THE REVIEW OF SYMBOLIC LOGIC
Volume 13, Number 1, March 2020

A RECOVERY OPERATOR FOR NONTRANSITIVE APPROACHES

EDUARDO ALEJANDRO BARRIO, FEDERICO PAILOS
and
DAMIAN SZMUC
University of Buenos Aires and IIF-SADAF (CONICET)

Abstract. In some recent articles, Cobreros, Egré, Ripley, & van Rooij have defended the idea that abandoning transitivity may lead to a solution to the trouble caused by semantic paradoxes. For that purpose, they develop the Strict-Tolerant approach, which leads them to entertain a nontransitive theory of truth, where the structural rule of Cut is not generally valid. However, that Cut fails in general in the target theory of truth does not mean that there are not certain safe instances of Cut involving semantic notions. In this article we intend to meet the challenge of answering how to regain all the safe instances of Cut, in the language of the theory, making essential use of a unary recovery operator. To fulfill this goal, we will work within the so-called Goodship Project, which suggests that in order to have nontrivial naïve theories it is sufficient to formulate the corresponding self-referential sentences with suitable biconditionals. Nevertheless, a secondary aim of this article is to propose a novel way to carry this project out, showing that the biconditionals in question can be totally classical. In the context of this article, these biconditionals will be essentially used in expressing the self-referential sentences and, thus, as a collateral result of our work we will prove that none of the recoveries expected of the target theory can be nontrivially achieved if self-reference is expressed through identities.

§1. Introduction. In some recent articles, Cobreros, Egré, Ripley, & van Rooij have defended the idea that abandoning transitivity may lead to a solution to the trouble caused by semantic paradoxes. For that purpose, they develop the Strict-Tolerant approach, which leads them to consider the logic **ST** used to later build a nontransitive theory of truth which we will hereinafter call **STTT**.[1] Throughout the authors' works, the nontransitive project has proved to have many attractive features. One of the most fundamental is something that might come as a surprise, namely that it gives up Cut without thereby abandoning Classical Logic, for the championed nontransitive theory of truth and Classical Logic coincide at the inferential level—that is, they have the same set of valid inferences.

Yet, Cut needs to be given up to nontrivially handle a naïve truth predicate: it is precisely this failure that saves the day when we attempt to reproduce a derivation of, e.g., the Liar

Received: September 23, 2017.
2010 *Mathematics Subject Classification*: 03B05, 03B47, 03B53.
Key words and phrases: substructural logics, cut rule, recovery operator, paradoxes.

[1] Nontransitive approaches to logical consequence were discussed, previously, in many works to which the authors refer in their articles. Some of these are due to Strawson (as referred in [43] and [39], and linked to **ST** for the first time in the recent article [13]), Tennant [40, 41], Weir [44], Cook [18] and Frankowski [23]. But it should be highlighted that the application of this logical approach to semantics paradoxical phenomena is original of Cobreros, Egré, Ripley, & van Rooij—although the approach applied by Zardini in [45] to deal with the sorites paradox is nontransitive, too.

© Association for Symbolic Logic, 2018

doi:10.1017/S1755020318000369

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

paradox. However, that Cut does not apply with full generality does not mean that there are no safe instances of Cut. This observation is crucial for many reasons, e.g., given the natural link Cut has with the transitivity of deduction and the decidability of certain systems. Thus, it will be interesting to highlight those places where it can be safely employed. We might want to get as much Cut as we can, even if we cannot get it all. The aim of this article is to meet the challenge of answering how to recover the safe instances of Cut, within the language of the theory itself.

To do this, we will expand the language of the target nontransitive theory of truth with a unary recovery operator, that will allow us to regain not only these safe instances of Cut, but also to recover other things that we may have lost when adopting **STTT**. Now, some may worry that introducing enough expressive resources to recover the safe instances of Cut will inevitably result in potential revenge paradoxes. We will take care of these worries with a great deal of detail, showing how the specifics of the (non)triviality results depend on the way self-reference is achieved. To this extent, we will be following Laura Goodship's recommendation that in order to have fully expressive nontrivial naïve theories (e.g., naïve theories of truth) it is sufficient to express these matters with a suitable conditional. However, we will diverge from Goodship in a relevant respect. While she suggested that for the conditional to be suitable it should invalidate either Modus Ponens or contraction—thus being sub-classical—we will show that it is possible to express self-reference with a completely classical conditional, if the underlying logic is nontransitive. As a collateral result of our work, we will prove that none of the recoveries expected of the target theory can be nontrivially achieved if self-reference is expressed through identities, instead of through equivalences.

The article is structured as follows. In §2 we present the logic **ST** and the nontransitive theory of truth **STTT**. In §3 we outline our proposal to recover the safe instances of Cut, along with the question of whether or not the project leads to triviality. In §4 we show that if self-referential sentences are expressed through equivalences the project can be nontrivially entertained, constructing furthermore a nontrivial valuation for the target theory, discussing its connection with the Goodship Project and drawing some philosophical reflections thereby. In §5 we provide a sound and complete three-sided sequent calculus for our system. Finally, in §6 we provide some concluding remarks.

§2. A nontransitive transparent theory of truth. In what follows we will consider the nontransitive *theory of truth* **STTT**. But, first, we need to understand its underlying *logic*, **ST**. It is usually presented as a first order logic, but for the sake of simplicity, we will deal with a propositional version of these systems and their expansions.[2]

To move forward, it will be helpful to fix some definitions first. If $\mathcal{L}$ is a propositional language, then with $\mathbf{FOR}(\mathcal{L})$ we will denote the absolutely free algebra of $\mathcal{L}$, whose universe is $FOR(\mathcal{L})$. As usual, we will let Γ, Δ, and other Greek capital letters represent sets of formulae, or sets of inferences, and Roman capital letters A, B, C represent formulae themselves, or inferences themselves.[3]

[2] For an extensive presentation of **ST**, see [14, 16, 36], and [37].

[3] Though we hope the context will make things clear enough, we will always clearly state whether, for example, Roman capital letters A, B, C represent formulae themselves, or inferences themselves.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

The logic **ST** can be presented as a *p*-logic, as devised by Frankowski in [23] as a means to characterize logical systems where valid derivations are such that the degree of strength of the conclusions can be smaller than that of the premises.[4]

DEFINITION 2.1 ([23]). *A p-consequence relation over a propositional language $\mathcal{L}$ is a relation $\vDash \subseteq \wp(FOR(\mathcal{L})) \times \wp(FOR(\mathcal{L}))$ obeying the following conditions for all $\Gamma, \Delta \subseteq FOR(\mathcal{L})$:*

1. $\Gamma \vDash \Delta$ *if* $\Gamma \cap \Delta \neq \emptyset$ (Reflexivity)
2. *If* $\Gamma \vDash \Delta$, *then* $\Gamma, \Gamma' \vDash \Delta, \Delta'$. (Monotonicity)

Additionally, a p-consequence relation $\vDash$ is substitution-invariant whenever if $\Gamma \vDash \Delta$, and σ is a substitution on $FOR(\mathcal{L})$, then $\{\sigma(B) \mid B \in \Gamma\} \vDash \{\sigma(A) \mid A \in \Delta\}$.

DEFINITION 2.2 ([23]). *A p-logic over a propositional language $\mathcal{L}$ is an ordered pair $(FOR(\mathcal{L}), \vDash)$, where $\vDash$ is a substitution-invariant p-consequence relation.*

In general, *p*-logics can be associated with *p*-matrices as follows, with the help of valuation functions, i.e., homomorphisms from $FOR(\mathcal{L})$ to the set of truth-values of the *p*-matrix in question.

DEFINITION 2.3 ([24]). *For $\mathcal{L}$ a propositional language, an $\mathcal{L}$-p-matrix is a structure $\langle \mathcal{V}, \mathcal{D}^+, \mathcal{D}^-, \mathcal{O} \rangle$, such that $\langle \mathcal{V}, \mathcal{O} \rangle$ is an algebra of the same similarity type as $\mathcal{L}$, with universe $\mathcal{V}$ and a set of operations $\mathcal{O}$, where $\mathcal{D}^+, \mathcal{D}^- \subseteq \mathcal{V}$ and $\mathcal{D}^+ \subseteq \mathcal{D}^-$.*

DEFINITION 2.4 ([21]). *An inference or sequent on $\mathcal{L}$, written $\Gamma \Rightarrow \Delta$, is an ordered pair (Γ, Δ) where $\Gamma, \Delta \subseteq FOR(\mathcal{L})$ are finite and possibly empty. $SEQ(\mathcal{L})$ is the set of all sequents on $\mathcal{L}$.*

DEFINITION 2.5. *For $\mathcal{M}$ a p-matrix, an $\mathcal{M}$-valuation v satisfies a sequent or inference $\Gamma \Rightarrow \Delta$, written $v \vDash_{\mathcal{M}} \Gamma \Rightarrow \Delta$, by letting*

$$v \vDash_{\mathcal{M}} \Gamma \Rightarrow \Delta \quad \text{iff} \quad \text{if } v(B) \in \mathcal{D}^+ \text{ for all } B \in \Gamma, \text{ then } v(A) \in \mathcal{D}^- \text{ for some } A \in \Delta.$$

A sequent or inference $\Gamma \Rightarrow \Delta$ is $\mathcal{M}$-valid, written $\vDash_{\mathcal{M}} \Gamma \Rightarrow \Delta$, by letting

$$\vDash_{\mathcal{M}} \Gamma \Rightarrow \Delta \quad \text{iff} \quad \text{for all } \mathcal{M}\text{-valuations } v,\ v \vDash_{\mathcal{M}} \Gamma \Rightarrow \Delta.$$

Notice, then, that we may interchangeably refer to an inference $\Gamma \Rightarrow \Delta$ that is valid in the logic **L** induced by the matrix $\mathcal{M}$, as $\Gamma \vDash_{\mathcal{M}} A$ or $\vDash_{\mathcal{M}} \Gamma \Rightarrow \Delta$. Moreover, when the matrix $\mathcal{M}$ is essentially associated with the logic **L** we may alternatively denote $\vDash_{\mathcal{M}}$ as $\vDash_{\mathbf{L}}$.

Furthermore, given the previous definitions, we can present **ST** as a *p*-matrix logic associated with the 3-element Kleene algebra.

DEFINITION 2.6. *The 3-element Kleene algebra is the structure*

$$\mathbf{K} = \langle \{1, \frac{1}{2}, 0\}, \{f^{\neg}_{\mathbf{K}}, f^{\wedge}_{\mathbf{K}}, f^{\vee}_{\mathbf{K}}\} \rangle,$$

[4] It might be argued that inferences involve formulae (and not sets of them) as conclusions. In this case, the conclusion should be read as a single formula, or the singleton of a single formula. All the results below carry over to this approach without loss of generality. The generalization of the Reflexivity and Monotonicity conditions on *p*-consequence, as they apply to the multiple-conclusion framework, was to the best of our knowledge first discussed in [8, p. 245], which surely builds on [38] and related works. Thanks to an anonymous referee for asking us for clarification in this regard.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

where the functions $f^{\neg}_{\mathbf{K}}$, $f^{\wedge}_{\mathbf{K}}$, $f^{\vee}_{\mathbf{K}}$ *are as follows*

$f^{\neg}_{\mathbf{K}}$	
1	0
$\frac{1}{2}$	$\frac{1}{2}$
0	1

$f^{\wedge}_{\mathbf{K}}$	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

$f^{\vee}_{\mathbf{K}}$	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

We take, additionally, the functions $\rightarrow$ and $\leftrightarrow$ to be definable via the usual definitions, i.e., $A \rightarrow B =_{\text{def}} \neg(A \wedge \neg B)$ and $A \leftrightarrow B =_{\text{def}} (A \wedge B) \vee (\neg A \wedge \neg B)$.

DEFINITION 2.7 ([14]). *A 3-valued* **ST***-matrix is a p-matrix*

$$\mathcal{M}_{\mathbf{ST}} = \langle \{1, \frac{1}{2}, 0\}, \{1\}, \{1, \frac{1}{2}\}, \{f^{\neg}_{\mathbf{K}}, f^{\wedge}_{\mathbf{K}}, f^{\vee}_{\mathbf{K}}\} \rangle$$

such that $\langle \{1, \frac{1}{2}, 0\}, \{f^{\neg}_{\mathbf{K}}, f^{\wedge}_{\mathbf{K}}, f^{\vee}_{\mathbf{K}}\} \rangle$ *is the 3-element Kleene algebra.*

Notice that for $\mathcal{M}$ a p-matrix, defining what it means for an $\mathcal{M}$-valuation v to *satisfy a formula* can be rather fishy. In particular, we may distinguish two different ways in which a formula might be satisfied by a valuation, i.e., if the said valuation assigns it a value in $\mathcal{D}^{+}$, or if the valuation assigns it a value in $\mathcal{D}^{-}$. These notions could be set to correspond, respectively, to whether the given formula is regarded as a premise or as a conclusion. Whether a formula is *valid*—i.e., whether it is satisfied by every valuation—will have, therefore, two corresponding precisifications. However, in what follows when we talk about satisfaction, satisfiability or validity of (a collection of) formulae, it will be the latter notion that we will be employing. These considerations, when instantiated in relation to systems extending **ST**, will render that a formula (collection of formulae) is satisfied by a valuation if and only if such a valuation assigns it (all of its members) the truth-value 1 or $\frac{1}{2}$. The usual definitions of satisfiability and validity then ensue from these clarifications.[5]

Let us now quickly move from the logic **ST** to the transparent theory of truth **STTT**. We will consider, thus, an expansion of the language $\mathcal{L}$ that also includes a unary truth predicate Tr, and we will furthermore require of **STTT** valuations that they satisfy Transparency, i.e., that for all A, where $\langle A \rangle$ is a name for sentence A

$$v(Tr\langle A \rangle) = v(A). \qquad \text{(Transparency)}$$

To valuations induced by the Kleene algebra, which additionally satisfy Transparency we refer to as *Kleene-Kripke valuations*. Understood this way, in particular, the following is a characterization of inferential **STTT** validity

$$\vDash_{\mathbf{STTT}} \Gamma \Rightarrow \Delta \textit{ iff } \text{for every valuation } v \begin{cases} \text{if } v(B) \in \{1\} & \text{for all } B \in \Gamma \\ \text{then } v(A) \in \{1, \frac{1}{2}\} & \text{for some } A \in \Delta \end{cases}.$$

This is not, however, the only interesting way to look at inferential validity for **STTT**. In fact, an alternative presentation of the previous semantic phenomenon highlights how closely it is related to Classical Logic (**CL**, for short).

[5] We would like to thank an anonymous referee for urging us to clarify these matters.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

$$\vDash_{\mathbf{STTT}} \Gamma \Rightarrow \Delta \textit{ iff } \text{there is no valuation } v \text{ such that } \begin{cases} v(B) \in \{1\} & \text{for all } B \in \Gamma \text{ and} \\ v(A) \in \{0\} & \text{for all } A \in \Delta \end{cases}.$$

Thus, as noted in [14, 16, 35], and [36] and many other places, these remarks imply among other things that every valid *inference* of Classical Logic is valid in this logic.

FACT 2.8. *An inference is valid in* **STTT** *iff it is valid in* **CL**.

However, **STTT** is not exactly every bit as classical as Classical Logic itself. Sure, since **CL** cannot nontrivially handle a transparent truth-predicate, as Tarski showed, some sacrifices must be made so that the inclusion of such a linguistic device does not break havoc. This is in fact the case, for in the context of **STTT** some classically valid *meta-inferences*—the most salient of which is the structural rule of Cut—are lost or rendered invalid, as the following remarks show.

DEFINITION 2.9 ([21]). *A meta-inference or meta-sequent on* $\mathcal{L}$*, written* $\Gamma \Rightarrow^1 A$*, is an ordered pair* (Γ, A) *where* $\Gamma \subseteq SEQ(\mathcal{L})$ *is finite and possibly empty and* $A \in SEQ(\mathcal{L})$.

DEFINITION 2.10. *For* $\mathcal{M}$ *a p-matrix, an* $\mathcal{M}$*-valuation* v *satisfies a meta-inference* $\Gamma \Rightarrow^1 A$*, written* $v \vDash_{\mathcal{M}} \Gamma \Rightarrow^1 A$*, by letting*

$$v \vDash_{\mathcal{M}} \Gamma \Rightarrow^1 A \quad \text{iff} \quad \textit{if } v \vDash_{\mathcal{M}} B \textit{ for all } B \in \Gamma\textit{, then } v \vDash_{\mathcal{M}} A.$$

A meta-inference $\Gamma \Rightarrow^1 A$ *is* $\mathcal{M}$*-valid, written* $\vDash_{\mathcal{M}} \Gamma \Rightarrow^1 A$*, by letting*

$$\vDash_{\mathcal{M}} \Gamma \Rightarrow^1 A \quad \text{iff} \quad \textit{for all } \mathcal{M}\textit{-valuations } v\textit{, } v \vDash_{\mathcal{M}} \Gamma \Rightarrow^1 A.$$

As is easy to check, this definition renders the structural rule of Cut invalid, i.e.,

$$\frac{\Gamma, A \Rightarrow \Delta \quad \Gamma \Rightarrow A, \Delta}{\Gamma \Rightarrow \Delta} \text{ Cut}$$

along with, e.g., the following meta-inferences, respectively, called Meta-Modus Ponens (MMP, for short) and Meta-Explosion (MEXP, for short)

$$\frac{\Gamma \Rightarrow A \to B, \Delta \quad \Gamma \Rightarrow A, \Delta}{\Gamma \Rightarrow B, \Delta} \text{ MMP} \qquad \frac{\Gamma \Rightarrow A, \Delta \quad \Gamma \Rightarrow \neg A, \Delta}{\Gamma \Rightarrow B, \Delta} \text{ MEXP}$$

which brings the opportunity to mention a number of things.

First, that as the discussion provided in [3], [35], and [21] shows, the invalid nature of these meta-inferences portrays **STTT** as being a close cousin to Priest's **LP**—or, perhaps more accurately, to the transparent theory of truth **LPTT**. In particular, in [3] the authors point to the fact that every invalid *inference* (e.g., transitivity for the conditional, Explosion, Modus Ponens, etc.) in the theory of truth based on this logic, i.e., in **LPTT**, has a corresponding invalid *meta-inference* in the nontransitive theory of truth **STTT** (e.g., Cut, Meta-Modus Ponens, Meta-Explosion, etc.). This is precisely established in [21] through a suitable translation.[6]

FACT 2.11. *A meta-inference is valid in* **STTT** *iff the corresponding inference is valid in* **LPTT**.

[6] For a more detailed account of this translation function and the importance it has in establishing the relevant results, see [21]. For a more detailed analysis of how the failure of, e.g., Meta-Explosion are linked with the paraconsistency of a given logic, see [2].

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

Second, that along with recovering the safe instances of Cut in **STTT**, we might also want to aim at recovering the safe instances of the previous invalid meta-inferences and, more generally, of all the meta-inferences that are invalid in **STTT** although valid in **CL**.

Third, that even if numerous classically valid meta-inferences are invalid in **STTT**, even Cobreros, Egré, Ripley and van Rooij are of the opinion that they should not be regarded as invalid *simpliciter*—which makes the recovery project even more interesting to pursue. In fact, we can see that, e.g., [16] makes some qualifications with regard to the failure of Cut.

> This nontransitivity, though, is quite limited. In fact, it is *restricted to* cases where *paradoxical sentences* rear their heads, in a quite particular way (...) So there is a fully precise sense in which the only failures of transitivity **STTT** allows for arise in paradoxical cases. Outside the realm of what Kripke calls *the paradoxical*, transitivity is perfectly *safe*. [16, our emphasis, p. 6].

This implies that restricted versions of the aforementioned meta-inferences where all or at least some crucial sentences are assumed to be nonparadoxical or nonpathological—and thus to receive classical truth-values—are indeed *valid*. Hence, it seems natural to wish that these facts can be properly expressed *within* the language of the theory itself, i.e., to have some means to express that when the involved sentences are assumed to be nonpathological, nothing logical needs to be lost. Sadly, this is something that cannot be done within **STTT**, for the system has no linguistic resources which can play this role. Designing a way to supplement this system in order to accomplish this task will be our goal in what follows.

§3. Recovering Cut (and much more). Our proposed way to achieve the goal of recovering Cut and all the classically valid meta-inferences that are invalid in **STTT** will be to supplement the target theory with a unary recovery operator $\circ A$ intended to mark those formulae to which Cut (and all other meta-inferences) can be safely applied—obtaining therefore a conservative expansion of the previous system.

First, regarding our recovery project, we shall require that we recover a particularly safe or *gentle* version of Cut, which is restricted to formulae that are assumed to receive a classical truth-value.[7]

$$\frac{\Gamma, A \Rightarrow \Delta \quad \Gamma \Rightarrow A, \Delta \quad \Rightarrow \circ A}{\Gamma \Rightarrow \Delta}\ \text{Gentle Cut}.$$

Second, we shall also demand that we recover safe or *gentle* versions of the otherwise invalid meta-inferences previously referred to as Meta-Modus Ponens and Meta-Explosion (we refer to these gentle versions, respectively, as GMMP and GMEXP, for short).

$$\frac{\Gamma \Rightarrow A \to B, \Delta \quad \Gamma \Rightarrow A, \Delta \quad \Rightarrow \circ A}{\Gamma \Rightarrow B, \Delta}\ \text{GMMP} \qquad \frac{\Gamma \Rightarrow A, \Delta \quad \Gamma \Rightarrow \neg A, \Delta \quad \Rightarrow \circ A}{\Gamma \Rightarrow B, \Delta}\ \text{GMEXP}.$$

[7] To answer an anonymous referee's question concerning how and where A needs to be marked with the recovery operator in the meta-inferences above, we shall highlight that the sentence $\circ A$ must always appear in the right-hand side of the sequent $\Rightarrow \circ A$, which gently accompanies the classically valid although **ST**-invalid meta-inferences discussed above. This is, precisely, because $\Rightarrow \circ A$ is meant to represent the fact that it is true that A has been assigned a classical value.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

Third, and finally, we shall prove that this can be done for every classically valid meta-inference that is invalid in **STTT**. But before turning to show that indeed all this can be accomplished in the system we will be presenting in this section, let us discuss a pressing question that might worry some readers: *Given a Kleene-Kripke valuation, which sentences deserve to be marked with the recovery operator*?

Since **STTT** is a theory of truth built using Kleene-Kripke valuations, this question is equivalent to the question: *Given a Kleene-Kripke valuation, which sentences are non-pathological*? It is a well-known fact that there is no general answer to this question. The particular answers to it depend, actually, on two choices. On the one hand, *the particular ground model* on top of which the Kleene-Kripke valuations are built, i.e., the set of sentences that are assumed from scratch to be true, or false. On the other hand, *the particular fixed-point* chosen to determine the extension and anti-extension of the truth-predicate, i.e., the set of sentences that are taken to be true, or false, or pathological, after the inductive definition of the truth predicate reaches a fixed-point.

With regard to the former choice, if we are working with an arithmetical theory as the base theory for a transparent theory of truth, it is usual to let the ground model be an arithmetical model and, hence, to let the true sentences of the ground model be those—and only those—that are true in a given set of arithmetical models. But this is not mandatory and nothing prevents us from including some pathological sentence, e.g., the Truth-Teller, among the sentences that are taken to be true from the very beginning. With regard to the latter choice—that of one fixed-point among the minimal, the intrinsic, the various maximal, etc.—it is well known that Kripke adopted an open-minded policy towards this issue, as registered for example in the following quote.[8]

> Nor is it even my present purpose to make any firm recommendation between the minimal fixed point of a particular valuation scheme and the various other fixed points (. . .) My purpose is rather to provide a family of flexible instruments which can be explored simultaneously and whose fertility and consonance with intuition can be checked. [28, p. 712].

The upshot of these reflections is, therefore, that there is no absolutely general way of determining which are the nonpathological sentences according to the Kleene-Kripke valuations and, thus, of determining which sentences ought to be marked with the recovery operator. At this point one may ask: does this mean the recovery project is pointless?

We do not think so. For, although we cannot exhaustively pin down a set of sentences that can be marked with the recovery operator, we can still recover all the meta-inferences valid in **CL** but not in **STTT** by *reasoning under the assumption* that some of the featured sentences are nonpathological. In this vein, the above gentle versions of Cut, Meta-Modus Ponens and Meta-Explosion can be seen as claiming that *if* the premise-sequents for the regular versions of these rules hold *and* additionally some of the sentences involved are nonpathological, *then* the conclusion-sequent holds.

This allows us to draw a genuine parallel with the work done on Logics of Formal Inconsistency (LFIs, for short) as carried out, e.g., in [9, 10, 11, 12].[9] Logicians working in

[8] Notwithstanding these remarks, let us notice that there are some sentences which can be undoubtedly classified as pathological, regardless of these particular choices. One such example is, as expected, the Liar sentence.

[9] LFIs are powerful paraconsistent logics that fix an interesting distinction between contradictions and inconsistencies. More formally, a logic **L** is an LFI iff there is some possibly empty set of formulae $\circ(A)$ depending on A such that the following conditions are met:

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

the LFIs tradition think of their logical frameworks as recovering classical reasoning when working *under the assumption* that some of the sentences in question are *consistent* [9, p. 18]. Similarly, work within our intended extension of **STTT** can be thought as recovering classical reasoning—i.e., as recovering all classical meta-inferences—when working *under the assumption* that some of the sentences in question are *nonpathological*.

Moving forward towards our aim, then, let **STTT**$^\circ$ be the extension of **STTT** with a unary operator $\circ A$, equipped with the following truth-table.

	f°
1	1
$\frac{1}{2}$	0
0	1

In what follows, we will prove that every meta-inference that is valid in **CL** and is invalid in **STTT** can be recovered in **STTT**$^\circ$. Such a general proof will have, as a corollary, that the gentle versions of Cut, Meta-Modus Ponens and Meta-Explosion detailed above are valid in **STTT**$^\circ$.

As an introduction to this general result, let us highlight that its comprehensive nature shows that we are offering nothing more than a sort of *Derivability Adjustment Theorem* (DAT, for short). In general, DATs show how a weaker logic can be used to recover the inferential behavior of a stronger logic, with the help of some assumptions concerning the sentences involved. The essential intuition behind such theorems was first incarnated in the Belgian approach with the aid of adaptive logics, e.g., in [4], and later in the Brazilian approach with the aid of LFIs. The former way permits to show how, assuming *in the meta-theory* the consistency of some sentences, classical reasoning can be recovered in the context of an otherwise nonmonotonic reasoning. As we said earlier, the latter way allows showing how, assuming *in the theory* the consistency of some sentences, classical reasoning can be recovered in the context of an otherwise paraconsistent reasoning.

In our case, we adopt a maneuver that explicitly states in the theory the assumption that some of sentences involved are nonpathological.[10] In doing so, we will prove that all classical meta-inferences can be recovered in the context of an otherwise nontransitive (and hence, meta-inferentially speaking, nonclassical) reasoning.

DEFINITION 3.1. *For every formula $C \in FOR(\mathcal{L})$, let $\mathsf{At}(C)$ be the set of propositional variables appearing in C. Additionally, let $\Gamma \subseteq SEQ(\mathcal{L})$ and $A \in SEQ(\mathcal{L})$ be such that $\Gamma = \{\Sigma_i \Rightarrow C_i \mid 1 \leq i \leq n\}$, $\Sigma_i = \{B_{1_i}, \ldots, B_{m_i}\}$ and $A = \Sigma \Rightarrow C$. From these, we define the following sets*

$$\mathsf{At}(\Sigma_i) = \bigcup\{\mathsf{At}(B_{1_i}), \ldots, \mathsf{At}(B_{m_i})\} \qquad \mathsf{At}(\Sigma_i \Rightarrow C_i) = \mathsf{At}(\Sigma_i) \cup \mathsf{At}(C_i)$$

There are some Γ, A, B such that: $\begin{cases} \nvDash_{\mathbf{L}} \Gamma, A, \neg A \Rightarrow B \\ \nvDash_{\mathbf{L}} \Gamma, \circ(A), A \Rightarrow B \\ \nvDash_{\mathbf{L}} \Gamma, \circ(A), \neg A \Rightarrow B. \end{cases}$

And, yet, for all Γ, A, B, it is the case that $\vDash_{\mathbf{L}} \Gamma, \circ(A), A, \neg A \Rightarrow B$.

LFIs go back to Newton da Costa's work on C-systems in [19] and [20], but were later developed in a systematic way by W. Carnielli, J. Marcos, M. Coniglio and others in their seminal articles [10], [11].

[10] We are not looking for *optimal* conditions for the DAT, but only for *sufficient* conditions. In doing so, we mimic the strategy followed by J.C. Beall in [5] and assume that every formula in the meta-inference gets a classical truth-value.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

$$\mathsf{At}(\Gamma) = \bigcup\{\mathsf{At}(\Sigma_1 \Rightarrow C_1), \ldots, \mathsf{At}(\Sigma_n \Rightarrow C_n)\} \qquad \mathsf{At}(\Gamma \Rightarrow^1 A) = \mathsf{At}(\Gamma) \cup \mathsf{At}(A).$$

FACT 3.2. *Let* $\mathsf{At}(\Gamma \Rightarrow^1 A)$ *be* $\{p_1, \ldots, p_k\}$*, then the following holds*

$$\vDash_{\mathbf{CL}} \Gamma \Rightarrow^1 A \quad \textit{iff} \quad \vDash_{\mathbf{STTT}^\circ} \Gamma, \Rightarrow\circ p_1, \ldots, \Rightarrow\circ p_k \Rightarrow^1 A.$$

Proof. Assume, without loss of generality, that $\Gamma = \{\Sigma_i \Rightarrow C_i \mid 1 \leq i \leq n\}$ and $A = \Sigma \Rightarrow C$.

LTR: Assume that for every valuation v of **CL**, if v satisfies every premise sequent $\Sigma_i \Rightarrow C_i$ in **CL**, then v satisfies the conclusion sequent $\Sigma \Rightarrow C$ in **CL**. Consider, now, a valuation v of $\mathbf{STTT}^\circ$ that satisfies every $\Sigma_i \Rightarrow C_i$ and also every $\Rightarrow\circ p_j$ (for $1 \leq j \leq k$). Thus, every propositional letter p_j (for $1 \leq j \leq k$) in the meta-inference will be such that either $v(p_j) = 1$ or $v(p_j) = 0$. It is easy to notice that, then, v is also a **CL** valuation which satisfies every $\Sigma_i \Rightarrow C_i$. Given this, we know that v satisfies the conclusion $\Sigma \Rightarrow C$ in **CL**, which implies that v satisfies $\Sigma \Rightarrow C$ in $\mathbf{STTT}^\circ$. Thus, v cannot be a counterexample to the meta-inference in $\mathbf{STTT}^\circ$, whence it is valid in it.

RTL: Assume that for every valuation v of $\mathbf{STTT}^\circ$, if v satisfies every premise sequent $\Sigma_i \Rightarrow C_i$ and $\Rightarrow\circ p_j$ (for $1 \leq j \leq k$) in $\mathbf{STTT}^\circ$, then v satisfies the conclusion sequent $\Sigma \Rightarrow C$ in $\mathbf{STTT}^\circ$. Consider, now, a valuation v of **CL** that satisfies every premise sequent $\Sigma_i \Rightarrow C_i$. It is easy to notice that this is also a $\mathbf{STTT}^\circ$ valuation which satisfies every $\Sigma_i \Rightarrow C_i$ and also every $\Rightarrow\circ p_j$ (for $1 \leq j \leq k$). From the latter we infer that $v(\circ p_j) = 1$, for $1 \leq j \leq k$. Thus, every propositional letter p_j (for $1 \leq j \leq k$) in the meta-inference will be such that either $v(p_j) = 1$ or $v(p_j) = 0$. Given our initial assumptions, we know that v satisfies $\Sigma \Rightarrow C$ in $\mathbf{STTT}^\circ$, which by the above information that no formula is assigned the value $\frac{1}{2}$ in v, entails that either $v(B) = 0$, for some $B \in \Sigma$, or $v(C) = 1$. But this will amount to the sequent $\Sigma \Rightarrow C$ being satisfied in **CL**. Whence, v cannot be a counterexample to the corresponding meta-inference in **CL**, whence it is valid in it. □

Thus, we have established that we are able to recover within $\mathbf{STTT}^\circ$ every classically valid inference that is invalid in the nontransitive theory of truth **STTT**. But as everyone acquainted with the literature on naïve theories can anticipate, this is not even nearly enough. Notwithstanding the importance of the previous remarks, there is still a crucial question which remains to be asked: is the resulting theory *nontrivial*? We will answer this shortly, but not before introducing some subtleties regarding the way in which self-reference is achieved in the nontransitive theory of truth in question. That is to say, the answer to the target question—i.e., *is* $\mathbf{STTT}^\circ$ *nontrivial*?—will depend on how self-reference and self-referential sentences, such as the Liar, the Truth-Teller, etc. are represented.

Thus, to formally mimic the self-referential character that some sentences like "This sentence is in English" or more prominently "This sentence is false" have, there seems to be two main technical options: through a *strong* or through a *weak* procedure. The latter option achieves this goal by requiring a self-referential sentence to be *equivalent* to a sentence that "talks about" the first one. The former involves an essential use of identities. This strong alternative, either (i) requires a term to be *identical* to the name of a sentence that "talks about" the first term, or (ii) involves a meta-linguistic denotation function from names to sentences of the language that have occurrences of that name in it. In what follows, we will take the notational liberty of calling a theory **Th** that uses a weak self-referential procedure as $\mathbf{Th_w}$ and—analogously—one that uses a strong self-referential procedure as $\mathbf{Th_s}$.

These options might be instantiated by a plethora of technical means, varying from one framework to another. Sometimes these procedures can be present "in" the theories in

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

question, e.g., if they are extensions of arithmetical theories codifying their own syntax. In such a case, self-reference may be either achieved through a strong or a weak procedure, i.e., through the strong Diagonal Lemma, or the weak Diagonal Lemma—as recently remarked by Lavinia Picollo in [31]. Thus, for example, a Liar sentence obtained through the strong Diagonal Lemma will be represented by a term l which is *identical* to the Gödel code $\ulcorner \neg Tr(l) \urcorner$. Whereas, a Liar sentence obtained through the weak Diagonal Lemma will be represented by a sentence L which is *equivalent* to the sentence $\neg Tr \ulcorner L \urcorner$.[11]

Nevertheless, in some other opportunities these ways of obtaining self-reference do not come with the theories, in which case they can be "imposed from the meta-language" to the theories in question, e.g., by restricting their valuations. For example, the way self-reference is achieved in, e.g., [17] and [36] constitutes for us a case of the implementation of a meta-linguistic strong procedure. In both of these works, the valuations of the target theory of truth are supplemented with a denotation function, later restricting the resulting valuations to those guaranteeing the existence of certain pathological sentences which denote sentences that talk about themselves. For instance, in the referred works, a Liar sentence is represented by a sentence L, only if the term $\langle L \rangle$ denotes the sentence $\neg Tr\langle L \rangle$ via the corresponding denotation function.

Alternatively, achieving self-reference by an implementation of a similar—although weak—meta-linguistic procedure may also be possible. To this extent, the valuations of the target theory of truth should be restricted to those guaranteeing the existence of certain pathological sentences which are equivalent to sentences that talk about themselves. For instance, in such a case a Liar sentence may be represented by a sentence L, only if L is equivalent to the sentence $\neg Tr\langle L \rangle$. This is, precisely, the route we are going to pursue in §4.

Going back to **STTT**, we may notice that it is often presented using a *strong* self-referential procedure to achieve self-reference. For instance, [36] implements a meta-linguistic denotation function working as detailed in the previous paragraphs—but for another alternative discussion of a strong self-referential procedure applied to **STTT**, see [22]. Whence, we can rightfully say that the discussion of this system has been actually focused, at least until now, in $\mathbf{STTT_s}$.[12]

These facts concerning how self-reference is achieved will be essential to our discussion of the triviality or nontriviality of $\mathbf{STTT}^{\circ}$. For, as we observe, the theory $\mathbf{STTT}^{\circ}_{\mathbf{s}}$, i.e., the system resulting from supplementing $\mathbf{STTT_s}$ with a unary recovery operator $\circ A$, is indeed trivial.

FACT 3.3. $\mathbf{STTT}^{\circ}_{\mathbf{s}}$ *is unsatisfiable.*

[11] This allows to answer an anonymous referee's question as to whether arithmetical theories codifying their own syntax provide strong or weak procedures to obtain self-reference.

[12] In their article [16], Cobreros, Egré, Ripley, & van Rooij work with **STTT** on top of an arithmetical theory capable of codifying its own syntax. Although they assume that such a theory can prove the existence of the usually problematic pathological sentences—allegedly, by instantiating either the strong or the weak Diagonal Lemma—they do not make any explicit case concerning the particular procedure employed in obtaining such results. In other words, the authors do not make any explicit clarification concerning whether the pathological sentences that they talk about (e.g., the Liar sentence) are obtained via the strong or the weak Diagonal Lemma. Our discussion and our results below can be taken as stating that if the former is the case—i.e., if in [16] the authors are entertaining $\mathbf{STTT_s}$—then the theory cannot be nontrivially extended with a recovery operator. This is, precisely, why in what remains of the article we recommend taking the latter route. We would like to thank an anonymous referee for urging us to clarify these issues.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

Proof. The proof is simple, by considering a usual Strengthened Liar sentence S, such that, e.g., $\langle S\rangle$ denotes $\neg Tr\langle S\rangle \wedge \circ Tr\langle S\rangle$. We will show that there is no stable assignment of truth values—i.e., no valuation—for $Tr\langle S\rangle$. If $v(Tr\langle S\rangle) = 0$, then $v(\neg Tr\langle S\rangle \wedge \circ Tr\langle S\rangle) = 1$, and so $v(Tr\langle S\rangle) = 1$. If $v(Tr\langle S\rangle) = 1$, then $v(\neg Tr\langle S\rangle \wedge \circ Tr\langle S\rangle) = 0$, and so $v(Tr\langle S\rangle) = 0$. Now, if $v(Tr\langle S\rangle) = \frac{1}{2}$, then $v(\neg Tr\langle S\rangle \wedge \circ Tr\langle S\rangle) = 0$, and so $v(Tr\langle S\rangle) = 0$. □

The previous triviality result for $\mathbf{STTT}^{\circ}_{\mathbf{s}}$ suggests that there are two immediate options

(I) Either giving up some expressive requirements (such as having a recovery operator).

(II) Or giving up the *strong* self-referential procedure.

In the rest of the article, we will discuss the latter option. In doing this, we will be exploring the prospect of changing the way to achieve self-reference in a nontransitive theory of truth, using some *weak* self-referential procedure, i.e., establishing self-reference via equivalences and not via identities.

§4. The Goodship project, revisited. In what follows, we are going to show how it is possible to prove the nontriviality of our target nontransitive theory of truth $\mathbf{STTT}^{\circ}_{\mathbf{w}}$, which is able to recover all the meta-inferences that are valid in **CL** although they are not valid in **STTT**.

Given the previous triviality results for $\mathbf{STTT}^{\circ}_{\mathbf{s}}$, as we recently advertised, our positive result will rely heavily on the implementation of a weak self-referential procedure to express or represent pathological sentences such as the Liar, the Truth-Teller, Curry and possibly others—including potential revenge-like sentences. By a *weak* procedure we mean that the self-referential character of these sentences will be portrayed via equivalences—through some suitable biconditional—that talk about them.

In this vein, it is important for us, before presenting the actual nontriviality result for $\mathbf{STTT}^{\circ}_{\mathbf{w}}$, to highlight how our effort here originates from a suggestion that Laura Goodship gave more than twenty years ago to build nontrivial naïve theories, e.g., transparent theories of truth.

In her 1996 article [25], Laura Goodship stated that there was an option for those working with naïve theories to avoid triviality: they should express the corresponding naïve comprehension axioms using a *suitable* biconditional. Given that $A \leftrightarrow B$ are standardly defined as the conjunction of the two conditionals $A \rightarrow B$ and $B \rightarrow A$, she emphasizes that what actually needs to be suitable is the corresponding conditional. But what does it mean for a conditional to be suitable, so that nontriviality is granted? By this Goodship means that the connective in question should be nonclassical to a certain extent: it should either (i) invalidate Modus Ponens, or (ii) invalidate Absorption (sometimes also called Contraction) for the conditional. In [5, p. 332], Beall refers to theories pursuing the former route, as theories located within what Priest and him call *the Goodship Project*.

Ever since Goodship outlined these options, many have followed her recommendations, whether implicitly or explicitly—the ones most explicitly inspired by these suggestions being [1, 29], and [34]. For example, Beall [5], [7] has recently advocated a detachment-free paraconsistent approach, where he embraces the failure of Modus Ponens displayed by paraconsistent logics such as **LP**. Similarly, Horsten [27] essentially advocates a dual strategy in a paracomplete setting, by abandoning Contraction for the conditional—although in his case this route is followed independently of the Goodship's suggestions. Omori [29]

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

also proposed an extension of paraconsistent set-theory based on **LP** which essentially employs a biconditional without Modus Ponens to express self-reference. This was later carried over in the field of transparent theories of truth by Barrio, Pailos, & Szmuc in [1], using a close cousin of **LP** as the underlying logic. Finally—and most saliently—recently Priest offered "an extended exploration of the strategy [of adopting the **LP**-conditional to formulate the comprehension axioms], its strengths, its weaknesses, and the various different ways in which it may be implemented" [34, p. 58].

In this essay, we do think of ourselves as continuing this tradition. However, we think we are offering a novel point of view within this interesting family of approaches. The results of the previous section allow us to conclude that employing a biconditional to express self-reference is essential for the nontriviality results—in that, we agree with Goodship. In fact, in the next section, we will be presenting a model for the transparent theory of truth $\mathbf{STTT}^{\circ}_{\mathbf{w}}$, where to grant nontriviality it is essential that self-reference be expressed through the **STTT** material conditionals. Nevertheless, given that the material conditional in **STTT** is *classical*, meaning that all the tautologies and inferences involving it which hold in **CL** also hold in **STTT**, this allows us to conclude that Goodship's recommendations were perhaps too narrow. There seems to be no need to employ a nonclassical conditional.

Let us emphasize this last point. According to our proposal, *the Goodship Project should be reformulated* in this way: nontrivial naïve theories should express self-reference with a suitable biconditional, avoiding to use identities or a meta-linguistic function. But, for the biconditional to be suitable, we must not require that it be nonclassical. In fact, it can be completely classical, if the *structural* properties of the logic are modified adequately. For instance, as the semantics for our theory $\mathbf{STTT}^{\circ}_{\mathbf{w}}$ witnesses below, the use of the biconditional in expressing self-reference is essential for the nontriviality of the theory, the biconditional employed is classical—but this combination is guaranteed not to be harmful because the logic is *nontransitive*.

This seems, to us, to offer a completely novel option within the Goodship Project. The attractive feature of this route is that it is not necessary to abandon the classicality of the target biconditional to generate self-reference in the context of naïve theories and, more specifically, of transparent theories of truth.

4.1. A nontriviality result for $\mathbf{STTT}^{\circ}_{\mathbf{w}}$**.** Our target theory of truth $\mathbf{STTT}^{\circ}_{\mathbf{w}}$ is nontransitive, has a transparent truth-predicate, and counts with a recovery operator that helps regain all the classically valid but **ST**-invalid meta-inferences. Still, we need to show that this system accomplishes this task while appropriately dealing with pathological phenomena.

To this extent, it may be helpful for the reader that we advance that establishing the nontriviality result below will require proving two things. On the one hand, that there is a $\mathbf{STTT}^{\circ}$ valuation compatible with the existence of pathological sentences constructed via a weak self-referential procedure and, on the other hand, that this valuation does not satisfy every formula.

Proving that there is an $\mathbf{STTT}^{\circ}$ valuation compatible with the existence of pathological sentences constructed via a weak self-referential procedure is not that straightforward. The reason is that we are not working with an extension of an arithmetical theory that codifies its own syntax, and thus must adopt some alternative technique to guarantee that pathological sentences are around. As we previously said, this will be done by restricting the valuations of the theory to those where some sentences are equivalent to sentences that talk about themselves. In such a framework, then, a Liar sentence will be represented by a sentence that is equivalent to a sentence saying that it is false. Similarly, a Curry sentence

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

will be represented by a sentence that is equivalent to a sentence saying that if it is true, then everything follows. And so on, and so forth.[13]

Thus, we will construct a set called PseudoDL (short for "Pseudo Diagonal Lemma"), including all sorts of equivalences between sentences, and sentences involving semantic predications about themselves. We will, furthermore, restrict the **STTT**$^{\circ}$ valuations to those satisfying PseudoDL. This will allow us to fulfill our first requirement, showing that there is a **STTT**$^{\circ}$ valuation compatible with the existence of pathological sentences constructed via a weak self-referential procedure. In what follows, we give the details of how to construct the set PseudoDL.[14]

First, we will need to distinguish an infinite proper subset of the set of propositional variables that we will later mark—in the metalanguage—with a *. This last move is not essential, though. But putting a mark to those distinguished propositional letters will make things easier to understand, as those propositional letters will play a key part in the self-referential procedure we are about to present. Thus, we will refer to the members of that set as, e.g., p^*, q^*, etc.

Second, we build a set called PseudoDL which only includes sentences of the form

$$p^* \leftrightarrow A_{p^*},$$

where p^* is a distinguished propositional variable and A_{p^*} is a sentence that has at least one instance of $Tr\langle p^*\rangle$ as a subformula. This approach allows for pathological sentences of all sorts. In fact, instantiation on A_{p^*} in the context of the above biconditional statement will exemplify every self-referential sentence that includes a truth predicate. For example, if $t^* \leftrightarrow Tr\langle t^*\rangle$ is in PseudoDL, then we are allowed to say that t^* is a Truth-Teller sentence, and if $c^* \leftrightarrow Tr\langle c^*\rangle \rightarrow \bot$ is in PseudoDL, then we are allowed to say that c^* is a Curry sentence. In particular, we will be able to express or represent the traditional pathological cases (The Liar, Curry, the Truth-Teller, etc.) but also possible *new pathological sentences* including the recovery operator. Thus, e.g., if $s^* \leftrightarrow \neg Tr\langle s^*\rangle \wedge \circ Tr\langle s^*\rangle$ is in PseudoDL, then s^* is a Strengthened Liar sentence.

Third, concerning the schema above, we will pair each distinguished propositional variable p^* with one and only one formula-schema A_{p^*}. Thus, if we have that $t^* \leftrightarrow Tr\langle t^*\rangle$ belongs to PseudoDL, we will not have that $t^* \leftrightarrow Tr\langle t^*\rangle \vee \bot$ belongs to PseudoDL. Moreover and simultaneously, we will pair each formula-schema A_{p^*} with *at least one* of the distinguished propositional variables. Thus, if $t^* \leftrightarrow Tr\langle t^*\rangle$ belongs to PseudoDL, this does not prevent, e.g., $u^* \leftrightarrow Tr\langle u^*\rangle$ from belonging to PseudoDL, too. This will render the existence of *at least* one Liar, one Truth-Teller, one Curry, one Strengthened Liar, etc., depending on the initial choices one makes.

Finally, we will simply restrict the set of the **STTT**$^{\circ}$ valuations to the ones that satisfy each member of PseudoDL. This allows us to show that there are **STTT**$^{\circ}$ valuations compatible with the existence of pathological sentences. Let us refer to the resulting theory as $\mathbf{STTT}^{\circ}_{\mathbf{w}}$. To show the *nontriviality* of this theory, we need something more. We need to show that there is at least one $\mathbf{STTT}^{\circ}_{\mathbf{w}}$ valuation which does not satisfy every sentence.

[13] By no means are we claiming that this is the best, the most elegant or the preferable way to achieve this goal. We are just saying that this is one possible way. Many other alternatives might be available or not, and we would be more than happy to consider them. But for the sake of the present discussion, the following will suffice.

[14] Of course, one may be even more liberal than we are in building PseudoDL. Here we just depict one possible approach.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

To put it in the terms of our recent discussion, we need to show that there is a **STTT**$^{\circ}$ valuation which satisfies every member of PseudoDL and yet does not satisfy every sentence. This is what we show next. Recall that, as per our discussion of satisfaction in this nontransitive approach in §2, a formula (collection of formulae) is satisfied by a valuation if and only if such a valuation assigns it (all of its members) the truth-value 1 or $\frac{1}{2}$.

FACT 4.1 (Nontriviality). *There is a* $\mathbf{STTT}^{\circ}_{\mathbf{w}}$ *valuation, for which not every sentence is satisfied.*

Proof. Consider the valuation v that assigns to every propositional letter the value $\frac{1}{2}$, and hence to every distinguished propositional letter p^* the value $\frac{1}{2}$. We first confirm that such a valuation satisfies every member of PseudoDL, i.e., we check that every biconditional $p^* \leftrightarrow A_{p^*}$ in PseudoDL will receive either the value 1 or the value $\frac{1}{2}$. This is true, precisely because of the way in which the **ST** biconditional is defined. Next, we show that not every sentence is satisfied by this valuation. In particular, let $r^* \leftrightarrow \neg Tr\langle r^*\rangle$ be a member of PseudoDL, so that r^* can be regarded as a Liar sentence. Then, consider the formula $\circ r^*$, for which we know that if $v(r^*) = \frac{1}{2}$, then $v(\circ r^*) = 0$. Thus, we have a formula that is not satisfied by v.[15] □

It is easy to see *that* the use of biconditionals in expressing self-reference is what puts this particular theory, $\mathbf{STTT}^{\circ}_{\mathbf{w}}$, on the path that Laura Goodship described to avoid triviality. After all, the difference between the triviality of $\mathbf{STTT}^{\circ}_{\mathbf{s}}$ and the nontriviality of $\mathbf{STTT}^{\circ}_{\mathbf{w}}$ is due, precisely, to the way in which self-reference is expressed. While in the former it is expressed through a strong procedure, in the latter it is expressed through a weak procedure. What is peculiar of this solution is, though, that the biconditional employed to express self-reference is just **ST**'s material biconditional, and not any other introduced further.

But it may be instructive to illustrate just *how* the use of biconditionals in expressing self-reference is what puts this particular theory, $\mathbf{STTT}^{\circ}_{\mathbf{w}}$, on the path that Laura Goodship described to avoid triviality. In other words, it may be interesting to understand how employing biconditionals in expressing self-reference in $\mathbf{STTT}^{\circ}_{\mathbf{w}}$ is the key to avoiding the triviality result of Fact 3.3. This can be exemplified as follows. Take, for example, a Strengthened Liar S like the one used to prove the triviality result for $\mathbf{STTT}^{\circ}_{\mathbf{w}}$ in the said Fact 3.3. If we assume, as it is done in that context, that S denotes $\neg Tr\langle S\rangle \wedge \circ Tr\langle S\rangle$, necessarily $Tr\langle S\rangle$ and $\neg Tr\langle S\rangle \wedge \circ Tr\langle S\rangle$ are assigned the *same truth-value*. Given that this is impossible for appropriately constructed revenge paradoxes, triviality ensues.

However, in the context of theories expressing self-reference through a weak procedure—an example being our own theory $\mathbf{STTT}^{\circ}_{\mathbf{w}}$—a Strengthened Liar sentence like the above would be represented by a sentence s^* assumed to be equivalent to $\neg Tr\langle s^*\rangle \wedge \circ Tr\langle s^*\rangle$. This, in itself, does *not* require that both of these sentence have the same truth-value, but merely that they are assigned some truth-values which guarantee the satisfaction of the equivalence $s^* \leftrightarrow \neg Tr\langle s^*\rangle \wedge \circ Tr\langle s^*\rangle$. (In the case of the valuation provided in the proof above, just assigning s^* the value $\frac{1}{2}$ will suffice. Then, even if the right side of the biconditional will get value 0, the biconditional itself will get value $\frac{1}{2}$.) The same holds for all the remaining pathological sentences which cause trouble for $\mathbf{STTT}^{\circ}_{\mathbf{s}}$ but not for $\mathbf{STTT}^{\circ}_{\mathbf{w}}$, as shown in the nontriviality result for $\mathbf{STTT}^{\circ}_{\mathbf{w}}$ above. There is a valuation

[15] We would like to thank Dave Ripley for discussion concerning this result.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

which guarantees that pathological sentences are around and that, still, not every sentence is satisfied.

The upshot of this section is that to obtain the desired nontriviality result for the target theory $\mathbf{STTT}^{\circ}_{\mathbf{w}}$, it is essential to employ a weak self-referential procedure to guarantee the existence of pathological sentences.[16] Nevertheless, nontriviality is a necessary but not a sufficient condition to obtain an interesting theory. In particular, one could question whether or not this theory is a better option than its actual competitors within the Goodship Project. To this discussion we now turn.

4.2. Philosophical reflections. Having proposed a new alternative within the Goodship Project, we must now discuss whether our theory $\mathbf{STTT}^{\circ}_{\mathbf{w}}$ fares better than the competitors in its own family. Prime among them is the nondetachable approach represented by **LP**, which is of special interest due to Graham Priest's dedicated study of this option, recently published as [34]. Thus, in what follows we will be comparing these two alternatives.

However, to compare different options one needs several criteria to rank them, and in fact comparisons between different approaches to transparent truth have appeared in the literature in several forms and shapes. As Hjortland [26] points out, it is usual to compare transparent theories of truth by evaluating their grades of classical involvement. According to him

> classical theories typically try to recapture the transparency of truth in special cases, while nonclassical theories try to *recapture classical principles* in special cases. This is a form of damage control. The target theories should revise the problematic principles while preserving sufficiently strong restricted versions. In short, most formal theories of truth are developed after the 'maxim of minimal mutilation'. [26, p. 1, our emphasis].

This criterion suggests that the nonclassical theories of truth should preserve properties of Classical Logic, if possible.[17] Thus, we might as well compare the Goodship Project based on **LP** to the alternative based on **ST**.

So let us focus, first, on the logical truths of Classical Logic: there, both the theories of truth based on **LP** and **ST** fare equally well, since both preserve all the tautologies of **CL**. Hence, at the level of the tautologies these approaches have nothing to recapture. However, when we look at the inferences that are valid in Classical Logic, things start to be different: while the inferences of **ST** are completely classical, meaning that they coincide with those valid in **CL**, this is *not* the case when we talk about **LP**. The upshot is that, regarding **ST**, there is nothing to recapture concerning the inferences, whereas in the case of **LP** this is

[16] An anonymous referee wonders whether or not this counts as an expressive limitation of this approach. In a way, our results imply that the use of a strong self-referential procedure to express self-reference—e.g., through identities—should be avoided, if one wishes to recover the classically valid meta-inferences without falling into triviality. This expressive limitation is a price to pay, if identity is a classical notion. If one wishes to express self-reference through a strong procedure, though, one available option would be to explore nonclassical accounts of identity, like the one discussed by Graham Priest in [33].

[17] Although, as remarked by an anonymous referee, some of the arguments presented above—e.g., those to the extent that our theory enjoys the unrestricted validity of Modus Ponens whereas rival theories do not—lend support to our approach even if we do not implicitly assume that Classical Logic sets some sort of standard that ought to be respected.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

mandatory—and, in fact, some efforts in this direction were done in naïve theories based on **LP** or some related systems, e.g., [1, 29], and [34].

Concerning the classically valid inferences that **LP**-based approaches fail to validate, we can focalize the discussion in a very important point, intimately related to the behavior of conditionals in the context of **LP**. As is well known, the conditional of **LP** is *too weak* to be taken seriously. More importantly, when talking about transparent theories of truth, it is not only in formulating the self-referential sentences that conditionals play an essential role, but also in formulating the T-Scheme. Thus, if the conditional involved in these tasks is without Modus Ponens, it does not seem to be a real conditional—as Priest himself acknowledged in a recent article.

> [T]he material conditional (and biconditional) do not detach. So, given only the material T-scheme, one cannot infer A from $Tr\langle A\rangle$, or vice versa. Now, we certainly do seem to reason using detachment in this way. Thus, for example, as many have noted, one function of the T-scheme would appear to be to make endorsement possible in contexts where this cannot be done explicitly. [34, p. 66].

In this vein—as an anonymous referee emphasizes—if a system counts with a conditional $\rightarrow$ that fails to validate Modus Ponens or the transitivity of the conditional, this raises the suspicion whether, within this theory, a statement of the form $A \leftrightarrow B$ expresses a real equivalence.[18]

Restricted universal quantification is another important aspect of a naïve theory of truth where the failure of detachment of the **LP**-based approach spills into. That this approach has difficulties in this regard is made evident when looking at the phenomenon of blind endorsement. That is, of the ascription of truth to sentences that are identified in a way which gives no clue about their content (e.g., when someone holds that everything the Bible says is true). For instance, while saying "All of the axioms of Peano Arithmetic are true"—formalized as $\forall x(Ax(x) \rightarrow Tr\langle \dot{x}\rangle)$—we would endorse all of an infinite number of claims, granted the axioms are infinite in number. Given that $\forall x \forall y(x' = y' \rightarrow x = y)$ is an axiom of Peano Arithmetic,[19] we may infer that $Tr\langle \forall x \forall y(x' = y' \rightarrow x = y)\rangle$, but if a nondetachable conditional is at play, one cannot be sure that it will be legitimate to infer that. Much to the contrary, our proposal based on **ST** avoids these difficulties completely. The conditional built in **ST** is essentially employed to express self-reference, however it not only allows Modus Ponens to apply unrestrictedly, but is also a completely classical conditional, at the inferential level. So, among other things, it interacts correctly with generalizations in contexts in which one uses transparent truth.

One may want to avoid these shortcomings by supplementing the **LP**-based approach with some sort of recovery technique. Among the salient options in this regard we find that of Beall [6], where a device based on Priest's [32] notion of "shrieking" is developed. Beall argues that one may adopt an extra-logical rule of inference—called a shriek rule—allowing one to express by meta-theoretical tools that a particular sentence A is consistent, by adding to the system the rule $A \wedge \neg A \vDash \bot$. Thus, we can recover the full steam of **CL** even for restricted universal quantifications, by applying the shriek rule to the formulae appearing in the antecedent of such quantifications.

---

[18] This idea parallels Zardini's claim in [45], with regard to vagueness phenomena, that the failure of Modus Ponens in **LP** deprives the validity of tolerance principles of their intended force.

[19] Where, of course, $'$ is the successor function and $=$ is the identity relation.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

Of course, it is easy to observe how this technique can be applied to areas like arithmetic, which are, allegedly, free of contradiction. However, it is a little more difficult to see how this mechanism can be applied to sentences involving the truth predicate, which constitute the most interesting cases for us. Presumably, there will be sentences appearing in the antecedent of restricted universal quantifications which feature the truth predicate and are, nevertheless, nonpathological. But there is no principled way to tell them apart from the pathological ones. Thus, the technique cannot be applied with full generality on pain of triviality. Consequently, no restricted universal quantification featuring the truth predicate in its antecedent will be able to detach its consequent. If this is the case, then, sentences like "Everything Aristotle says which is true, is also clear"—formalized as $\forall x((Arist(x) \wedge Tr\langle \dot{x} \rangle) \rightarrow C(x))$—cannot be properly handled inside his account. Much to the contrary, nothing like that happens in our theory, as our conditional is inferentially speaking perfectly classical, even in contexts where the truth-predicate appears. This, we think, is yet another advantage of our account against the remaining alternatives.

In sum, the Goodship Project based on **ST** seems to fare better, if we take into consideration the criterion of minimal mutilation. As Goodship predicted, it is essential to express self-reference through equivalences and not through identities in order to stay away from the troubles caused by paradoxes and, specially, revenge paradoxes. Our work shows that this also holds of substructural theories, when they are supplemented with enough expressive resources such as recovery operators. However, contrary to Goodship's predictions, this move need not be so problematic, for the conditional involved in expressing self-reference must not necessarily be nonclassical, i.e., nondetachable or noncontractive. We can, indeed, stick to Goodship's solution employing a fully classical conditional, *if the underlying logic is nontransitive*.

As a closing remark, let us mention that for those interested in the generality of the nontransitive approach as a solution for many problematic phenomena like semantic paradoxes, vagueness and more,[20] our discussion seems to provide evidence for the fact that the failure of Cut is crucial to keep naïve theories from falling into triviality. Here we showed that the substructural character of nontransitive theories is also tied to the way in which self-reference is expressed within the system.

§5. A sequent calculus. In this section, we introduce a three-sided sequent system for $\mathbf{STTT}^{\circ}_{\mathbf{w}}$ that we will call $\mathbf{LSTTT}^{\circ}_{\mathbf{w}}$, later proving soundness and completeness of $\mathbf{LSTTT}^{\circ}_{\mathbf{w}}$ with respect to $\mathbf{STTT}^{\circ}_{\mathbf{w}}$. We will be implementing a system of three-sided sequent, endowed with a *disjunctive* reading, as suggested in, e.g., [36]. We will first present this for **STTT** and then work our way up to $\mathbf{STTT}^{\circ}_{\mathbf{w}}$.

DEFINITION 5.1. *A three-sided disjunctive sequent* $\Gamma \mid \Sigma \mid \Delta$ *is satisfied by a valuation* v *if and only if* $v(A) = 0$ *for some* $A \in \Gamma$, *or* $v(B) = \frac{1}{2}$ *for some* $B \in \Sigma$, *or* $v(C) = 1$ *for some* $C \in \Delta$. *A sequent is valid if and only if it is satisfied by every valuation. A valuation is a counterexample to a sequent if the valuation does not satisfy the sequent.*

It is easy to see that there is a strong relation between valid disjunctive three-sided sequents and valid inferences of **STTT**: $\vDash_{\mathbf{STTT}} \Gamma \Rightarrow \Delta$ if and only if $\Gamma \mid \Gamma, \Delta \mid \Delta$ is valid.

[20] For example, Cobreros, Egré, Ripley, & van Rooij have pointed out that the nontransitive approach can be applied to some metaphysical issues in [15], whereas [42] analyzes how to apply the approach to Fitch's paradox of knowability.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

This fact follows from the definition of **STTT**'s validity and the definition of validity of a three-sided sequent. Moreover, this facts holds the same for all the expansions of **STTT** that we will present here: all of them will have theirs corresponding three-sided sequent versions.

The proof system we are about to present, called **LSTTT**, includes some axioms and rules, as usual. A sequent is provable if and only if it follows from the axioms by some number (possibly zero) of applications of the rules. As we are working with sets, the effects of the structural rules of Exchange and Contraction are built in, and Weakening is built into the axioms.

We will have three versions of a three-sided Cut rule, and also a Derived Cut rule (that can be inferred from the three basic rules of Cut) and that will also play a key role in the completeness proof presented below, following [37]. Id is the only axiom-scheme of **LSTTT**. Cut 1, Cut 2, Cut 3 and Derived Cut are structural rules. The rest of them are **LSTTT**'s operational rules. Given that the connectives for which we do not give any rules ($\vee$, $\rightarrow$ and $\leftrightarrow$) can be defined in terms of the ones for which we do provide rules, this succinct presentation is just as good.[21]

$$\frac{}{A, \Gamma \mid A, \Sigma \mid A, \Delta}\ \text{Id}$$

$$\frac{\Gamma, A \mid \Sigma \mid \Delta \quad \Gamma \mid \Sigma, A \mid \Delta}{\Gamma \mid \Sigma \mid \Delta}\ \text{Cut1} \qquad \frac{\Gamma \mid \Sigma \mid \Delta, A \quad \Gamma \mid \Sigma, A \mid \Delta}{\Gamma \mid \Sigma \mid \Delta}\ \text{Cut2} \qquad \frac{\Gamma, A \mid \Sigma \mid \Delta \quad \Gamma \mid \Sigma \mid \Delta, A}{\Gamma \mid \Sigma \mid \Delta}\ \text{Cut3}$$

$$\frac{\Gamma, A \mid \Sigma, A \mid \Delta \quad \Gamma \mid \Sigma, A \mid \Delta, A \quad \Gamma, A \mid \Sigma \mid \Delta, A}{\Gamma \mid \Sigma \mid \Delta}\ \text{Derived Cut}$$

$$\frac{\Gamma \mid \Sigma \mid \Delta, A}{\Gamma, \neg A \mid \Sigma \mid \Delta}\ L\neg \qquad \frac{\Gamma \mid \Sigma, A \mid \Delta}{\Gamma \mid \Sigma, \neg A \mid \Delta}\ M\neg \qquad \frac{\Gamma, A \mid \Sigma \mid \Delta}{\Gamma \mid \Sigma \mid \Delta, \neg A}\ R\neg$$

$$\frac{\Gamma, A, B \mid \Sigma \mid \Delta}{\Gamma, A \wedge B \mid \Sigma \mid \Delta}\ L\wedge \qquad \frac{\Gamma \mid \Sigma \mid \Delta, A \quad \Gamma \mid \Sigma \mid \Delta, B}{\Gamma \mid \Sigma \mid \Delta, A \wedge B}\ R\wedge$$

$$\frac{\Gamma \mid \Sigma, A \mid \Delta, A \quad \Gamma \mid \Sigma, B \mid \Delta, B \quad \Gamma \mid \Sigma, A, B \mid \Delta}{\Gamma \mid \Sigma, A \wedge B \mid \Delta}\ M\wedge$$

$$\frac{\Gamma \mid \Sigma \mid \Delta, A}{\Gamma, \mid \Sigma \mid \Delta, Tr\langle A\rangle}\ \text{R}Tr \qquad \frac{\Gamma \mid \Sigma, A \mid \Delta}{\Gamma, \mid \Sigma, Tr\langle A\rangle \mid \Delta}\ \text{M}Tr \qquad \frac{\Gamma, A \mid \Sigma \mid \Delta}{\Gamma, Tr\langle A\rangle \mid \Sigma \mid \Delta}\ \text{L}Tr.$$

Moving forward to our target proof system, $\mathbf{LSTTT}^{\circ}_{\mathbf{w}}$ works exactly like **LSTTT**, with the addition of two rules for the recovery operator

$$\frac{\Gamma, A \mid \Sigma \mid \Delta, A}{\Gamma \mid \Sigma \mid \Delta, \circ A}\ R\circ \qquad \frac{\Gamma \mid \Sigma, A \mid \Delta}{\Gamma, \circ A \mid \Sigma \mid \Delta}\ L\circ$$

[21] This is just a version of the three-sided sequent calculi Ripley presents in [36]. None of the differences are, for what we understand, significant ones. For example, Ripley restricts the identity axioms to atomic formula, while our presentation does not have that restriction. (Actually, every instance of Identity will have a proof in Ripley's system.)

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

and two axioms-schemes for the self-referential sentences, that we will call $PseudoDL_1$ and $PseudoDL_2$. Notice that, in $PseudoDL_1$ and $PseudoDL_2$, the expression p^* (alternatively, A_{p^*}) must be the left (alternatively, right) term of a biconditional $p^* \leftrightarrow A_{p^*}$ that is a member of the set PseudoDL.

$$\frac{}{\Gamma, p^* \mid \Sigma, p^*, A_{p^*} \mid \Delta, A_{p^*}}\ PseudoDL_1 \qquad \frac{}{\Gamma, A_p^* \mid \Sigma, p^*, A_{p^*} \mid \Delta, p^*}\ PseudoDL_2.$$

Let us comment briefly on the semantic interpretation of the rules $PseudoDL_1$ and $PseudoDL_2$. In a nuthsell, the validity of these rules is equivalent to the satisfaction of all the members of the set PseudoDL. In other words, the validity of these rules is equivalent with there being pathological sentences around when working with $\mathbf{LSTTT}^{\circ}_{\mathbf{w}}$—a requirement indispensable for our nontriviality result to be meaningful.[22]

Finally, the following are the main results concerning $\mathbf{LSTTT}^{\circ}_{\mathbf{w}}$ and $\mathbf{STTT}^{\circ}_{\mathbf{w}}$.

THEOREM 5.2 (Soundness). *If a sequent* $\Gamma \mid \Sigma \mid \Delta$ *is provable in* $\mathbf{LSTTT}^{\circ}_{\mathbf{w}}$*, then it is valid in* $\mathbf{LSTTT}^{\circ}_{\mathbf{w}}$.

Proof. The axioms are valid, and validity is preserved by the rules, as can be checked without too much trouble. □

THEOREM 5.3 (Completeness). *If a sequent* $\Gamma \mid \Sigma \mid \Delta$ *is valid in* $\mathbf{LSTTT}^{\circ}_{\mathbf{w}}$*, then it is provable in* $\mathbf{LSTTT}^{\circ}_{\mathbf{w}}$.

Proof. In the Appendix. □

§6. Conclusion. Throughout this article we have shown how to recover the safe instances of Cut and of all the classically valid meta-inferences that are invalid in the nontransitive transparent theory of truth **STTT**, with the help of a unary recovery operator $\circ A$. We have defended that the best way to do this is by expressing self-reference through equivalences, thus placing ourselves in the context of the so-called Goodship Project. However, even if our current approach can be located within the Goodship Project, the point of view and the results we present are quite novel to this work. This is so because Laura Goodship suggested that, to achieve a nontrivial naive theory, it was essential to express all the self-referential sentences through nonclassical biconditionals—invalidating either Modus Ponens or Contraction. Contrary to this, we showed how this can be properly achieved by using a biconditional that is completely classical, inferentially speaking, provided the underlying logic is allowed to be nontransitive. Finally, we presented a three-sided sequent calculus which was shown to be sound and complete with regard to our target theory $\mathbf{STTT}^{\circ}_{\mathbf{w}}$.

[22] This can be, furthermore, observed by noting that these rules are interderivable with the rule

$$\frac{}{\Gamma \mid \Sigma, p^* \leftrightarrow A_{p^*} \mid \Delta, p^* \leftrightarrow A_{p^*}}$$

where, yet again, the sentences of the form $p^* \leftrightarrow A_{p^*}$ are members of the set PseudoDL. Notwithstanding this fact, this version of our system will be less elegant that the one that we have presented above. This new axiom-scheme will include a logical constant—a biconditional—whereas the rules $PseudoDL_1$ and $PseudoDL_2$ can be considered as structural, because no logical constant appears in them.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

§7. Appendix: Completeness proof. We will use the method of reduction trees,[23] that allows to build for any given sequent, either a proof of that sequent, or a counterexample to it. The method also provides of a way of building the potential counterexample. We will introduce the notions of subsequent and sequent union, that will be used in the proof:

DEFINITION 7.1. *A sequent* $S = \Gamma \mid \Sigma \mid \Delta$ *is a subsequent of a sequent* $S' = \Gamma' \mid \Sigma' \mid \Delta'$ (*written* $S \sqsubseteq S'$) *if and only if* $\Gamma \sqsubseteq \Gamma'$, $\Sigma \sqsubseteq \Sigma'$, *and* $\Delta \sqsubseteq \Delta'$.

DEFINITION 7.2. *A sequent* $S = \Gamma \mid \Sigma \mid \Delta$ *is the sequent union of a set of sequents* $[\Gamma_i \mid \Sigma_i \mid \Delta_i]_{i \in I}$ (*written* $S = \sqcup[\Gamma_i \mid \Sigma_i \mid \Delta_i]_{i \in I}$) *if and only if* $\Gamma = \sqcup_{i \in I} \Gamma_i$, $\Sigma = \sqcup_{i \in I} \Sigma_i$ *and* $\Delta = \sqcup_{i \in I} \Delta_i$.

The construction starts from a root sequent $S_0 = \Gamma_0 \mid \Sigma_0 \mid \Delta_0$, and then builds a tree in stages, applying at each stage all the operational rules that can be applied, plus Derived Cut "in reverse", i.e., from the conclusion sequent to the premise(s) sequent(s).[24] Thus, stage 0 will just be the root sequent S_0. If it is an axiom, the branch is closed. For any stage $n + 1$, one of two following things might happen:

1. For all branches in the tree after stage n, if the tip—e.g., the sequent that is being reduced—is an axiom, the branch is closed.
2. For open branches: For each formula A in a sequent position in each open branch, if A already occurred in that sequent position in that branch (i.e., A has not been generated during stage $n + 1$), and A has not already been reduced during stage $n + 1$, then reduce A as is shown below. There are three possible positions in which a formula can appear in a sequent: either (i) on the left side, or (ii) on the middle, or (iii) on the right side. We need to consider all these possible cases.

- If A is a negation $\neg B$, then: if A is in the left/ middle/ right position, extend the branch by copying its current tip and adding B to the right/ middle/ left position.
- If A is a conjunction $B \wedge C$, then: (i) if A is in the left position, extend the branch by copying its current tip and adding both B and C to the left position. (ii) If A is in the middle position, split the branch in three: extend the first by copying the current tip and adding B to both the middle and right positions; extend the second by copying the current tip and adding C to the middle and right positions; and extend the third

[23] For similar proofs, see [30] and [36].

[24] For the proof, we use an enumeration of the formulae and an enumeration of names. We will reduce, at each stage, all the formulae in the sequent, starting from the one with the lowest number, then continuing with the formula with the second lowest number, and moving on in this way until the formula with the highest number in the sequent is reduced. In the case a formula appears in more than one side of the sequent, we will start by reducing the formula that appears on the left side and then proceed to the middle and the right side, respectively. The final step, at each stage n of the reduction process, will be an application of the Derived Cut rule to the nth-formula in the enumeration. If we apply a multi-premise rule, we will generate more branches that will need to be reduced. If we apply a single-premise rule, we just extend the branch with one more leave. We will only add formulae at each stage, without erasing any of them. As a result of the process just described, every branch will be ordered by the subsequent relation. Any branch that has an axiom as it topmost sequent will be closed. A branch that is not closed is considered open. This procedure is repeated until every branch is closed, or until there is an infinite open branch. If every branch is closed, then the resulting tree itself is a proof of the root sequent. If there is an infinite open branch Y, we can use it to build a counterexample to the root sequent.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

by copying the current tip and adding both B and C to the middle position. (iii) If A is in the right position, split the branch in two: extend the first by copying the current tip and adding B to the right position; and extend the second by copying the current tip and adding C to the right position.

- If A is a formula $\circ B$, then: (i) if A is in the left position, extend the branch by copying its current tip and adding B to the middle position. (ii) If A is in the right position, extend the branch by copying its current tip and adding B to the right and left positions. (iii) If A is in the middle position, then do nothing.
- If A is a formula $Tr\langle B\rangle$, then if A is in the left / middle / right position, extend the branch by copying its current tip and adding B to the left / middle / right position.[25]

We will also apply the Derived Cut rule at each step. Consider the nth formula in the enumeration of formulae and call it A. Now extend each branch using the Derived Cut rule. For each open branch, if its tip is $\Gamma \mid \Sigma \mid \Delta$, split it in three and extend the new branches with the sequent $\Gamma, A \mid \Sigma, A \mid \Delta$, the sequent $\Gamma, A \mid \Sigma \mid \Delta, A$, and the sequent $\Gamma \mid \Sigma, A \mid \Delta, A$, respectively.

Now we need to repeat this procedure until every branch is closed, or, if that does not happen, until there is an infinite open branch. If the first scenario is the actual one, then the tree itself is a proof of the root sequent, because each step will be the result of an application of a structural or operational rule to the previous steps. If the second scenario is the actual one, we can use the infinite open branch to build a counterexample.[26]

If in fact there is an infinite open branch Y, then the Derived Cut rule will have been used infinitely many times. Thus, every formula will appear at some point in the branch for the first time, and will remain in every step afterwards. Now, we first collect all sequents of the infinite open branch Y into one single sequent $S_\omega = \Gamma_\omega \mid \Sigma_\omega \mid \Delta_\omega = \sqcup \{S \mid S \text{ is a sequent of } Y\}$. Notice that, as Derived Cut has been applied infinitely many times in the construction of the branch, every formula will occur in exactly two places in S_ω.[27] Thus, there will be a valuation such that no formula in the sequent gets the value associated with the place where it occurs (i.e., 0 if the formula occurs in the left, $\frac{1}{2}$ if it occurs in the middle, 1 if it occurs in the right). Hence, for each formula A in the sequent, v will give to A a value different from the ones corresponding to the sides where A appears in the sequent. But that includes all the formulae in the initial and finite sequent S_0. That valuation, then, will also

[25] As formulae that are disjunctions, conditionals and biconditionals can be defined in terms of negations and conjunctions (remember that the conditional is just the material one), it will not be necessary to specify special rules for them. Those cases will be subsumed in the ones already specified.

[26] As we have already mentioned, *the tip* is the sequent that is being reduced. If it is an axiom, then this branch will be closed. If it is not, then the reduction process will go on. We will present a toy example to understand exactly what *the tip* is. Take, for instance, the sequent $\oslash \mid \neg B \mid \neg A$. In this case, this is *the tip* we have mentioned. The first step of the reduction will give us the sequent $\oslash \mid B, \neg B \mid \neg A$. But this is not an axiom either. So the reduction process will keep on rolling. $\oslash \mid B, \neg B \mid \neg A$ is the new *tip*, and the next step of the reduction will give us the sequent $A \mid B, \neg B \mid \neg A$. Finally, as this sequent is not an axiom either, we will split this new *tip* in three when we apply the Derived Cut rule in reverse, and the reduction process will continue. Thanks to an anonymous referee for pushing us to clarify this point.

[27] It cannot occur in the three places, because then there will be some finite stage n where the formula appears for the first time in the branch in the three sides. But then that sequent will be an axiom, and therefore the branch will be closed.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

be a counterexample to S_0. Therefore that valuation will be a counterexample to the sequent being considered.

Thus, for atomic formulae A (propositional letters and truth assertions), $v(A) = 0$ or $\frac{1}{2}$ or 1, respectively, if and only if A does not appear in Γ_ω or Σ_ω or Δ_ω, respectively.[28]

The rules for reducing formulae can be used to show by induction, that, if none of the components of complex formulae (negations, conjunctions, disjunctions, conditionals, biconditional, consistency and truth assertions) receive the value associated with any place in which they appear in S_ω, neither will the compound. We will not see, due to limitations of space, how this method works in detail. For conjunctions, negations and truth assertions, we proceed exactly as is shown in [36]. The new cases are that of formulae $\circ A$. Thus, we will just check how that type of assertion can be reduced.

In the cases of formulae of the form $\circ A$, no formula like this can appear both on the left and the right side, because then they would have appeared for the first time in the same leaf of the branch. When you reduced them at the next stage, you'll get a sequent with A at the three sides of the sequent. That sequent will be an axiom, and so will not be part of an infinite open branch. So (i) either $\circ A$ is both in the left and the middle sides of the sequent, or (ii) it is both in the middle and the right side of the sequent.

Let us start with (i). At some point, $\circ A$ will be reduced. It appears in the middle side of the sequent, so nothing is supposed to be done when this happens. But it also appears on the left side. If that happens, then A will appear on the middle side of the sequent on the next stage of the construction. At some point, an application of Derived Cut will also introduce A in the left or in the right side, and so they will also appear in that place in S_ω. If A appears on the left, then A will get value 1, an also will $\circ A$. If it appears on the right, then it will get value 0, and $\circ A$ will get value 1. Thus, in none of this cases $\circ A$ receives a value associated with one the sides where it appears. Now consider (ii). Once $\circ A$ is reduced, as it is on the right side of the sequent, that will get A on both the left and the right sides. So by inductive hypothesis, A will get value $\frac{1}{2}$, and $\circ A$ will get value 0. So, again, $\circ A$ will not receive a value associated with the sides where it appears.

By completing the induction along these lines, we can show that we can construct a valuation such that no formula receives the value associated with any place where it appears in S_ω. But, as we know, that includes all the formulae in the initial and finite sequent S_0. That valuation, then, will also be a counterexample to S_0, which is what we were looking for. Thus, for any sequent S, either it has a proof or it has a counterexample.

§8. Acknowledgments. The material included in this article has been presented at the Logic and Metaphysics Seminar in CUNY, at the 3rd CLE - Buenos Aires Logic Group Workshop in Buenos Aires, and at the 2nd Workshop on Dialetheism and Paraconsistency in Kyoto. Thanks to the audiences of these events for their comments and helpful feedback. In addition, we are thankful to thank Rohan French, Hitoshi Omori, Graham Priest,

[28] Does A appears in exactly the places where $Tr\langle A\rangle$ appears? Yes. As any formula in sequent that corresponds to an infinite open branch, A appears in exactly two places. If $Tr\langle A\rangle$) appears in the only place where A does not appear, then, as $Tr\langle A\rangle$ will eventually be reduced, A will appear in the only place where it does not appear until that moment in branch. But then that sequent will be an axiom, and thus the branch will be closed. This is the only possibility that we need to consider. $Tr\langle A\rangle$ cannot appear in less places that A: as any formula in a sequent corresponding to an infinite open branch, it has to appear in exactly two places.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

Dave Ripley, and the members of the Buenos Aires Logic Group for discussing previous versions of this work. Finally, we would like to express our gratitude to two anonymous reviewers for this journal, whose suggestions helped to substantially improve the article.

BIBLIOGRAPHY

[1] Barrio, E., Pailos, F., & Szmuc, D. (2017). A paraconsistent route to semantic closure. *Logic Journal of the IGPL*, **25**(4), 387–407.

[2] Barrio, E., Pailos, F., & Szmuc, D. (2018). What is a paraconsistent logic? In Carnielli, W. and Malinowski, J., editors. *Between Consistency and Inconsistency*. Trends in Logic. Dordrecht: Springer, pp. 89–108.

[3] Barrio, E., Rosenblatt, L., & Tajer, D. (2015). The logics of strict-tolerant logic. *Journal of Philosophical Logic*, **44**(5), 551–571.

[4] Batens, D. (1989). Dynamic dialectical logics. In Priest, G., Sylvan, R., and Norman, J., editors. *Paraconsistent Logic. Essays on the Inconsistent*. München: Philosophia, pp. 187–217.

[5] Beall, J. C. (2011). Multiple-conclusion LP and default classicality. *Review of Symbolic Logic*, **4**(2), 326–336.

[6] Beall, J. C. (2013). Shrieking against gluts: The solution to the 'just true' problem. *Analysis*, **73**(3), 438–445.

[7] Beall, J. C. (2014). Finding tolerance without gluts. *Mind*, **123**(491), 791–811.

[8] Blasio, C., Marcos, J., & Wansing, H. (2018). An inferentially many-valued two-dimensional notion of entailment. *Bulletin of the Section of Logic*, **46**, 233–262.

[9] Carnielli, W. & Coniglio, M. (2016). *Paraconsistent Logic*: *Consistency, Contradiction and Negation*. Dordrecht: Springer.

[10] Carnielli, W., Coniglio, M., & Marcos, J. (2007). Logics of formal inconsistency. In Gabbay, D. and Guenthner, F., editors. *Handbook of Philosophical Logic*, Vol. 14. Dordrecht: Springer, pp. 1–93.

[11] Carnielli, W. & Marcos, J. (2002). A taxonomy of C-systems. In Carnielli, W., Coniglio, M., and D'Ottaviano, I., editors. *Paraconsistency: The Logical Way to the Inconsistent*. New York: Marcel Dekker, pp. 1–94.

[12] Carnielli, W., Marcos, J., & De Amo, S. (2004). Formal inconsistency and evolutionary databases. *Logic and Logical Philosophy*, **8**, 115–152.

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

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

[15] Cobreros, P., Egré, P., Ripley, D., & van Rooij, R. (2013). Identity, Leibniz's law and non-transitive reasoning. *Metaphysica*, **14**(2), 253–264.

[16] Cobreros, P., Egré, P., Ripley, D., & van Rooij, R. (2014). Reaching transparent truth. *Mind*, **122**(488), 841–866.

[17] Cook, R. T. (2004). Patterns of paradox. *Journal of Symbolic Logic*, **69**(3), 767–774.

[18] Cook, R. T. (2005). What's wrong with tonk (?). *Journal of Philosophical Logic*, **34**(2), 217–226.

[19] Da Costa, N. (1974). On the theory of inconsistent formal systems. *Notre Dame Journal of Formal Logic*, **15**(4), 497–510.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

[20] Da Costa, N. (1993). *Sistemas Formais Inconsistentes* (*Inconsistent Formal Systems, in Portuguese*). Ph.D. Thesis, Universidade Federal do Paraná, Curitiba, Brazil, 1963.

[21] Dicher, B. & Paoli, F. (forthcoming). ST, LP, and tolerant metainferences. In Başkent, C. and Ferguson, T. M., editors. *Graham Priest on Dialetheism and Paraconsistency*. Dordrecht: Springer.

[22] Fjellstad, A. (2016). Omega-inconsistency without cuts and nonstandard models. *The Australasian Journal of Logic*, **13**(5), 96–122.

[23] Frankowski, S. (2004). Formalization of a plausible inference. *Bulletin of the Section of Logic*, **33**(1), 41–52.

[24] Frankowski, S. (2004). *p*-Consequence versus *q*-Consequence operations. *Bulletin of the Section of Logic*, **33**(4), 197–207.

[25] Goodship, L. (1996). On dialethism. *Australasian Journal of Philosophy*, **74**(1), 153–161.

[26] Hjortland, O. (2017). Theories of truth and the maxim of minimal mutilation. *Synthese*. Advance online publication, doi: 10.1007/s11229-017-1612-8.

[27] Horsten, L. (2009). Levity. *Mind*, **118**(471), 555–581.

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

[29] Omori, H. (2015). Remarks on naive set theory based on LP. *Review of Symbolic Logic*, **8**(2), 279–295.

[30] Paoli, F. (2013). *Substructural Logics*: *A Primer*. Dordrecht: Springer.

[31] Picollo, L. (2018). Reference in arithmetic. *Review of Symbolic Logic*, **11**(3), 573–603.

[32] Priest, G. (2006). *In Contradiction*: *A Study of the Transconsistent*. Oxford: Oxford University Press.

[33] Priest, G. (2014). *One*: *Being an Investigation into the Unity of Reality and of its Parts, including the Singular Object which is Nothingness*. Oxford: Oxford University Press.

[34] Priest, G. (2017). What if? The exploration of an idea. *The Australasian Journal of Logic*, **14**(1), 54–127.

[35] Pynko, A. (2010). Gentzen's cut-free calculus versus the logic of paradox. *Bulletin of the Section of Logic*, **39**(1/2), 35–42.

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

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

[38] Scott, D. (1971). On engendering an illusion of understanding. *Journal of Philosophy*, **68**(21), 787–807.

[39] Sharvit, Y. (2017). A note on (Strawson) entailment. *Semantics and Pragmatics*, **10**(1), 1–38, 6.

[40] Tennant, N. (1982). Proof and paradox. *Dialectica*, **36**(2–3), 265–296.

[41] Tennant, N. (2015). A new unified account of truth and paradox. *Mind*, **124**(494), 571–605.

[42] van Rooij, R. (2017). Nonmonotonicity and knowability: As knowable as possible. In Başkent, C., Moss, L. S., and Ramanujam, R., editors. *Rohit Parikh on Logic, Language and Society*. Cham: Springer, pp. 53–65.

[43] von Fintel, K. (1999). NPI licensing, Strawson entailment, and context dependency. *Journal of Semantics*, **16**(2), 97–148.

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press

[44] Weir, A. (2005). Naive truth and sophisticated logic. In Beall, J. C. and Armour-Garb, B., editors. *Deflationism and Paradox*. Oxford: Oxford University Press, pp. 218–249.

[45] Zardini, E. (2008). A model of tolerance. *Studia Logica*, **90**(3), 337–368.

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF BUENOS AIRES
PUAN 480, C1420 BUENOS AIRES
ARGENTINA
and
IIF-SADAF (CONICET)
BULNES 642, C1176ABL BUENOS AIRES
ARGENTINA
E-mail: eabarrio@gmail.com
E-mail: fpailos@hotmail.com
E-mail: szmucdamian@gmail.com

https://doi.org/10.1017/S1755020318000369 Published online by Cambridge University Press