\frac{1}{2}$, i.e., a counterexample that is a classical valuation. Hence, if there is no classical counterexample, then there is no ST counterexample. Indeed, as long as we don't have sentences like λ in the language the consequence relation of ST coincides with classical consequence.

In order to be prepared for what is coming, we should also define the logic TS, which is like ST except that in the definition of validity, the strict and the tolerant standards are flipped.

Definition 3. TS-*validity*: $\models_{TS} \subseteq \mathcal{P}(\mathfrak{L}) \times \mathcal{P}(\mathfrak{L})$ such that $\Gamma \models_{TS} \Delta$ iff for every valuation, v, such that $\forall \gamma \in \Gamma\ (v(\gamma) \in \{\frac{1}{2}, 1\})$ there is a $\delta \in \Delta$ such that $v(\delta) = 1$.

The logic TS has no theorems, like K3, but it also has no valid inferences (unless we add truth-constants or the like to the language). The valuation that gives every sentence value $\frac{1}{2}$ is a counterexample to every inference (French, 2016).

2.2 The Strict/Tolerant Hierarchy

Let us now look at the second development, viz. the generalization of the idea behind ST to meta-inferences of higher levels. I will first explain the general idea and then give some technical details.

This second development can be seen as a continuation of a movement that leads from LP to ST. It is well known that LP includes as theorems all the theorems of classical logic, but it doesn't validate all classical inferences. For example, LP invalidates disjunctive syllogism. To see this, if we let $v(\lambda) = \frac{1}{2}$ and $v(A) = 0$, then this is a counterexample that shows $\lambda \vee A, \neg\lambda \not\models_{LP} A$.[3] One way to look at ST is to see it as doing something similar, as it were, "one level up": ST includes all classical theorems and all classical inferences, but unlike classical logic it doesn't validate Cut and similar principles governing relations between inferences. For example, what we might call "meta-disjunctive-syllogism" fails in ST, i.e., for any A, we have $\emptyset \models_{ST} \lambda \vee A$ and $\emptyset \models_{ST} \neg\lambda$ (because there is no counterexample as long as $v(\lambda) = v(\neg\lambda) = \frac{1}{2}$ in all valuations), but in many cases $\emptyset \not\models_{ST} A$.

It seems fairly natural at this point to think of sentences as, roughly, meta-inferences of level -1, to think of ordinary inferences as meta-inferences of level 0, and think of Cut and meta-disjunctive-syllogism and the like as meta-inferences of level 1. Hence, LP agrees with classical logic up to level -1 but no further, and ST agrees with classical logic up to level 0 but no

[3] In the language I am using here modus ponens corresponds to disjunctive syllogism, as the language has no conditional. The failure of modus ponens in LP is what corresponds most closely to the failure of Cut in ST.

further. And once we have this—at this point admittedly still vague—idea, we can ask: Is there a logic that agrees with classical logic up to level 1, or 2, or 3, ... or perhaps even some arbitrarily high level n? Over the last couple of years, Barrio, Pailos, and Szmuc have shown that the (astonishing) answer is: Yes, there are such logics for any finite n, and we can even push this to the first limit ordinal ω and agree with classical logic at all finite levels, while still having a transparent truth-predicate (Barrio et al., 2019a; Pailos, 2020; Barrio and Pailos, 2022).[4]

To see how this hierarchy of logics is constructed, we must define what meta-inferences of arbitrary levels are and how we can define logics governing such objects.

Definition 4. *Meta-inferences*: A meta-inference of level -1 is a sentence (of $\mathfrak{L}$). A meta-inference of level $n+1$, written $\Gamma \overset{n+1}{\Rightarrow} \Delta$ (or just $\Gamma \Rightarrow \Delta$ if the level is clear from context), is a pair of sets of meta-inferences of level n, namely the premise meta-inferences Γ and the conclusion meta-inferences Δ.

To define consequence relations for meta-inferences, we say that valuations that are not counterexamples to an inference *satisfy* that inference. We can now quantify over valuations to define validity at arbitrary levels. This notion of validity for meta-inferences is often called the "local" notion.

Definition 5. *Meta-inferential validity (local)*: A meta-inference of level n, $\Gamma \Rightarrow \Delta$, is valid according to the pair of logics L_1 and L_2, written $\models_{L_1 L_2} \Gamma \Rightarrow \Delta$, iff, for all valuations v, if v satisfies every $\gamma \in \Gamma$ according to L_1, then v satisfies at least one $\delta \in \Delta$ according to L_2. We say that a meta-inference is valid in classical logic (CL) iff it is valid according to the pair of CL and CL.

[4]Scambler (2020b) has shown how to extend this result into the transfinite.

For level -1, we say that a meta-inference of level -1, i.e. a sentence, is satisfied according to "strict logic" S (and in CL, with valuations restricted to classical ones) by a valuation iff the valuation gives it value 1; and it is satisfied according to "tolerant logic" T by a valuation iff the valuation gives it value $\frac{1}{2}$ or 1. Thus, we have, e.g., $\Gamma \models_{ST} \Delta$ iff $\models_{ST} \Gamma \overset{0}{\Rightarrow} \Delta$ and $\Gamma \models_{TS} \Delta$ iff $\models_{TS} \Gamma \overset{0}{\Rightarrow} \Delta$. But we can now also evaluate meta-inferences of higher levels. And we can express the failure of Cut in ST as follows: $\not\models_{STST} \{\emptyset \overset{0}{\Rightarrow} \lambda, \lambda \overset{0}{\Rightarrow} \emptyset\} \overset{1}{\Rightarrow} \{\emptyset \overset{0}{\Rightarrow} \emptyset\}$. This says that the meta-inference from $\emptyset \overset{0}{\Rightarrow} \lambda$ and $\lambda \overset{0}{\Rightarrow} \emptyset$ to $\emptyset \overset{0}{\Rightarrow} \emptyset$ is not valid by the standards of STST. This is so because there is a valuation that satisfies $\emptyset \overset{0}{\Rightarrow} \lambda$ and $\lambda \overset{0}{\Rightarrow} \emptyset$ according to ST but does not satisfy $\emptyset \overset{0}{\Rightarrow} \emptyset$ according to ST; indeed every valuation is like that.

Notice that while our instance of Cut is invalid in STST, it is valid in TSST. After all, since $v(\lambda) = \frac{1}{2}$ in every valuation, there is no valuation that satisfies $\emptyset \overset{0}{\Rightarrow} \lambda$ and $\lambda \overset{0}{\Rightarrow} \emptyset$ according to TS. Indeed, TSST agrees with classical logic up to level 1 (Barrio et al., 2019a).

Since, for every inference, a valuation satisfies the inference according to TS only if it also satisfies the inference according to ST but not vice versa, TS is a stricter standard for inferences than ST. Hence, if we want to evaluate our premises by a stricter standard than our conclusions, as does ST at the inferential level, it may seem that we should be partial to TSST and prefer it to STST at the meta-inferential level 1. Following this idea up the inferential levels leads to the following definition of the strict/tolerant hierarchy.

Definition 6. *Strict/Tolerant hierarchy*: The strict/tolerant hierarchy of logics ST^n is the collection of logics for meta-inferences such that, for every level n, the hierarchy includes exactly $\models_{ST^n} \Gamma \overset{n}{\Rightarrow} \Delta$, where ST^n is defined inductively thus: $\mathsf{ST}^{n+1} = \mathsf{TS}^n\mathsf{ST}^n$ and $\mathsf{TS}^{n+1} = \mathsf{ST}^n\mathsf{TS}^n$, and $\mathsf{ST}^0 = \mathsf{ST}$ and $\mathsf{TS}^0 = \mathsf{TS}$, as defined above. A meta-inference is valid in ST^ω iff it is valid in some logic ST^j $(-1 \leq j < \omega)$.

The astonishing results that Barrio, Pailos, and Szmuc obtained are (Barrio et al., 2019a; Pailos, 2020):

Fact 7. *(General Collapse Result) For every level n* ($-1 \leq j < \omega$)*, a meta-inference,* $\Gamma \overset{n}{\Rightarrow} \Delta$*, of level n is valid according to* CL *if and only if it is valid in* ST^n*.*

(Absolute General Collapse Result) For every level n ($-1 \leq j < \omega$)*, a meta-inference,* $\Gamma \overset{n}{\Rightarrow} \Delta$*, of level n is valid according to* CL *if and only if it is valid in* ST^ω*.*

(Truth Result) ST^ω *can be expanded with a transparent truth-predicate, i.e., the resulting theory is satisfiable.*

Thus, we have a a logic ST^ω, which is the union of the consequence relations at each level of the strict-tolerant hierarchy, that agrees with classical logic at every finite meta-inferential level while allowing us to add a transparent truth-predicate.

2.3 Classicality and the Strict/Tolerant Hierarchy

The logic ST^ω can easily look like the "holy grail" theorists of truth have been searching at least since Tarski. After all, insofar as ST is superior to LP because it agrees with classical logic not merely in its theorems but also in its valid inferences, any ST^{n+1} can seem superior to ST^n in a parallel way: it agrees with classical logic not only on all meta-inferences up to level n but also on the meta-inferences of level $n+1$. And ST^ω can seem thus superior to all ST^n with finite n. That is (roughly) how Barrio, Pailos, and Calderón interpret their results. They call this the "Buenos Aires Plan." They write (with their labels mildly adjusted to fit mine):

> Classical logic is the logic widely accepted. There are not enough good reasons to doubt the correctness of classical logic as applied to our mathematical, physical, and other scientific theories. [...] It is methodologically possible to revise CL to deal

> with paradoxes. [...] The *Buenos Aires Plan* gives an alternative explanation. The logic ST^ω (based on the hierarchy) is fully compatible with transparent truth and does not have to recommend giving up any classical reasoning patterns [... By the] *maxim of minimal mutilation*, one concludes that the hierarchy of ST theories is better than sub-classical and ST alternatives. One can deal with naive truth without rejecting classical patterns of reasoning. (Barrio et al., 2021, sec. 1)

Thus, the key virtue that Barrio, Pailos, and Calderón see in ST^ω is that it agrees with classical logic at all meta-inferential levels.

Chris Scambler (2020a; 2020b) doesn't endorse any logic in the strict/tolerant hierarchy, but he nevertheless agrees with the Buenos Aires Plan that anyone tempted by ST should feel forced to climb the hierarchy. Here is Scambler, with his notation changed to fit mine.

> It seems to me that if Ripley's use of [ST] is attractive, one can make a case that each theory [ST^n] for $n > 1$ is still more attractive, because it gets us more classical logic. If it was a good idea to expand the horizons of classicality from mere [LP] to include $MP(\lambda)_1$ [i.e. $\{\lambda, \neg\lambda \vee \bot\} \overset{1}{\Rightarrow} \{\bot\}$][5] in the theory [ST], why isn't it good to have $MP(\lambda)_2$ [i.e. $\{\overset{1}{\Rightarrow} \lambda, \overset{1}{\Rightarrow} \neg\lambda \vee \bot\} \overset{2}{\Rightarrow} \{\overset{1}{\Rightarrow} \bot\}$] as well, as in the theory [ST^2], pushing back the boundaries of non-classicality to the third level and denying only $MP(\lambda)_3$ and related principles? Once one starts down this road, there is no natural stopping point; accepting $MP(\lambda)_{9336}$ but denying $MP(\lambda)_{9337}$ would be bizarre. (Scambler, 2020a, 367)

Here Scambler says that it would be bizarre to find some logic at level n in the strict/tolerant hierarchy attractive, for the reasons that Ripley's first-level ST is attractive, without finding the logic at level $n + 1$ even more

[5] Scamber uses the material conditional to formulate MP. I move to disjunctive syllogism because my official language here doesn't have a conditional.

attractive. In his paper, he follows this up with an argument that aims to show that we should not endorse ST^{ω}. And he takes these considerations together to show that even the first step on the strict/tolerant hierarchy, the one to ST (or perhaps already the adoption of LP) is a mistake. So Scambler and Barrio and his colleagues hold that an advocate of ST should feel forced to climb the hierarchy because, at least by her own lights, she should consider the higher levels closer and closer to classical logic.

David Ripley (2021) disagrees and holds that the basic aims and ideas behind ST and the non-transitive approach don't push us "up the hierarchy." Rather, we can and should stop at level 1, i.e., with ST. As already intimated, I will side with Ripley below.

From what Scambler and Barrio and colleagues say, it is clear that a crucial question for this debate is whether it is a virtue in a logical theory—or at least in a logical theory that is intended to address the paradoxes—if the theory validates as much as possible of classical logic at meta-inferential levels 1 and above (and if so, how much weight this virtue has in comparisons between theories). One reason why the answer to this question isn't obvious is that it is unclear what pre-theoretical or intuitive phenomenon we are modeling or theorizing by giving formal accounts of meta-inferences of levels 1 and above.

To see this, it may be helpful to notice that the ST^{ω}-theorist and the ST-theorist agree on the following verdicts: $\varnothing \overset{0}{\Rightarrow} \lambda$ and $\lambda \overset{0}{\Rightarrow} \varnothing$ are both valid but $\varnothing \overset{0}{\Rightarrow} \varnothing$ is not valid. They disagree, however, on whether $\{\varnothing \overset{0}{\Rightarrow} \lambda, \lambda \overset{0}{\Rightarrow} \varnothing\} \overset{1}{\Rightarrow} \{\varnothing \overset{0}{\Rightarrow} \varnothing\}$ is valid. Interestingly, they agree on the set of valuations that we must look at in order to determine whether this meta-inference is valid. Moreover, they agree, for each meta-inferential level, what relations hold between the inferences in terms of valuations; they agree on all the mathematical facts, described as patterns in valuations (without using "consequence"-talk). However, they disagree on which of these patterns we should call a "counterexample" at a given level.

They thus merely disagree on which of the relations defined in terms of the agreed-upon valuations we should call the "meta-inferential consequence relation at level n."

This disagreement is intractable until and unless we know the significance of saying that a particular meta-inference is invalid (i.e. has a counterexample) in terms that are independent of our formal theories. We need a philosophical, non-technical account of what meta-inferences are and what it means that they are valid or invalid. Ideally such an account would give us substantive constraints on which meta-inferences at level 1 and above are valid, which we could then use to arbitrate the dispute between the ST^{ω}-theorist and the ST-theorist.

One important question for such an account will be whether we should treat inferences and meta-inferences alike. For if we think that the inferences and meta-inferences differ in important respects, then it is not *prima facie* implausible to hold that staying close to classical logic may be a virtue in inferential consequence relations but not in meta-inferential consequence relations. Below I will cast some doubt on the idea that we must treat the inferential and the meta-inferential levels alike.

Since I don't have a general account of what we model by consequence relations of different levels, I will try to make progress in a more limited way, namely by formulating the same issue in a different formalism, i.e., in the truth-maker semantics below. I think that this is helpful because as long as we have only strong Kleene models on the table, the disagreement between the ST^{ω}-theorist and the ST-theorist looks like a merely technical disagreement regarding which valuations should count as counterexamples. And it is hard to see what philosophical considerations could speak for or against the different views. Staying as close as possible to classical logic may then look like one of the very few plausible metrics on which we can compare the views. This temptation will be significantly reduced, I think, once we can look at the issues from another perspective.

3 Truth-Maker Semantics for ST

In this section, I present a truth-maker semantics for ST. The basic setup of the semantic theory is due to Fine (2017a). The two main differences are, firstly, my definition of consequence and, secondly, my rejection of what Fine calls "Exhaustivity." Under my definition of consequence, Exhaustivity turns out to be a version of Cut. That is why rejecting Exhaustivity yields ST in this framework. The resulting truth-maker semantics for ST is more flexible than the usual three-valued semantics for ST.

3.1 Semantic Setup

The basic idea behind exact truth-maker semantics is that sentences are made true or verified by some states and made false or falsified by others. Exact truth-makers and falsity-makers are exact in the sense that they must be wholly relevant to the truth or, respectively, falsity of the sentence.[6] A state that includes an exact truth-maker of a sentence as a proper part need not be an exact truth-maker of the sentence; I'll call it an inexact truth-maker, and similarly for falsity-makers. States can be parts of other states, and some of them are possible and others impossible. To see how this works, let's start with what Fine (2017a, 647) calls a "modalized state space":

Definition 8. *Modalized state space*: A modalized state space is a triple, $\langle S, S^{\Diamond}, \sqsubseteq \rangle$, such that S is a non-empty set of states, $S^{\Diamond} \subseteq S$ (intuitively: the possible states), and $\sqsubseteq$ is a partial order on S (intuitively: parthood), such that all subsets of S have a least upper bound.

[6] Exact truth-makers are similar to what are sometimes called "minimal truth-makers," i.e., truth-makers of sentences such that no proper part of them is a truth-maker of the respective sentence (O'Conaill and Tahko, 2016; Armstrong, 2004). There are, however, differences. The fusion of truth-makers of each disjunct of a disjunction is an exact truth-maker of the disjunction—by the third disjunct of the clause (or+) below—while having two proper parts that are also truth-makers of the disjunction.

The least element, $\square$, is the "null state" that is part of every state. We define fusions of states as the least upper bounds of our partial order (Fine, 2017a, 646).

Definition 9. *Fusion*: The fusion of a set of states $T = \{t_1, t_2, t_3, ...\}$, written $t_1 \sqcup t_2 \sqcup t_3 ...$ or $\bigsqcup T$, is the least upper bound of T with respect to $\sqsubseteq$.

Let $\mathfrak{L}$ be as above. A model is a modalized state space together with an interpretation function, $|\cdot|$, that assigns to each sentence A its interpretation $|A|$, which is a pair consisting of the set of A's exact verifiers, written $|A|^+$, and the set of A's exact falsifiers, written $|A|^-$.

Definition 10. *Model*: Given a language $\mathfrak{L}$, a model, $\mathcal{M}$, is a quadruple $\langle S, S^\Diamond, \sqsubseteq, |\cdot| \rangle$, where $\langle S, S^\Diamond, \sqsubseteq \rangle$ is a modalized state space and $|\cdot|$ is a interpretation function, such that $|A| = \left\langle |A|^+, |A|^- \right\rangle \in \mathcal{P}(S) \times \mathcal{P}(S)$.

We write $\mathcal{M}, s \Vdash A$ if a state s exactly verifies a sentence A in model $\mathcal{M}$, and if no risk of confusion arises, simply $s \Vdash A$. Similarly, $s \dashV A$ says that s is an exact falsifier of sentence A (in model $\mathcal{M}$). We require that interpretation functions satisfy the following semantic clauses for the logical connectives.

(atom+) $s \Vdash p$ iff $s \in |p|^+$

(atom−) $s \dashV p$ iff $s \in |p|^-$

(neg+) $s \Vdash \neg B$ iff $s \dashV B$

(neg−) $s \dashV \neg B$ iff $s \Vdash B$

(and+) $s \Vdash B \wedge C$ iff $\exists u, t\, (u \Vdash B$ and $t \Vdash C$ and $s = u \sqcup t)$

(and−) $s \dashV B \wedge C$ iff $s \dashV B$ or $s \dashV C$ or $\exists u, t\, (u \dashV B$ and $t \dashV C$ and $s = u \sqcup t)$

(or+) $s \Vdash B \vee C$ iff $s \Vdash B$ or $s \Vdash C$ or $\exists u, t\, (u \Vdash B$ and $t \Vdash C$ and $s = u \sqcup t)$

(or−) $s \dashV B \vee C$ iff $\exists u, t\, (u \dashV B$ and $t \dashV C$ and $s = u \sqcup t)$

It will prove convenient below to extend the definition of (exact) truth-makers and falsity-makers to sets of sentences as follows:

Definition 11. *Truth- and falsity-makers of sets*: $u \Vdash \Gamma$ iff $u \Vdash \bigwedge \Gamma$, unless $\{x : x \Vdash \bigwedge \Gamma\} = \emptyset$ in which case $\square$ and nothing else makes Γ true. And $t \dashVv \Delta$ iff $t \dashVv \bigvee \Delta$, unless $\{x : x \dashVv \bigvee \Delta\} = \emptyset$ in which case $\square$ and nothing else makes Δ false.

So far, we have put no constraints on the possible states. Fine often imposes the following constraints, and they will become important below:

Downward-Closure: If $s \in S^{\Diamond}$ and $t \sqsubseteq s$, then $t \in S^{\Diamond}$.

Exclusivity: If $s \in |p|^{+}$ and $t \in |p|^{-}$, then $\forall u(s \sqcup t \sqcup u \notin S^{\Diamond})$.[7]

Exhaustivity: $\forall u \in S^{\Diamond}$, either $\exists s \in |p|^{+} \ (u \sqcup s \in S^{\Diamond})$ or $\exists t \in |p|^{-} \ (u \sqcup t \in S^{\Diamond})$.

Downward-Closure says that all parts of a possible state are possible. Exclusivity says that if you take any atomic[8] sentence, then if you fuse one of its verifiers with one of its falsifiers together with any state, you always get an impossible state. And Exhaustivity says that if you have a possible state and an atomic sentence, then you can extend it to a possible state either by fusing it with a verifier of the sentence or by fusing it with a falsifier of the sentence.

So far, (almost) everything is familiar from Fine's work (Fine and Jago, 2018; Fine, 2017a,b, 2016, 2014). However, we can now define consequence in a new way:

Definition 12. *Truth-Maker Validity* ($\models_{TM}$): $\Gamma \models_{TM} \Delta$ iff, in every model, $s \notin S^{\Diamond}$ for all $s = u \sqcup t$ such that $u \Vdash \Gamma$ and $t \dashVv \Delta$.

---

[7]This formulation differs from Fine's in the quantification over further states u. In the presence of Downward-Closure, the two formulations are equivalent.

[8]Stipulating these constraints for atomic sentences suffices (given the semantic clauses) to enforce them for the whole language.

If there is a possible state $s = u \sqcup t$ such that $u \Vdash \Gamma$ and $t \dashVv \Delta$, then that state is an exact counterexample to $\Gamma \models_{TM} \Delta$. A possible state that includes an exact counterexample as a part is called an "inexact counterexample." The idea behind truth-maker validity is that what it means for Δ to follow from Γ is that there isn't any possible state that makes everything in Γ true but everything in Δ false. In other words, any state that makes all the premises true makes it impossible for all the conclusions to be false.

Note that this notion of consequence captures ideas that have been independently motivated. In particular, modalists about consequence hold that B follows from A iff the conjunction of A and the negation of B is impossible (Bueno and Shalkowski, 2013, 11-12). Since the states that verify A and the negation of B are the fusions of a verifier of A and a falsifier of B and we can generalize this to multiple premises and conclusions, truth-maker validity is recognizably a version of the idea behind modalism. According to truth-maker validity, consequence is an essentially modal notion that holds if the truth of the premises is incompatible with the falsity of the conclusions.

It is easy to see that if we impose all three of Fine's constraints on possible states, then $\models_{TM}$ coincides with classical (propositional) logic (I'll leave this as an exercise to the reader). More importantly for our purposes, rejecting Exhaustivity yields ST, and that is the topic of the next subsection.

3.2 Recasting ST in Truth-Maker Theory

It turns out that the three constraints in the truth-maker theory just presented are versions of the structural rules of Monotonicity, Reflexivity, and Cut, respectively. Since ST results from dropping Cut from classical logic, we are only interested in the truth-maker version of Cut, i.e., Exhaustivity. From now on, I will use $\models_{TM}$ for truth-maker validity over models in which Downward-Closure and Exclusivity hold, while Exhaustivity can fail.

Let's start by showing that we really get ST by dropping Exhaustivity. The following lemmas will prove useful for that and more below:

Lemma 13. *There is a strong Kleene valuation, v, that partitions all sentences such that $v(\gamma) = 1$ iff $\gamma \in \Gamma$, and $v(\delta) = 0$ iff $\delta \in \Delta$, and $v(\theta) = \frac{1}{2}$ iff $\theta \in \Theta$, if and only if there is a possible state in a truth-maker model that is an inexact truth-maker of all $\gamma \in \Gamma$, an inexact falsity-maker for all $\delta \in \Delta$, and neither an inexact truth-maker nor an inexact falsity-maker for all $\theta \in \Theta$.*

Proof. Left-to-right: Suppose that there is a strong Kleene valuation that partitions all sentences such that $v(\gamma) = 1$ iff $\gamma \in \Gamma$, and $v(\delta) = 0$ iff $\delta \in \Delta$, and $v(\theta) = \frac{1}{2}$ iff $\theta \in \Theta$. Take what Fine (2017a, 647) calls the "modalized canonical space," S_c, over our set of atoms, $\mathtt{At}$, namely: S is the powerset of the literals over $\mathtt{At}$, i.e., $S = \mathcal{P}(\mathtt{At} \cup \{\neg p : p \in \mathtt{At}\})$. The only exact truth-maker of an atom p is $\{p\}$ and the only exact falsity-maker is $\{\neg p\}$. Satisfying Exclusivity, we let the possible states be exactly those that don't include an atomic sentence and its negation; i.e., $S^{\Diamond} = \{s \in S : s$ does not contain both a sentence p and $\neg p\}$. And parthood is set-theoretic inclusion: $\sqsubseteq = \{\langle s, s' \rangle : s \subseteq s' \subseteq S\}$. In this truth-maker model, we take the state, s_v, such that, for all atomic sentences, $p \in s$ iff $v(p) = 1$ and $\neg p \in s$ iff $v(p) = 0$. We leave out all the atoms who have value $\frac{1}{2}$ in v. Now, s is possible because there isn't any atom for which $v_s(p) = 1$ and $v_s(p) = 0$. Moreover, s is an inexact truth maker for all $\gamma \in \Gamma$, an inexact falsity-maker for all $\delta \in \Delta$, and neither an inexact truth-maker nor an inexact falsity-maker for all $\theta \in \Theta$. To see this, we argue by induction on the complexity of a sentence, A, that there is a $t \sqsubseteq s$ such that $t \Vdash A$ iff $v(A) = 1$, and there is a $t \sqsubseteq s$ such that $t \dashV A$ iff $v(A) = 0$. Our proposition is immediate for the base case in which $A \in \mathtt{At}$. For the induction step, we first look at the cases where $v(A) = 1$: If $A = B \wedge C$, then $v(A) = 1$ iff $1 = \min(v(B), v(C))$. Hence, by our induction hypothesis, $\exists t, u \sqsubseteq s$ such that

$t \Vdash B$ and $u \Vdash C$. If $A = B \vee C$, then $v(A) = 1$ iff $1 = \max(v(B), v(C))$. Hence, by our induction hypothesis, $\exists t \sqsubseteq s$ such that $t \Vdash B$ or $t \Vdash C$. If $A = \neg B$, then $v(A) = 1$ iff $1 = v(\neg B)$ iff $0 = v(B)$. Hence, by our induction hypothesis, $\exists t \sqsubseteq s$ such that $t \dashV B$. The cases where $v(A) = 0$ are analogous. So the remaining sentences, i.e. those in Θ, have neither an inexact truth-maker nor an inexact falsity-maker in s.

Right-to-left: Suppose that there is a possible state in a truth-maker model that is an inexact truth-maker of all $\gamma \in \Gamma$, an inexact falsity-maker for all $\delta \in \Delta$, and neither an inexact truth-maker nor an inexact falsity-maker for all $\theta \in \Theta$. We now define a strong Kleene valuation, v_s, on the basis of the state s in the following way: $v_s(A) = 1$ iff there is a $t \sqsubseteq s$ such that $t \Vdash A$. $v_s(A) = 0$ iff there is a $t \sqsubseteq s$ such that $s \dashV A$. And $v_s(A) = \frac{1}{2}$ otherwise. We don't have to worry about the case where s includes a verifier and also a falsifier of A because, given Exclusivity, such states are impossible (as an induction on sentence complexity shows). To see that v_s is indeed a strong Kleene valuation, note that $v_s(A \wedge B) = \min(v_s(A), v_s(B))$. For, by (and+), $\exists t \sqsubseteq s$ such that $t \Vdash A \wedge B$ iff $\exists u \sqsubseteq s$ such that $u \Vdash A$ and $\exists w \sqsubseteq s$ such that $w \Vdash B$; hence, $v_s(A \wedge B) = \min(v_s(A), v_s(B)) = 1$. And, by (and$-$), $\exists t \sqsubseteq s$ such that $t \dashV A \wedge B$ iff $\exists u \sqsubseteq s$ such that $u \dashV A$ or $\exists w \sqsubseteq s$ such that $w \dashV B$; hence, $v_s(A \wedge B) = \min(v_s(A), v_s(B)) = 0$. And in all other cases $v_s(A \wedge B) = \min(v_s(A), v_s(B)) = \frac{1}{2}$. Similarly, $v_s(A \vee B) = \max(v_s(A), v_s(B))$. For, by (or+), $\exists t \sqsubseteq s$ such that $t \Vdash A \vee B$ iff $\exists u \sqsubseteq s$ such that $u \Vdash A$ or $\exists w \sqsubseteq s$ such that $w \Vdash B$. By (or$-$), $\exists t \sqsubseteq s$ such that $t \dashV A \vee B$ iff $\exists u \sqsubseteq s$ such that $u \dashV A$ and $\exists w \sqsubseteq s$ such that $w \dashV B$. And in all other cases $v_s(A \vee B) = \min(v_s(A), v_s(B)) = \frac{1}{2}$. Moreover, $v_s(\neg A) = 1 - v_s(A)$. For, $\exists t \sqsubseteq s$ such that $t \Vdash \neg A$ iff $\exists t \sqsubseteq s$ such that $t \dashV A$. So, $v_s(\neg A) = 1$ iff $v_s(A) = 0$. And $\exists t \sqsubseteq s$ such that $t \dashV \neg A$ iff $\exists t \sqsubseteq s$ such that $t \Vdash A$. So, $v_s(\neg A) = 0$ iff $v_s(A) = 1$. Otherwise, $v_s(\neg A) = \frac{1}{2}$. ■

Lemma 14. *There is a state in a truth-maker model that is an inexact counterexample to all and only the inferences in a collection of inferences $\mathscr{I}$ iff there is a strong Kleene valuation that is an* ST *counterexample to all and only the inferences in $\mathscr{I}$.*

Proof. A counterexample in a truth-maker model is a state that inexactly verifies all the premises and inexactly falsifies all the conclusions of all and only the inferences in $\mathscr{I}$. By Lemma 13, for any possible state there is a strong Kleene model such that it assigns 1 to all the sentences that the state inexactly verifies, assigns 0 to all the sentences that the state inexactly falsifies, and assigns $\frac{1}{2}$ to all other sentences. And Lemma 13 also allows us to go in the other direction. ■

Proposition 15. *If we impose Downward-Closure and Exclusivity but not Exhaustivity, then $\Gamma \models_{TM} \Delta$ iff $\Gamma \models_{ST} \Delta$.*

Proof. Immediate from Lemma 14: there is a truth-maker counterexample to $\Gamma \models_{TM} \Delta$ iff there is a strong Kleene counterexample to $\Gamma \models_{ST} \Delta$. ■

I have now reached the first aim of this paper, namely to give a truth-maker semantics for ST. Before moving on to meta-inferences, let's consider how this works without thinking about the usual three-valued semantics.

To see how Cut fails in $\models_{TM}$, let's add the paradoxical sentence λ again. Recall that we require that $v(\lambda) = v(\neg\lambda)$. In our truth-maker context, this means that λ should be verified whenever $\neg\lambda$ is verified and vice versa. But, by (neg+), a state verifies $\neg\lambda$ iff it falsifies λ. So, a state verifies λ iff it falsifies λ. Hence, in our new setting, we must require that, for any state s in any truth-maker model, $s \Vdash \lambda$ iff $s \dashv\!\mid\!\mid \lambda$.

Proposition 16. *For paradoxical sentences—i.e. sentences such that $s \Vdash A$ iff $s \dashv\!\mid\!\mid A$—Cut fails, and more precisely: $\Gamma \models_{TM} A, \Delta$ and $\Gamma, A \models_{TM} \Delta$ but not always $\Gamma \models_{TM} \Delta$.*

Proof. $\Gamma \models_{TM} A, \Delta$ and $\Gamma, A \models_{TM} \Delta$ because $s \Vdash A$ iff $s \dashv\!\!\vert\!\vert A$. And since Exhaustivity applies to arbitrary sentences and not only atomic sentences, $\forall u(s \sqcup u \notin S^{\Diamond})$ if $s \Vdash A$. But we just have to set $\Gamma = \Delta = \emptyset$ and look at a model in which $\square$ is possible to show that it is not generally true that $\Gamma \models_{TM} \Delta$. ∎

What this shows is that, in the truth-maker version of ST, Cut fails because paradoxical sentences can neither be verified nor falsified by any part of a possible state. Exhaustivity tells us that for any possible state and any sentence, we can always extend the state with either a truth-maker or a falsity-maker of the sentence without reaching an impossible state. But that fails for paradoxical sentences because adding a truth-maker of a paradoxical sentence to a state is to also add a falsity-maker of the sentence, and a state that includes a verifier and also a falsifier of the same sentence is always impossible.

Notice that this explanation of why Cut fails is isomorphic to Ripley's (2013; 2015) explanation of why Cut fails in terms of normative bilateralism. As one of the chief advocates of ST, Ripley says that Cut fails for paradoxical sentences because what it means for Δ to follow from Γ is for it to be out-of-bounds (i.e. normatively ruled out by a coherence norm governing speech acts) to assert everything in Γ and also deny everything in Δ but it is always out-of-bounds to assert or deny paradoxical sentences. Hence, it can happen that asserting everything in Γ and also denying everything in Δ is in-bounds (so $\Gamma \not\models \Delta$) but adding an assertion or a denial of λ will make the collection of assertions and denials out-of-bounds (so $\Gamma, \lambda \models \Delta$ and $\Gamma \models \lambda, \Delta$). Similarly, in truth-maker semantics, failures of Cut are cases in which we can extend a possible state neither with a truth-maker nor with a falsity-maker of a sentence. Just as Ripley says that an in-bounds collection of assertions and denials can neither include an assertion nor a denial of the liar sentence, so truth-maker semantics says that a possible state can neither include a truth-maker nor a falsity-maker of the liar sentence.

Thus, the truth-maker semantics seems closer to the philosophical ideas behind ST than the usual three-valued semantics, at least for advocates of ST who are normative bilateralists, like Ripley. This suggests that the truth-maker semantics may offer a perspective on meta-inferences that is more in line with the philosophical ideas behind ST than that offered by the strong Kleene semantics. To investigate this suggestion, we must see how meta-inferences look from the perspective of truth-maker semantics. So, let's turn to meta-inferences.

3.3 Meta-Inferences in Truth-Maker ST

In this subsection I will first define local meta-inferential validity for level 1 in our truth-maker semantics. I will then discuss how this definition could be extended to meta-inferential validity for higher levels, and I will show what it would mean to climb the strict/tolerant hierarchy from the perspective of truth-maker semantics.

As we have seen, in the truth-maker version of ST, counterexamples are not valuations but states in models. We can say that a state satisfies an inference iff the state is not an inexact counterexample to the inference. With this notion in hand, we can adapt the local notion of meta-inferential validity that we encountered above as follows.

Definition 17. *Truth-maker meta-inferential validity (local)*: A meta-inference of level 1, $\Gamma \Rightarrow \Delta$, is valid according to $\models_{TM}$ iff, for all states s in all truth-maker models, if s satisfies every inference $\gamma \in \Gamma$, then s satisfies at least one inference $\delta \in \Delta$.

Like the local notion of meta-inferential validity in the three-valued setting, the definition says that a meta-inference is valid iff all counterexamples to all the conclusion-inferences are counterexamples to at least one premise-inference, in the only sense of "counterexample" we have (so far) used in our truth-maker theory, namely that of a state that inexactly verifies all the premises and inexactly falsifies all the conclusions. Now, this notion of meta-inferential validity coincides with STST above.[9]

Proposition 18. *A meta-inference of level 1, $\Gamma \Rightarrow \Delta$, is valid according to $\models_{TM}$ iff it is valid according to* STST.

Proof. The proposition follows from Lemma 14. For, the lemma implies that there is a strong Kleene valuation that is a counterexample to all inferences in Δ but not to any inference in Γ iff there is a state in a truth-maker model that is also a counterexample to all inferences in Δ but not to any inference in Γ. ∎

Note that nothing about this definition of meta-inferential validity suggests that we treat inferences and meta-inferences alike. For, a state verifying a sentence and a state satisfying an inference differ markedly. Let me bring this out more explicitly. A state makes a sentence true by being the worldly item on which the truth of the sentence turns and being such that the truth of the sentence is settled in the positive. By contrast, a state satisfies an inference merely by not being a possible state that includes a verifier for each premise and a falsifier for each conclusion. A state that is entirely foreign to the subject matter of all the sentences in an inference will satisfy the inference.

Moreover, the modal aspect of the validity of inferences is captured, in my truth-maker semantics, by the fact that all the states that are counterexamples to a valid inference are impossible. Nothing like that is going on in the case of meta-inferences. Insofar as meta-inferential validity has a modal aspect, it is captured by quantifying over states, some of which

[9] As is well known, there is a translation between LP and STST (see Barrio et al., 2015; Dicher and Paoli, 2019). Hence, the same translation also maps truth-maker meta-inferential validity, as just defined, into LP. I don't think that this shows that ST is "really" LP, but I won't engage this debate here.

may be counterexamples. The two notions of validity look *prima facie* very different. Hence, according to my truth-maker semantics, it is *prima facie* at least not implausible that staying close to classical logic may be a virtue in our treatment of inferential validity while not being a virtue in our treatment of meta-inferential validity.

With these differences firmly in mind, let's consider how one might climb the strict/tolerant hierarchy from the perspective of truth-maker semantics. In order to climb the hierarchy, we have to define TSST at level 1. In order to do this, we first need to define TS in terms of the same models we used to define ST.[10] This can be done in the following way:

Definition 19. *Truth-maker validity for* TS ($\models_{TM.TS}$): $\Gamma \models_{TM.TS} \Delta$ iff, in every model, for all $s \in S^{\Diamond}$, if $\neg\exists u \sqsubseteq s$, $\exists \gamma \in \Gamma$ such that $u \dashv\!\| \gamma$, then $\exists u \sqsubseteq s$,$\exists \delta \in \Delta$ such that $u \|\!\vdash \delta$.

Proposition 20. *If we impose Downward-Closure and Exclusivity but not Exhaustivity, then* $\Gamma \models_{TM.TS} \Delta$ *iff* $\Gamma \models_{TS} \Delta$.

Proof. It suffices to adjust the proof of Lemma 14 above in appropriate ways. Recall that $\Gamma \models_{TS} \Delta$ iff, for every strong Kleene valuation, v, such that $\forall \gamma \in \Gamma$ $(v(\gamma) \in \{\frac{1}{2}, 1\})$ there is a $\delta \in \Delta$ such that $v(\delta) = 1$. Given Lemma 13, this holds just in case for every possible state in all truth-maker models, if the state does not inexactly falsify any $\gamma \in \Gamma$, then it inexactly verifies some $\delta \in \Delta$. ■

Note that, in our truth-maker framework, TS-validity doesn't seem pre-theoretically well motivated. The definition says that an inference is TS-valid iff all possible states, in all models, that don't make any of the premises false (inexactly) make at least one conclusion true (inexactly).

[10] I show elsewhere that one way to define TS is to reject Exclusivity while holding on to Exhaustivity (Hlobil, 2022). But this won't help with our current problem because we want to use the same space of models for both logics, just as we do in the strong Kleene semantics.

For example, a state that is entirely foreign to the inference, in the sense that it neither verifies nor falsifies any of the premises or conclusions, is a counterexample to the inference according to TS-validity.

By contrast, we have seen above that the notion of truth-maker validity that gives us ST is a version of the independently motivated idea of modalism about consequence. As far as I can see, there is no similarly nice thing that we can say about TS-validity in our truth-maker theory. It is, of course, a perfectly well-behaved mathematical relation. However, it seems very difficult to find any independent motivation for the view that it captures anything that we could plausibly mean, in a pre-theoretical sense, by "consequence." After all, once we accept the idea that some states make some sentences neither true nor false, could we reasonably object, e.g., to the inference from "It is raining" to "It is not the case that it is not raining" by appeal to an arbitrary state that neither verifies nor falsifies that it is raining, such as the state of seven being prime? That doesn't seem pre-theoretically plausible. Thus, the truth-maker version of TS-validity seems pre-theoretically less plausible and less motivated than the truth-maker version of ST-validity.

Once we have a definition of TS over the same truth-maker models as ST, we can define TSST in the usual way.

Definition 21. Truth-Maker-TSST: A meta-inference of level 1, $\Gamma \Rightarrow \Delta$, is valid according to Truth-Maker-TSST iff, for all states s in all truth-maker models, if s is a counterexample to all $\delta \in \Delta$ according to $\models_{TM}$, then it is a counterexample to at least one $\gamma \in \Gamma$ according to $\models_{TM.TS}$.

This gives us the second step in the strict/tolerant hierarchy within the truth-maker framework, i.e., TSST. It is now easy to define Truth-Maker-STTS by flipping the two standards:

Definition 22. Truth-Maker-STTS: A meta-inference of level 1, $\Gamma \Rightarrow \Delta$, is valid according to Truth-Maker-STTS iff, for all states s in all truth-maker models, if s is a counterexample to all $\delta \in \Delta$ according to $\models_{TM.TS}$, then it is a counterexample to at least one $\gamma \in \Gamma$ according to $\models_{TM}$.

It is now easy to see how we can generalize this step. Thus, we can climb the hierarchy by defining Truth-Maker-ST^n and Truth-Maker-TS^n as follows.

Definition 23. Truth-Maker-ST^n: A meta-inference of level n, $\Gamma \Rightarrow \Delta$, is valid according to Truth-Maker-ST^n iff, for all states s in all truth-maker models, if s is a counterexample to all $\delta \in \Delta$ according to Truth-Maker-ST^{n-1}, then it is a counterexample to at least one $\gamma \in \Gamma$ according to Truth-Maker-TS^{n-1}.

Truth-Maker-TS^n: A meta-inference of level n, $\Gamma \Rightarrow \Delta$, is valid according to Truth-Maker-TS^n iff, for all states s in all truth-maker models, if s is a counterexample to all $\delta \in \Delta$ according to Truth-Maker-TS^{n-1}, then it is a counterexample to at least one $\gamma \in \Gamma$ according to Truth-Maker-ST^{n-1}.

As we did in the Kleene semantics, we can say that an inference is valid according to Truth-Maker-ST^ω if, for some n, it is valid according to Truth-Maker-ST^n. This yields that truth-Maker-ST^ω coincides with ST^ω above. This is what it looks like, from the perspective of truth-maker theory, to climb the strict/tolerant hierarchy.

Let's take stock. I have presented a truth-maker semantics for ST. The key idea behind that semantics is that what it means for an inference to be valid is that any state that combines truth-makers for all the premises and falsity-makers for all the conclusions is an impossible state. That idea is a variant of modalism about consequence, which has been advocated on independent grounds. Moreover, we have seen that this semantics treats sentences and inferences very differently: sentences are made true or false by states, whereas the validity of inferences brings in modality. We can define the strict/tolerant hierarchy in this truth-maker semantics. As I will argue in the next section, however, climbing the hierarchy doesn't look very attractive or illuminating from the perspective of truth-maker semantics.

4 Philosophical Upshot

In this section, I turn to my second aim, namely to cast some doubt on climbing the strict/tolerant hierarchy. More precisely, I want to argue that the truth-maker semantics above suggests a philosophical interpretation of consequence and meta-inferential consequence at level 1 that doesn't support the idea that staying close to classical logic is a virtue in meta-inferential consequence relations. The overall argument of my second project may be summarized as follows:

(P1) Which properties of meta-inferential consequence relations we should deem theoretical virtues depends on what we are modeling or theorizing by such meta-inferential consequence relations.

(P2) The truth-maker semantics above makes a suggestion about what we are modeling or theorizing by meta-inferential consequence relations, and it doesn't suggest that staying close to classical consequence is a virtue in such meta-inferential consequence relations.

(P3) The suggestion that comes out of truth-maker semantics is no worse than any of the available alternatives.

(P4) If (P1)–(P3) are true, then we should have some doubts about climbing the strict/tolerant hierarchy.

(C) Therefore, we should have some doubts about climbing the strict/tolerant hierarchy.

Notice that my conclusion is a rather weak claim. It is compatible, e.g., with these doubts being ultimately overruled by other considerations. Note also that I don't claim that the philosophical understanding of meta-inferential validity suggested by truth-maker semantics is clearly the correct or uniquely best one. I merely claim that it is no worse than its competitors. Moreover, I focus on just one potential theoretical virtue of meta-inferential consequence relations, namely closeness to classical logic. My unwillingness to make stronger claims reflects my conviction that the currently available philosophical research on the nature and role of logic, and especially the nature and role of meta-inferential consequence relations, doesn't allow for more definite and stronger pronouncements.

I will start by saying something about (P1) and (P4). I will then devote separate subsections to (P2) and (P3).

I have, in effect, already presented some reasons for accepting (P1) in Subsection 2.3 above. In general, what is a virtue in a formal theory depends, at least in part, on the point or purpose of the formal theory, and this point or purpose crucially involves what phenomenon we are trying to model or theorize by our formal theory. The question doesn't arise for formal theories considered merely as bits of pure mathematics, which don't model or theorize anything beyond their mathematical content. But when we wonder whether we should adopt a logical theory as (part of) our response to the truth-theoretic paradoxes, we don't consider the logical theory merely as a bit of pure mathematics.

In the case at hand, it is clear that the strict/tolerant hierarchy is an interesting mathematical structure. What is at issue is whether we should accept ST^{ω} as a response to the truth-theoretic paradoxes. Now, how plausible ST^{ω} is as a response to the paradoxes will depend on what we take ST^{ω} to tell us about the paradoxes at a conceptual or philosophical level. It will be plausible if, in accepting it, we attribute pre-theoretically plausible or illuminating properties to the relation that is modeled or theorized by the consequence relation of ST^{ω}, including meta-inferential consequence relations at all finite levels. In particular, whether closeness to classical logic should count as a virtue in ST^{ω} depends on this relation between the target phenomena that we are trying to understand and the logical theory.

Staying close to classical logic is not always a virtue in a logical theory. In formulating the Lambek calculus, e.g., we are trying to model or theorize expressions of which grammatical types are formed by concatenating expressions of certain grammatical types. And in light of this target phenomenon, closeness to classical logic isn't a virtue in a theory like the Lambek calculus. Since the type of expression depends on the order of the concatenated expressions, e.g., the structural rule of permutation isn't a virtue in such a theory. I cannot see any reason to think that this dependence of what counts as a virtue on the target phenomenon should fail for meta-inferential consequence relations.

Turning to (P4), we should first note that formal theories are not only assessed in light of their target phenomena, but that they can also make a potential target phenomenon salient and thus suggest a target phenomenon. A game-semantics for different logics may suggest, e.g., that the target phenomena of these logics have to do with broadly dialogical situations (see Dutilh Novaes, 2020). Moreover, if there are equally good rival views on what the target phenomenon of a logical theory is, then we should not arbitrate the dispute by appeal to alleged theoretical virtues that advocates of one of the rival views shouldn't deem theoretical virtues (in light of their view). Hence, in such a case, we should have doubts about any such rival view whose chief argument rests on an appeal to such a controversial theoretical virtue. But if the rest of my premises are true, then this is our situation with respect to climbing the strict/tolerant hierarchy. Nothing in the truth-maker semantics suggests that the target phenomenon of meta-inferential consequence relations is such that closeness to classical logic is a theoretical virtue. Indeed, nothing in this semantics suggests that inferential and meta-inferential consequence relations concern a shared target phenomenon; and what the target phenomenon of meta-inferential consequence relations is and, hence, what should count as virtues in such consequence relations is very much an open question, from the perspective of truth-maker semantics. But the chief argument of the advocate of ST^{ω} rests on an appeal to this closeness to classical logic being a theoretical virtue in meta-inferential consequence relations. So, if my other premises are true, we should have doubts about climbing the strict/tolerant hierarchy, which is what (P4) says.

Since I already defended (P1), it remains for me to argue for (P2) and (P3). I will take these two premises in turn.

4.1 Truth-Maker Semantics and Classicality

I claim, in (P2), that my truth-maker semantics suggests a way to think about the target phenomenon of logics that address the truth-theoretic paradoxes in such a way that closeness to classical logic doesn't appear to be a theoretical virtue in meta-inferential consequence relations.

As already intimated, the truth-maker semantics that I have presented above suggests a pre-theoretical understanding of inferential validity that is a version of modalism about consequence (Bueno and Shalkowski, 2013). What it means for an inference to be valid is for the premises to be such that any worldly state that makes all the premises true is such that it cannot possibly make all the conclusions jointly false. And in logically valid inferences this holds in virtue of the meanings of the logical connectives, which is why we quantifier over all models. Thus the target phenomenon that we are modeling or theorizing by inferential consequence relations is the relation in which the truth of the premises render the falsity of (all the) conclusions impossible (and do so in virtue of their logical form, if we talk about specifically logical inferences).

Moreover, the truth-maker semantics suggests a pre-theoretical understanding of meta-inferential validity of level 1 as Truth-Maker-STST: What it is for a meta-inference of level 1 to be valid is that if a state isn't a counterexample (i.e., a possible state that includes truth-makers for all the premises and falsity-makers for all the conclusions) to any of the premise-inferences, then it isn't a counterexample to at least one conclusion-inference. This is a useful relation because we can track it, e.g., in sequent calculi in an easy and straightforward way (see Ré et al. (2021) for the relation of derivable rules and local meta-inferential validity). While this relation can be understood in a pre-theoretical way, its interest derives from its use in theorizing inferential consequence.

The idea is that it is only inferential consequence that is bound up with our "mathematical, physical, and other scientific theories" (Barrio et al., 2021, sec. 1). In pursuing these theories we aim at understanding the natures of the phenomena that we thereby theorize, and that means that we come to understand what is possible and impossible in the domain of these phenomena. Classical logic can serve as a lower bound on what is impossible in any such domain, namely the combination of the truth of premises and the falsity of conclusions that follow classically from the premises. This doesn't require that we rule out that there are sentences such that their truth-makers and their falsity-makers are both incompatible with any possible state. And it is a virtue and not a vice if this phenomenon comes out clearly in the relation among inferences that we logical theorists use, e.g., in sequent calculi.

Note that if someone asks what it means for a meta-inference of level 2 to be valid, there is no suggestion forthcoming from the truth-maker semantics. That strikes me as a feature and not a bug. The truth-maker semantics can capture all kinds of complicated relations among inferences in virtue of which states are counterexamples for them (as we can see in the definition of the strict/tolerant hierarchy in truth-maker semantics), but nothing in the theory forces us to choose one of these relations and call it "*the* meta-inferential consequence relation at level 2."

The ST^{ω}-theorist can define the strict/tolerant hierarchy in my truth-maker semantics. But we should immediately ask, why we should do this. That is, why should we call the thus defined relations the "meta-inferential consequence relations" for each level? The ST^{ω}-theorist answers that this allows us to stay close to classical logic. But why should this move us? For note what the ST^{ω}-theorist is clearly not offering. First, the ST^{ω}-theorist is not offering a relation among (meta-)inferences that is useful in theorizing inferential consequence, e.g., by being particularly suitable for codification in a sequent calculus or the like. Second, the ST^{ω}-theorist is not offering us a relation among (meta-)inferences that is of the same kind as the inferential consequence relation at the ground level. To do the latter, the ST^{ω}-theorist would have to define a relation, at level 1, such that the relation holds iff any fusion of states that make the premise-inferences valid and make the conclusion-inferences invalid is an impossible state. But to do that, the ST^{ω}-theorist would have to give us an account of which states make which inferences valid or invalid. Now, the question which worldly states (if any) make an inference valid or invalid is a big question in the philosophy of logic, and I think the ST^{ω}-theorist is well advised not to make any answer to it part of her core commitments. Hence, I think that once we adopt the truth-maker semantics above, the ST^{ω}-theorist shouldn't even be tempted to offer a relation at the meta-inferential level that is of the same kind as inferential consequence.

If what I just said is correct, then the ST^{ω}-theorist's insistence on closeness to classical logic cannot be justified by appeals to theoretical usefulness or continuity between the consequence relations of different levels. I cannot see any other plausible motivation for treating closeness to classical logic as a virtue in meta-inferential consequence relations.

If what I said in this section is right, then the three-valued semantics for ST may be misleading. It suggests that we should have views about meta-inferential validity at arbitrarily high levels and that such meta-inferential consequence relations are continuous with the consequence relation at the ground level. But this suggestion now looks like an artefact of the strong Kleene semantics, while the truth-maker semantics suggests that meta-inferential consequence differs markedly from inferential consequence in its role and nature—not just technically but in their relation to the target phenomena. Once we move to truth-maker semantics, the question what we think about meta-inferential validity in a way that is continuous with inferential consequence is revealed to concern what worldly states make it the case that inferences are valid or invalid. It seems to me that this is a huge philosophical issue on which the advocate of ST can and should (absent independent considerations) remain silent. This concludes my defense of (P2).

4.2 No Better Alternative

What I said in the previous subsection wouldn't be a problem for the ST^ω-theorist if she could easily dismiss the suggestion about what we model or theorize by consequence relations at different levels that comes out of the truth-maker semantics. But in order to dismiss this suggestion, the ST^ω-theorist would need a superior alternative view about what the target phenomenon of meta-inferential consequence relations is. In this subsection, I will argue that there is no such superior alternative and, hence, (P3) is true.

Sometimes Barrio and his collaborators talk as if closeness to classical logic is a virtue in meta-inferential consequence relations independently of the phenomenon we are trying to model or theorize by such meta-inferential consequence relations, and I have already explained why I think that is wrong. However, Barrio and his collaborators sometimes offer a philosophical account of the target phenomenon. In particular, they sometimes suggest that a philosophical account broadly similar to the one offered by normative bilateralism (e.g. by Restall and Ripley) for inferential validity can be extended to meta-inferential validity. Thus, Barrio, Pailos, and Calderón (2021, Sec 4.3) suggest that we interpret meta-inferences—as we move up the hierarchy—as more and more complicated claims about which patterns of assertions and denials are not ruled out by coherence norms.[11] They write:

> For $\Gamma_1 \models \Delta_1, ..., \Gamma_n \models \Delta_n \Longrightarrow \Sigma_1 \models \Pi_1, ..., \Sigma_j \models \Pi_j$ to be valid in TS/ST [i.e. TSST] means that, for every possible epistemic scenario—represented by valuations—if, for every $\Gamma_i \models \Delta_i$, either some $\gamma \in \Gamma_i$ is not accepted (i.e., does not get value 1 in the valuation) or some $\delta \in \Delta'$[12] is not rejected (i.e., does not get value 0 in the valuation), then some $\Sigma_i \models \Pi_i$ is such that either some $\sigma \in \Sigma'$ is not accepted (i.e., does not get value 1 in the valuation) or some $\pi \in \Pi'$ is not rejected (i.e., does not get value 0 in the valuation). (Barrio et al., 2021, Sec 4.3)

The idea here seems to be that the target phenomenon that we are modeling or theorizing by meta-inferential consequence relations are complex patterns of acceptances and rejections that would be correct or adequate in possible epistemic scenarios. And I take this to mean acceptances and rejections that an ideally rational agent could adopt. Is this a philosophical conception of meta-inferential validity that is superior to the one suggested by truth-maker semantics?

First of all, Barrio and collaborators commit a mistake in the passage above: What they are describing here is really STST and not TSST, as they intend to do. They should write: "... if, for every $\Gamma_i \models \Delta_i$, either some $\gamma \in \Gamma_i$ is rejected (i.e., gets value 0 in the valuation) or some $\delta \in \Delta'$ is accepted (i.e., gets value 1 in the valuation), then ..." And that is how

[11] Barrio and Pailos (2022, 94) hint at a similar philosophical interpretation of the hierarchy in terms of weak and strong acceptance and rejection.

[12] I take it that Δ' should be Δ_i, and similarly for the other "primed" uppercase Greek letters in this quote.

they present the situation one paragraph later, when they discuss the liar sentence.

> If both λ is accepted—i.e., gets value 1—and rejected—i.e., gets value 0—, then either some premise of the empty inference is not accepted—i.e., does not get value 1—or some conclusion of the empty inference is not rejected—i.e., it does not get value 0. And this is the case—i.e., this conditional is (classically) true—just because it is not the case that the Liar sentence is either [sic] accepted and rejected—in fact, as it is neither accepted nor rejected, both conjuncts are false, and therefore, the antecedent is false. (Barrio et al., 2021, Sec 4.3)

The mistake in the earlier quote makes it seem as if TSST-consequence tracks the preservation of a status of inferences, as does STST, but that is of course not the case. In TSST, the patterns of acceptance and rejection that count as a counterexample for inferences on the left are not those that count as a counterexample for inferences on the right. As explained above, that is why the STST relation is easier to codify, e.g., in a calculus. So, to correct the mistake in the quoted passage, let's be clear that TSST picks out one fairly complex relation among patterns of acceptances and rejections of ideally rational agents among a multitude of such relations, which are all already fixed by our valuations, and the relation isn't a particularly easy one for theorists to work with, as, in contrast to STST, it isn't a relation of preservation of a particular status of inferences relative to valuations or the like.

We can define a multitude of relations among inferences in terms of strong Kleene valuations. All of these relations capture facts about "complex epistemic relationships" (Barrio et al., 2021, Sec 4.3), as modeled by strong Kleene valuations. Why does the TSST-relation, among all of them, deserve the label "the consequence relation at level 1"? The only answer that advocates of the Buenos Aires Plan offer is that it deserves the label because it allows us to accept the classical structural rule of Cut. But that simply isn't a responsive answer when we are wondering whether closeness to classical logic is a virtue in meta-inferential consequence relations.

Furthermore, it is unclear why we should be interested in the more and more complex patterns of epistemic relationships that are modeled by meta-inferential consequence relations of higher and higher levels, according to this view. Introducing these more complicated relations doesn't, e.g., seem to increase our expressive power. For each epistemic scenario, modeled by a strong Kleene valuation, we can already say which combinations of acceptances and rejections are possible for ideally rational agents. We can, e.g., formulate an inferential sequent that codifies the fact that an ideally rational agent cannot accept and reject the liar sentence, namely $\lambda \models \lambda$. And if that is all that is conveyed by the validity of $\{\emptyset \overset{0}{\Rightarrow} \lambda, \lambda \overset{0}{\Rightarrow} \emptyset\} \overset{1}{\Rightarrow} \{\emptyset \overset{0}{\Rightarrow} \emptyset\}$, as Barrio and his collaborators suggest in the quote above, then we may wonder why it is so important to call TSST the "consequence relation" at level 1.

To sum up, the suggestion that the target phenomenon that we are trying to model or theorize by meta-inferential consequence relations are complex patterns of acceptances and rejections of ideally rational agents isn't superior to the suggestion that comes out of the truth-maker semantics. For, (a) the suggestion doesn't provide my opponents with any reason to choose one of the many candidate meta-inferential relations as the consequence relation at a given level. And (b) the suggestion doesn't explain why we should be interested in all of these complex patterns or why closeness to classical logic is a virtue in a theory of these patterns.

5 Conclusion

Let's take stock. I have presented a truth-maker semantics for the logic ST that is an alternative to the usual three-valued semantics. The notion of (local) meta-inferential validity for level 1 that immediately suggests itself on this view is STST. All meta-inferential consequence relations that are defined in terms of the models that define inferential consequence appear markedly different from inferential consequence in their technical definition and philosophical significance. The meta-inferential consequence relation STST isn't privileged by its relation to the target phenomenon of our logical theory; rather, it is merely a nice and useful relation to work with for us theorists. The consequence relations in the strict/tolerant hierarchy are also not privileged by their relation to any target phenomenon, but they are not even particularly nice or useful for the work of us theorists. In general, the philosophical significance of meta-inferential consequence relations is limited to their significance for theorists; they don't have any clear philosophical significance with respect to the target phenomenon that we are trying to understand in our logical theories. At least, this is the view that is suggested by the truth-maker semantics that I have presented here.

Acknowledgements

For invaluable comments and discussion, I would like to thank Lucas Rosenblatt, Robert Brandom, Daniel Kaplan, Ryan Simonelli, Rea Golan, Shuhei Shimamura, Viviane Fairbank, and audiences at the University of Connecticut and at the Tenth Workshop on Philosophical Logic at IIF-SADAF-CONICET. Work on this paper was supported by the EXPRO grant No. 20-05180X of the Czech Science Foundation.

Statements and Declarations

Work on this paper was supported by the EXPRO grant No. 20-05180X of the Czech Science Foundation. There author has no conflict of interests of any kind.

References

Armstrong, D. M. (2004). *Truth and Truthmakers*. Cambridge University Press, Cambridge.

Barrio, E. and Pailos, F. (2022). Validities, antivalidities and contingencies: A multi-standard approach. *Journal of Philosophical Logic*, 51(1):75–98.

Barrio, E. A., Pailos, F., and Calderón, J. T. (2021). Anti-exceptionalism, truth and the BA-plan. *Synthese*, 199(5-6):12561–12586.

Barrio, E. A., Pailos, F., and Szmuc, D. (2019a). A hierarchy of classical and paraconsistent logics. *Journal of Philosophical Logic*, 49(1):1–28.

Barrio, E. A., Pailos, F., and Szmuc, D. (2019b). (Meta)inferential levels of entailment beyond the Tarskian paradigm. *Synthese*, 198(S22):5265–5289.

Barrio, E. A., Rosenblatt, L., and Tajer, D. (2015). The logics of strict-tolerant logic. *Journal of Philosophical Logic*, 44(5):551–571.

Bueno, O. and Shalkowski, S. A. (2013). Logical constants: A modalist approach. *Noûs*, 47(1):1–24.

Cobreros, P., Egré, P., Ripley, D., and van Rooij, R. (2012). Tolerant, classical, strict. *Journal of Philosophical Logic*, 41(2):347–385.

Cobreros, P., Egré, P., Ripley, D., and van Rooij, R. (2013). Reaching transparent truth. *Mind*, 122(488):841–866.

Cobreros, P., Egré, P., Ripley, D., and van Rooij, R. (2020a). Inferences and metainferences in ST. *Journal of Philosophical Logic*, 49(6):1057–1077.

Cobreros, P., Rosa, E. L., and Tranchini, L. (2020b). (I can't get no) antisatisfaction. *Synthese*, 198(9):8251–8265.

Cobreros, P., Rosa, E. L., and Tranchini, L. (2021). Higher-level inferences in the strong-Kleene setting: A proof-theoretic approach. *Journal of Philosophical Logic*, pages 1–36.

Da Ré, B., Pailos, F., Szmuc, D., and Teijeiro, P. (2020). Metainferential duality. *Journal of Applied Non-Classical Logics*, 30(4):312–334.

Dicher, B. and Paoli, F. (2019). ST, LP and tolerant metainferences. In Bakent, C. and Ferguson, T. M., editors, *Graham Priest on Dialetheism and Paraconsistency*, pages 383–407. Springer Verlag.

Dutilh Novaes, C. (2020). *The Dialogical Roots of Deduction: Historical, Cognitive, and Philosophical Perspectives on Reasoning*. Cambridge University Press, Cambridge.

Fine, K. (2014). Truth-maker semantics for intuitionistic logic. *Journal of Philosophical Logic*, 43(2-3):549–577.

Fine, K. (2016). Angellic content. *Journal of Philosophical Logic*, 45(2):199–226.

Fine, K. (2017a). A theory of truthmaker content I: Conjunction, disjunction and negation. *Journal of Philosophical Logic*, 46(6):625–674.

Fine, K. (2017b). A theory of truthmaker content II: Subject-matter, common content, remainder and ground. *Journal of Philosophical Logic*, 46(6):675–702.

Fine, K. and Jago, M. (2018). Logic for exact entailment. *Review of Symbolic Logic*, pages 1–21.

Fitting, M. (2021a). A family of strict/tolerant logics. *Journal of Philosophical Logic*, 50(2):363–394.

Fitting, M. (2021b). Strict/tolerant logics built using generalized weak kleene logics. *Australasian Journal of Logic*, 18(2).

French, R. (2016). Structural reflexivity and the paradoxes of self-reference. *Ergo: An Open Access Journal of Philosophy*, 3.

Hlobil, U. (2022). The laws of thought and the laws of truth as two sides of one coin. *Journal of Philosophical Logic*, forthcoming.

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, 72(19):690–716.

O'Conaill, D. and Tahko, T. E. (2016). Minimal truthmakers. *Pacific Philosophical Quarterly*, 97(2):228–244.

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

Ré, B. D., Szmuc, D., and Teijeiro, P. (2021). Derivability and metainferential validity. *Journal of Philosophical Logic*, forthcoming.

Ripley, D. (2012). Conservatively extending classical logic with transparent truth. *Review of Symbolic Logic*, 5(2):354–378.

Ripley, D. (2013). Paradoxes and failures of Cut. *Australasian Journal of Philosophy*, 91(1):139–164.

Ripley, D. (2015). Anything goes. *Topoi*, 34(1):25–36.

Ripley, D. (2021). One step is enough. *Journal of Philosophical Logic*, forthcoming:1–27.

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

Scambler, C. (2020b). Transfinite meta-inferences. *Journal of Philosophical Logic*, 49(6):1079–1089.