Journal of Philosophical Logic (2022) 51:201–218
https://doi.org/10.1007/s10992-021-09625-5

Non-reflexivity and Revenge

Julien Murzi[1] · **Lorenzo Rossi**[2]

Received: 12 June 2020 / Accepted: 19 June 2021 / Published online: 12 November 2021

Abstract
We present a revenge argument for non-reflexive theories of semantic notions – theories which restrict the rule of assumption, or (equivalently) initial sequents of the form $\varphi \vdash \varphi$. Our strategy follows the general template articulated in Murzi and Rossi [21]: we proceed via the definition of a notion of paradoxicality for non-reflexive theories which in turn breeds paradoxes that standard non-reflexive theories are unable to block.

Keywords Liar paradox · Curry's Paradox · Validity Curry · Non-reflexive logics · Revenge

Jc Beall and Julien Murzi [2] argue that, if the semantic paradoxes are to be solved via a revision of classical logic, paradoxes of *naïve validity* such as the Validity-Curry require a restriction of one of the classical *structural rules*.[1] Building on [23], [22] consider the possibility of restricting certain instances of *structural reflexivity*

$$\frac{}{\varphi \vdash \varphi}\ \mathsf{SRef}$$

as a way of treating the paradoxes of naïve validity, and the semantic paradoxes more generally. This paves the way for a consistent and broadly Kripkean account of naïve

[1] For more discussion of Beall and Murzi's argument, see [8] and [22].

✉ Julien Murzi
j.murzi@gmail.com

Lorenzo Rossi
lorenzo.rossi@lrz.uni-muenchen.de

[1] University of Salzburg, Salzburg, Austria

[2] Munich Center for Mathematical Philosophy, LMU, München, Germany

validity, *grounded validity*, as Carlo Nicolai and Lorenzo Rossi call it. In this note, we show that a broad family of semantic theories based on *non-reflexive logics*, including Nicolai and Rossi's theory of grounded validity, breed revenge paradoxes they are unable to block.[2] The revenge strategy we follow is but an instance of the general revenge argument discussed in [21]. In a nutshell, the strategy is this. Non-classical theories such as Nicolai and Rossi's restrict the application of classical principles to paradoxical sentences. However, there exist finitely many principles $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ such that a sentence satisfies $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ just in case it satisfies all the principles of classical logic. This allows one to characterise *paradoxical* sentences as the ones that only satisfy $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ on pain of triviality. Unsurprisingly perhaps, such a notion of paradoxicality – which articulates a basic *fact* about non-classical theories – in turn gives rise to new paradoxes that can't be blocked by restricting, in the non-reflexive case, instances of SRef.

1 Paradoxes of Truth and Naïve Validity

Let $\mathcal{L}$ be a first-order language encoding a minimal amount of syntax theory. More specifically, we require that the following two requirements are satisfied:

(i) There is a function $\ulcorner\urcorner$ such that for every sentence φ, $\ulcorner\varphi\urcorner$ is a closed term – informally, a name of φ.
(ii) For every open formula $\varphi(x)$ there is a term t_φ such that $\ulcorner\varphi(t_\varphi/x)\urcorner$ denotes the same element as t_φ, in a fixed acceptable model of the language, where $\varphi(t_\varphi/x)$ is the result of replacing every occurrence of x with t_φ in φ.

Theories that are rich enough to satisfy (i) and (ii) include, for instance, first-order arithmetic in the standard signature $\{0, S, +, \times\}$ plus primitive symbols and defining formulae for all the primitive recursive functions.[3] We let $\mathcal{L}_{\mathsf{Tr}}$ be $\mathcal{L}$ plus a primitive unary predicate Tr, for 'is true'.

We now specify the rules of classical propositional logic we will employ (henceforth, CPL).[4] We formulate them in a sequent-style natural deduction calculus where structural rules are explicitly formulated (see [25, Ch. 2]). A *sequent* is an expression

[2] Other non-reflexive approaches to truth and paradox have been developed, for instance, in [10, 19], and [9]. The approaches discussed in those papers are somewhat different from the one we discuss (for instance, [9] formulates his theory via many-sided sequents). However, the revenge paradox to be developed in what follows uniformly applies to all the theories (and subtheories) satisfying the (minimal) requirements formulated in Definition 1, Definition 4, and that are encoded in the proof of Proposition 5.

[3] See [12] for more details.

[4] Since our main result only requires propositional logical rules, this suffices for the purposes of this paper. A double line indicates that a rule can be read in both directions. Our axiomatization is redundant, but simplifies the subsequent derivations.

of the form $\Gamma \vdash \varphi$, where Γ is a finite multiset of sentences. A rule is an *inference* if its premises are empty, and a *meta-inference* otherwise.

$$\frac{}{\varphi \vdash \varphi}\ \mathsf{SRef} \qquad \frac{\Gamma \vdash \chi}{\Gamma, \varphi \vdash \chi}\ \mathsf{SWeak} \qquad \frac{\Gamma, \varphi, \varphi \vdash \chi}{\Gamma, \varphi \vdash \chi}\ \mathsf{SContr}$$

$$\frac{\Gamma \vdash \varphi \qquad \Delta, \varphi \vdash \psi}{\Gamma, \Delta \vdash \psi}\ \mathsf{Cut}$$

$$\frac{\Gamma \vdash \varphi \qquad \Delta \vdash \psi}{\Gamma, \Delta \vdash \varphi \wedge \psi}\ \wedge\text{-I} \qquad \frac{\Gamma \vdash \varphi \wedge \psi}{\Gamma \vdash \varphi}\ \wedge\text{-E}_1 \qquad \frac{\Gamma \vdash \varphi \wedge \psi}{\Gamma \vdash \psi}\ \wedge\text{-E}_2$$

$$\frac{\Gamma \vdash \varphi}{\Gamma \vdash \varphi \vee \psi}\ \vee\text{-I}_1 \qquad \frac{\Gamma \vdash \psi}{\Gamma \vdash \varphi \vee \psi}\ \vee\text{-I}_2$$

$$\frac{\Gamma \vdash \varphi \vee \psi \qquad \Delta_0, \varphi \vdash \chi \qquad \Delta_1, \psi \vdash \chi}{\Gamma, \Delta_0, \Delta_1 \vdash \chi}\ \vee\text{-E}$$

$$\frac{\Gamma, \varphi \vdash \psi}{\Gamma \vdash \varphi \to \psi}\ \to\text{-I} \qquad \frac{\Gamma \vdash \varphi \qquad \Delta \vdash \varphi \to \psi}{\Gamma, \Delta \vdash \psi}\ \to\text{-E}$$

$$\frac{\Gamma \vdash \varphi \to \psi}{\Gamma \vdash \neg\psi \to \neg\varphi}\ \mathsf{Contr}$$

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

In order to present the Liar and other truth-theoretic paradoxes, we use the following introduction and elimination rules for naïve truth (for convenience, we assume both positive and negative forms):

$$\frac{\Gamma \vdash \varphi}{\Gamma \vdash \mathsf{Tr}(\ulcorner\varphi\urcorner)}\ \mathsf{Tr}\text{-I} \qquad \frac{\Gamma \vdash \mathsf{Tr}(\ulcorner\varphi\urcorner)}{\Gamma \vdash \varphi}\ \mathsf{Tr}\text{-E} \qquad \frac{\Gamma \vdash \neg\varphi}{\Gamma \vdash \neg\mathsf{Tr}(\ulcorner\varphi\urcorner)}\ \neg\text{-}\mathsf{Tr}\text{-I}$$

$$\frac{\Gamma \vdash \neg\mathsf{Tr}(\ulcorner\varphi\urcorner)}{\Gamma \vdash \neg\varphi}\ \neg\text{-}\mathsf{Tr}\text{-E}$$

Now let $\neg\mathsf{Tr}(t_\lambda)$ be a sentence – a liar sentence – s.t. $t_\lambda = \ulcorner\neg\mathsf{Tr}(t_\lambda)\urcorner$ and let λ be a shorthand for $\neg\mathsf{Tr}(t_\lambda)$. We may then (paradoxically) reason thus. We first prove $\mathsf{Tr}(\ulcorner\lambda\urcorner) \vdash \bot$:

$$\dfrac{\dfrac{}{\mathsf{Tr}(\ulcorner\lambda\urcorner) \vdash \mathsf{Tr}(\ulcorner\lambda\urcorner)}\ \mathsf{SRef} \qquad \dfrac{\dfrac{\dfrac{}{\mathsf{Tr}(\ulcorner\lambda\urcorner) \vdash \mathsf{Tr}(\ulcorner\lambda\urcorner)}\ \mathsf{SRef}}{\mathsf{Tr}(\ulcorner\lambda\urcorner) \vdash \lambda}\ \mathsf{Tr}\text{-E}}{\mathsf{Tr}(\ulcorner\lambda\urcorner) \vdash \neg\mathsf{Tr}(\ulcorner\lambda\urcorner)}\ \text{Definition of } \lambda}{\dfrac{\mathsf{Tr}(\ulcorner\lambda\urcorner), \mathsf{Tr}(\ulcorner\lambda\urcorner) \vdash \bot}{\mathsf{Tr}(\ulcorner\lambda\urcorner) \vdash \bot}\ \mathsf{SContr}}\ \neg\text{-E}$$

Call the above derivation $\mathcal{D}_1$. We then derive $\mathsf{Tr}(\ulcorner\lambda\urcorner)$ from $\mathcal{D}_1$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\mathcal{D}_1}{\mathsf{Tr}(\ulcorner\lambda\urcorner) \vdash \bot}}{\vdash \neg\mathsf{Tr}(\ulcorner\lambda\urcorner)}\ \neg\text{-I}}{\vdash \lambda}\ \text{Definition of } \lambda}{\vdash \mathsf{Tr}(\ulcorner\lambda\urcorner)}\ \mathsf{Tr}\text{-I}$$

Call this derivation $\mathcal{D}_2$. $\mathcal{D}_1$ and $\mathcal{D}_2$ can now be combined together to yield a proof of absurdity, courtesy of Cut:

$$\dfrac{\begin{array}{cc}\mathcal{D}_2 & \mathcal{D}_1 \\ \vdash \mathsf{Tr}(\ulcorner\lambda\urcorner) & \mathsf{Tr}(\ulcorner\lambda\urcorner) \vdash \bot\end{array}}{\vdash \bot}\ \mathsf{Cut}$$

Given $\bot$-E, the foregoing Liar reasoning yields a proof of *any sentence* φ, thus trivialising any theory in which the paradox can be derived.

Triviality can also be established without making use of $\bot$-E, via Curry's Paradox [6]. The paradox involves a sentence $\mathsf{Tr}(t_\kappa) \to \psi$ (where ψ is any sentence) s.t. $t_\kappa = \ulcorner \mathsf{Tr}(t_\kappa) \to \psi \urcorner$. Now let κ etc. be shorthand for $\mathsf{Tr}(t_\kappa) \to \psi$. We can then essentially replicate the Liar reasoning to yield a 'proof' of ψ. We first 'prove' $\mathsf{Tr}(\ulcorner\kappa\urcorner) \vdash \psi$ – call this derivation derivation $\mathcal{D}_3$:

$$\dfrac{\dfrac{\dfrac{}{\mathsf{Tr}(\ulcorner\kappa\urcorner) \vdash \mathsf{Tr}(\ulcorner\kappa\urcorner)}\ \mathsf{SRef}}{\mathsf{Tr}(\ulcorner\kappa\urcorner) \vdash \kappa}\ \mathsf{Tr\text{-}E} \qquad \dfrac{}{\mathsf{Tr}(\ulcorner\kappa\urcorner) \vdash \mathsf{Tr}(\ulcorner\kappa\urcorner)}\ \mathsf{SRef}}{\mathsf{Tr}(\ulcorner\kappa\urcorner) \vdash \mathsf{Tr}(\ulcorner\kappa\urcorner) \to \psi}\ \text{Definition of } \kappa \qquad \to\text{-E}}{\dfrac{\mathsf{Tr}(\ulcorner\kappa\urcorner), \mathsf{Tr}(\ulcorner\kappa\urcorner) \vdash \psi}{\mathsf{Tr}(\ulcorner\kappa\urcorner) \vdash \psi}\ \mathsf{SContr}}$$

We then use $\mathcal{D}_3$ to derive $\mathsf{Tr}(\ulcorner\lambda\urcorner)$:

$$\dfrac{\dfrac{\dfrac{\begin{array}{c}\mathcal{D}_3 \\ \mathsf{Tr}(\ulcorner\kappa\urcorner) \vdash \psi\end{array}}{\vdash \mathsf{Tr}(\ulcorner\kappa\urcorner) \to \psi}\ \to\text{-I}}{\vdash \kappa}\ \text{Definition of } \kappa}{\vdash \mathsf{Tr}(\ulcorner\kappa\urcorner)}\ \mathsf{Tr\text{-}I}$$

Call this derivation $\mathcal{D}_4$. $\mathcal{D}_3$ and $\mathcal{D}_4$ can now be combined together to yield a proof of absurdity, courtesy of Cut:

$$\dfrac{\begin{array}{cc}\mathcal{D}_4 & \mathcal{D}_3 \\ \vdash \mathsf{Tr}(\ulcorner\kappa\urcorner) & \mathsf{Tr}(\ulcorner\kappa\urcorner) \vdash \psi\end{array}}{\vdash \psi}\ \mathsf{Cut}$$

On a revisionary approach to paradox,[5] naïve principles such as Tr-I and Tr-E are non-negotiable: they're both required in order for the truth predicate to play its *inferential role*. For instance, the following reasoning is usually taken to motivate the unrestricted rule of Tr-E:

(AGREEMENT) All the theorems of Peano Arithmetic (PA) are true. φ is a theorem of PA. Hence, φ is true. Therefore (by Tr-E), φ.

Similar kinds of reasoning are standardly cited in support of a naïve conception of truth and, therefore, of a revisionary approach to paradox.

[5] See e.g. [7], [1, §§1.1-1.2], [5, 13].

Several revisionary theories of truth modify the logical rules in order to non-trivially retain naïve truth-theoretical principles, such as Tr-I, Tr-E, ¬-Tr-I, ¬-Tr-E. In light of the Liar and Curry's Paradox, one might think that revising *operational* rules is enough to avoid semantic paradox altogether. Indeed, if one's only naïve semantic notion is truth, then provably non-trivial approaches are available that only restrict logical rules. However, truth is not the only meaningful semantic notion.

In particular, *substructural* approaches to the Liar and the semantic paradoxes, i.e. approaches that restrict one of the classical structural rules, are required if one's semantic theory is to be rich enough to express a naïve conception of validity, obeying the following principles:

(*Validity Proof*)
If ψ follows from φ, then the argument from φ to ψ is naïvely valid.
(*Validity Detachment*)
ψ follows from φ and the claim that the argument from φ to ψ is naïvely valid.

Both principles are intuitively compelling. What is more, they can also be motivated by the need to model reasonings involving expressions of agreement and disagreement, just like the principles for naïve truth. For instance, the following reasoning can be taken to motivate VD (see [22] for more details):

(AGREEMENT*) All the inferences performed by Mary are naïvely valid. Mary infers χ from φ and ψ. Both φ and ψ hold. Therefore (by Validity Detachment), χ also holds.

However, as in the case of truth, *Validity Proof* and *Validity Detachment* give rise to a version of Curry's Paradox, called v-Curry, that only employs structural and semantic principles ([2]).[6]

To see this, let $\mathsf{Val}(x, y)$ be understood as 'the argument from x to y is naïvely valid'. Then, the above principles can be respectively formalized as follows:

$$\dfrac{\varphi \vdash \psi}{\vdash \mathsf{Val}(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)}\ \mathsf{VP} \qquad \dfrac{\Gamma \vdash \mathsf{Val}(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \qquad \Delta \vdash \varphi}{\Gamma, \Delta \vdash \psi}\ \mathsf{VDm}.$$

We can now reason thus. Let π be a sentence identical to $\mathsf{Val}(\ulcorner\pi\urcorner, \ulcorner\bot\urcorner)$ (so that π says of itself that it entails absurdity), and let $\mathcal{D}_5$ be the following derivation of $\mathsf{Val}(\ulcorner\pi\urcorner, \ulcorner\bot\urcorner)$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\pi \vdash \pi}\ \mathsf{SRef}}{\pi \vdash \mathsf{Val}(\ulcorner\pi\urcorner, \ulcorner\bot\urcorner)}\ \text{Def. of } \pi \qquad \dfrac{}{\pi \vdash \pi}\ \mathsf{SRef}}{\pi, \pi \vdash \bot}\ \mathsf{VDm}}{\pi \vdash \bot}\ \mathsf{SContr}}{\vdash \mathsf{Val}(\ulcorner\pi\urcorner, \ulcorner\bot\urcorner)}\ \mathsf{VP}$$

[6] Note that our formalization of 'naïvely valid' in AGREEMENT* as a binary predicate is only one amongst several options. Another possibility is to formalize it as a unary predicate, taking arguments which (in the relevant applications) code arguments, i.e. codes of pairs of sentences. Alternatively, yet another option is to formalize it as a predicate applying to codes of pairs of multisets, for more structured arguments (this is done, for instance, in [23]). Nothing crucial hinges on this choice, however: in each of these cases, one can formulate analogues of VP and VD that give rise to essentially the same (naïve) validity-involving version of Curry's Paradox.

Then, using $\mathcal{D}_5$, we can then 'prove' $\bot$:

$$\dfrac{\dfrac{\mathcal{D}_5}{\vdash \mathsf{Val}(\ulcorner \pi \urcorner, \ulcorner \bot \urcorner)} \qquad \dfrac{\dfrac{\mathcal{D}_5}{\vdash \mathsf{Val}(\ulcorner \pi \urcorner, \ulcorner \bot \urcorner)}}{\vdash \pi}\,\text{Definition of } \pi.}{\vdash \bot}\,\mathsf{VDm}$$

Since the argument makes no assumptions about the logic of negation and the conditional, it resists *fully structural* revisionary treatments, i.e. treatments that retain all of SRef, SContr, and Cut. As [2] point out, on a revisionary approach to the semantic paradoxes, the v-Curry Paradox calls for a *substructural* treatment – one on which one between SRef, SContr, and Cut is restricted.[7]

2 Fixed Points for Naïve Validity

Substructural approaches to the semantic paradoxes mostly fall into two main camps: SContr-free (see e.g. [16, 34–38]) and Cut-free (see e.g. [4, 5, 26–28]).[8] However, both kinds of approaches have been criticised as being both radical and unnecessary. For, a classically minded theorist might reason, how could a difference in the number of times a given premise is used in the course of an argument affect that argument's validity? And how could consequence fail to be transitive? Moreover, one might legitimately ask, is there really a coherent notion of naïve validity for which *Validity Proof* and *Validity Detachment* are sound and yet non-trivial? In particular, ([7] pp. 10–11) argues that substructural approaches to paradox hard to get one's head around and alleges [8] that there can be no coherent notion of validity satisfying both *Validity Proof* Proof and *Validity Detachment*.

However, Field's assessment is too bleak. Nicolai and Rossi [23] offer a SRef-free theory of naïve truth and naïve validity (called *grounded validity*), based on a family of fixed point constructions that can be seen as a generalisation of Kripke's original construction for truth. As we're about to see, the theory validates versions of *Validity Proof* and *Validity Detachment*. Since the new construction is but a natural generalisation of Kripke's original one, [22] argue, in response to Field, that grounded validity is just as legitimate as a naïve conception of truth based on Kripke's

[7] Strictly speaking, Cut is not employed in the above derivation. This is because the formulation of VDm essentially already incorporates a form of Cut (or meta-inferential *modus ponens*), and is therefore unavailable to non-transitive theories. More generally, the principle of *Validity Detachment* can be (and has been) formulated in two different ways in the literature: as an inference,

$$\dfrac{\varphi,\ \mathsf{Val}(\ulcorner \varphi \urcorner, \ulcorner \psi \urcorner \vdash \psi}{}\,\mathsf{VD}$$

and as a meta-inference, as in our VDm (the 'm' stands for 'meta-inference'). A derivation of $\bot$ via a v-Curry reasoning involving both Cut and VD can be easily obtained via a simple modification of the above derivation (see [22], §1.3). We here focus on the meta-inferential VDm, because our target are non-reflexive theories such as the one developed in Section 2, which can be consistently closed under VDm but not VD.

[8] For general background on revisionary approaches to semantic paradox, see [3]. For a comprehensive critical comparison of SContr-free and Cut-free, see [29].

original construction. In what follows, we offer an informal presentation of Nicolai and Rossi's construction – the KV-construction.[9]

2.1 The KV-Construction

Let $\mathcal{L}_{\mathsf{V}}$ be $\mathcal{L}$ plus the designated predicate $\mathsf{Val}(x, y)$, for 'is naïvely valid'. We don't need to add a predicate for truth, since naïve truth can be defined putting $\mathsf{Tr}(\ulcorner\varphi\urcorner) := \mathsf{Val}(\ulcorner\top\urcorner, \ulcorner\varphi\urcorner)$. Similarly, negation can be defined via the validity predicate, letting $\neg\varphi := \mathsf{Val}(\ulcorner\varphi\urcorner, \ulcorner\bot\urcorner)$, and conditionals can be defined via negation and disjunction, setting $\varphi \rightarrow \psi := \neg\varphi \vee \psi$. For this reason, we freely employ the truth predicate, negations, and conditionals in what follows, even though the semantic construction to be developed in this section only contains explicit clauses for the naïve validity predicate.[10]

Let $\mathcal{M}$ be a countable acceptable model of the non-semantic fragment of $\mathcal{L}_{\mathsf{V}}$.[11] We also assume that $\mathcal{L}_{\mathsf{V}}$ has constants for all the elements in $\mathcal{M}$. The KV-construction defines an increasing succession of sets

$$\mathsf{I}^0_\Psi \subsetneqq \mathsf{I}^1_\Psi \subsetneqq \ldots \subsetneqq \mathsf{I}^{\alpha+1}_\Psi \subsetneqq \ldots$$

which ultimately determines the extension of the predicate Val. However, rather than building a succession of sets of sentences, the KV-construction builds a succession of sets of multiple conclusion *sequents*, i.e. objects of the form $\Gamma \vdash \Delta$, which are then internalized via naïve clauses for the predicate Val, thus building its extension.[12] The succession of sets of inferences is formalized by an inductive definition, and therefore it reaches a fixed point, i.e. for some ordinal δ:

$$\mathsf{I}^0_\Psi \subsetneqq \mathsf{I}^1_\Psi \subsetneqq \ldots \subsetneqq \mathsf{I}^{\alpha+1}_\Psi \subsetneqq \ldots \subsetneqq \mathsf{I}^{\delta}_\Psi = \mathsf{I}^{\delta+1}_\Psi = \ldots$$

The starting point of the construction is just the empty set: $\mathsf{I}^0_\Psi = \varnothing$. At stage 1, it contains only the inferences that are validated by the base model $\mathcal{M}$. Therefore, I^1_Ψ contains:

$$\begin{array}{ll} \Gamma \vdash \varphi, \Delta & \text{where } \varphi \text{ is an atomic } \mathcal{L}\text{-sentence and } \mathcal{M} \models \varphi, \\ \Gamma, \psi \vdash \Delta & \text{where } \psi \text{ is an atomic } \mathcal{L}\text{-sentence and } \mathcal{M} \not\models \psi \end{array}$$

[9] Our discussion here presupposes familiarity with Kripke's theory (strong Kleene version). For a technically informed presentation, see ([17] Chapters 3 and 4); for a less technical presentation, see ([32] Chapters 4 and 5).

[10] As it turns out, the so-defined material conditional has the same truth-conditions of the corresponding Val-statement in the KV-construction. In other words, $\varphi \rightarrow \psi$ is validated by the KV-construction whenever $\mathsf{Val}(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ is. This means that, on the semantics on offer, the material conditional can simply be taken to be the 'connective version' of the naïve validity predicate. To be sure, *qua* predicate, the validity predicate can be used by itself to formulate paradoxical sentences (unlike the material conditional).

[11] An acceptable model has an in-built coding scheme and a sub-structure which is isomorphic to $\mathbb{N}$ (the standard model of arithmetic) and $\leq_{\mathbb{N}}$ (its standard well-ordering). See ([20] p. 22) for more details.

[12] Using multiple conclusion sequents simplifies some of the proofs of the properties of the KV-construction.

At stage $\alpha + 1$, the construction only contains the sequents that are obtained from sequents in stage α, by closing it under all the classically valid meta-inferences. More precisely, $\Gamma \vdash \Delta \in \mathsf{I}_\Psi^{\alpha+1}$ if:

$\Gamma \vdash \Delta$ is in I_Ψ^α, or

$\Gamma \vdash \Delta$ is $\Gamma \vdash \varphi \wedge \psi, \Delta_0$ and $\Gamma \vdash \varphi, \Delta_0$ is in I_Ψ^α and $\Gamma \vdash \psi, \Delta_0$ is in I_Ψ^α, or

$\Gamma \vdash \Delta$ is $\Gamma_0, \varphi \wedge \psi \vdash \Delta$ and $\Gamma_0, \varphi, \psi \vdash \Delta$ is in I_Ψ^α, or

$\Gamma \vdash \Delta$ is $\Gamma \vdash \varphi \vee \psi, \Delta_0$ and $\Gamma \vdash \varphi, \psi, \Delta_0$ is in I_Ψ^α, or

$\Gamma \vdash \Delta$ is $\Gamma_0, \varphi \vee \psi \vdash \Delta$ and $\Gamma_0, \varphi \vdash \Delta$ is in I_Ψ^α and $\Gamma_0, \psi \vdash \Delta$ is in I_Ψ^α, or

$\Gamma \vdash \Delta$ is $\Gamma \vdash \forall x \varphi(x), \Delta_0$ and for every closed $\mathcal{L}_\mathsf{V}$-term s: $\Gamma \vdash \varphi(s), \Delta_0$ is in I_Ψ^α, or

$\Gamma \vdash \Delta$ is $\Gamma_0, \forall x \varphi(x) \vdash \Delta$ and for some closed $\mathcal{L}_\mathsf{V}$-term s: $\Gamma_0, \varphi(s) \vdash \Delta$ is in I_Ψ^α, or

$\Gamma \vdash \Delta$ is $\Gamma \vdash \mathsf{Val}(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner), \Delta_0$ and $\Gamma, \varphi \vdash \psi, \Delta_0$ is in I_Ψ^α, or

$\Gamma \vdash \Delta$ is $\Gamma_0, \mathsf{Val}(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \vdash \Delta$ and $\Gamma_0 \vdash \varphi, \Delta$ is in I_Ψ^α and $\Gamma_0, \psi \vdash \Delta$ is in I_Ψ^α

Finally, at limit stages we take unions. For β a limit ordinal:

$$\mathsf{I}_\Psi^\beta = \bigcup_{\alpha<\beta} \mathsf{I}_\Psi^\alpha$$

We indicate with I_Ψ the (smallest) fixed point of the above construction.

Just like Kripke's original construction, I_Ψ validates the intersubstitutivity of φ and $\mathsf{Tr}(\ulcorner\varphi\urcorner)$ in all non-opaque contexts. More precisely, recalling that $\mathsf{Tr}(\ulcorner\varphi\urcorner)$ is defined as $\mathsf{Val}(\ulcorner\top\urcorner, \ulcorner\varphi\urcorner)$, the construction clearly validates the following equivalences:[13]

$$\Gamma \vdash \mathsf{Tr}(\ulcorner\varphi\urcorner), \Delta \in \mathsf{I}_\Psi \text{ iff } \Gamma \vdash \mathsf{Val}(\ulcorner\top\urcorner, \ulcorner\varphi\urcorner), \Delta \in \mathsf{I}_\Psi \text{ iff } \Gamma \vdash \varphi, \Delta \in \mathsf{I}_\Psi$$

$$\Gamma, \mathsf{Tr}(\ulcorner\varphi\urcorner) \vdash \Delta \in \mathsf{I}_\Psi \text{ iff } \Gamma, \mathsf{Val}(\ulcorner\top\urcorner, \ulcorner\varphi\urcorner) \vdash \Delta \in \mathsf{I}_\Psi \text{ iff } \Gamma, \varphi \vdash \Delta \in \mathsf{I}_\Psi$$

In addition, it also validates the following versions of *Validity Proof* and *Validity Detachment*:

(VP) if $\Gamma, \varphi \vdash \psi$ is in I_Ψ, then $\Gamma \vdash \mathsf{Val}(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ is in I_Ψ.

(VDm) if $\Gamma \vdash \mathsf{Val}(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ is in I_Ψ and $\Delta \vdash \varphi$ is in I_Ψ, then $\Gamma, \Delta \vdash \psi$ is in I_Ψ.

2.2 A Non-reflexive Theory of Naïve Validity

The set of sequents and the extension of Val provided by I_Ψ yields a *non-reflextive* theory of naïve validity (and truth). Structural reflexivity is the only classically valid structural rule that in general fails in I_Ψ, since $\pi \vdash \pi$ (or equivalently $\varnothing \vdash \mathsf{Val}(\ulcorner\pi\urcorner, \ulcorner\pi\urcorner)$) is not in I_Ψ. More generally, I_Ψ does not validate any inference: it is *not* the case that, if a schematic inference $\Gamma \vdash \Delta$ is classically valid, then all its instances are in I_Ψ. However, I_Ψ is closed under all the classically valid *meta-inferences*: if the meta-inference

$$\frac{\Gamma_0 \vdash \Delta_0, \ \ldots, \ \Gamma_n \vdash \Delta_n}{\Gamma \vdash \Delta}$$

[13] Proving this requires establishing some closure properties of I_Ψ. See [23] for details.

is classically valid, then whenever $\Gamma_0 \vdash \Delta_0, \ldots, \Gamma_n \vdash \Delta_n$ are in I_Ψ, so is $\Gamma \vdash \Delta$. Since structural reflexivity is the only inference among the classically valid structural rules (i.e. the other structural rules are all meta-inferences), it is the only structural rule that's not unrestrictedly validated.

The theory of naïve validity provided by I_Ψ implicitly adopts a *tolerant-strict* (or TS) notion of consequence [4, 5]. Simplifying a bit, a TS-consequence relation interprets the logical constants as in a strong Kleene evaluation, and therefore makes use of three semantic values (0, 1/2, and 1). A set of sentences Δ is said to TS-*follow* from a set of sentences Γ if, for any given strong Kleene evaluation, whenever all the sentences in Γ have value 1 or 1/2, then at least one of the sentences in Δ has value 1. Once validity is so defined, the failure of structural reflexivity is easy to see: since π receives value 1/2 in the strong Kleene evaluation associated with I_Ψ, the sequent $\pi \vdash \pi$ is not TS-valid.

I_Ψ guarantees the consistency and non-triviality of several non-reflexive axiomatic theories of naïve truth and validity. Here, we (informally) present one very simple such theory:[14]

Definition 1 (TSTV) TSTV is provided by:

- Primitive name-forming;
- CPL (defined as above), *minus* SRef, *plus* the following principle:

$$\frac{\Gamma \vdash \varphi \rightarrow \varphi}{\Gamma, \varphi \vdash \varphi}\ \mathsf{WSRef}$$

- The naïve truth rules Tr-I, Tr-E, ¬Tr-I, ¬Tr-E, and the following two rules for naïve validity:

$$\frac{\Gamma, \varphi \vdash \psi}{\Gamma \vdash \mathsf{Val}(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)}\ \mathsf{VP} \qquad \frac{\Gamma \vdash \mathsf{Val}(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \qquad \Delta \vdash \varphi}{\Gamma, \Delta \vdash \psi}\ \mathsf{VDm}$$

TSTV is an extremely weak axiomatic theory: to wit, it does not have compositional rules for truth or validity, nor does it have strong base-theoretic axioms, such as the instances of the induction schema. For all its weakness, however, TSTV is prone to revenge – and so are, *a fortiori*, all of its stronger extensions.

2.3 Towards Revenge

As far as we can see, *pace* [8], I_Ψ and TSTV characterise a coherent notion of validity – one that satisfies forms of *Validity Proof* and *Validity Detachment* and that also justifies, for this reason, the adoption of a substructural logic of paradox.[15] To be sure, restrictions of SRef might seem pretty drastic. After all, SRef simply codifies, in a sequent setting, the natural deduction *rule of assumption*, which says that we may assume,

[14] A much richer, axiomatic non-reflexive theory of naïve truth and validity is developed in Nicolai and Rossi, [23]. Systems for non-reflexive consequence (under review).

[15] We defend this claim at length in [22]. See also footnote 7, where we explicitly consider different formalisations of *Validity Detachment*.

for the sake of argument, any sentence we like. And, it might be objected, if we can't make assumptions for the sake of argument for any sentence, how are we to reason? In response to this (understandable) worry, we'd like to make two observations.

To begin with, consider non-classical theories of naïve truth whose underlying logic is the three-valued strong Kleene logic K3, or some extension thereof [7, 13]. In these theories, validity is defined as preservation of value 1, and meta-inferences such as $\rightarrow$-I and $\neg$-I fail. Although in K3-based theories all instances of SRef hold, the assumptions that are thereby licensed cannot really be used for $\rightarrow$-I and $\neg$-I, since these rules do not hold in K3-based theories. Thus, it seems to us, our capacity to make assumptions is *already compromised* in well-known and widely accepted non-classical theories, such as the axiomatic theory PKF [11, 13, 14].

The second point is that, although SRef isn't valid in general, many of its instances are perfectly harmless. More precisely, in the foregoing setting, SRef is correctly applied to any sentence with a classical value. To be sure, we don't always know if a sentence actually has a classical value. For instance, natural languages allow the construction of *contingent liars* – sentences such as 'What Jane said is not true', whose paradoxical status depends on contingent facts, such as whether 'What Jane said is not true' *is* the semantic value of 'what Jane said' (see [15] p. 692 and ff.). If it is, then, in the foregoing framework, 'What Jane said is not true' has value 1/2, and one may not apply SRef to it. However, one might reason, what's the worst that could happen if we applied SRef to a paradoxical instance of 'What Jane said is not true'? Arguably, just this: *on such an assumption*, we'd derive (via familiar Liar-like reasoning) that what Jane said is both true and not true, i.e. (courtesy of $\neg$-E) we'd be in a position to derive $\bot$. But rather than trivialising our theory, this would simply give us a reason to discharge the assumption that 'What Jane said is not true' satisfies SRef and assert that 'What Jane said is not true' is paradoxical, and shouldn't be reasoned with classically. That is, a principle to the effect that if $\bot$ can be derived from the assumption that φ satisfies $\varphi \vdash \varphi$ (or some equivalent principle), then φ is paradoxical, would help us deciding *which sentences* are paradoxical and hence cannot be reasoned with classically. For instance, if the assumption that λ satisfies $\lambda \vdash \lambda$ (or some equivalent principle) leads to absurdity, then we know that λ is paradoxical and hence may not be reasoned with classically (we return to this point in Section 4).

However, this is precisely where TSTV and its extension reveal their limits. Murzi and Rossi [21] show that four families of revisionary theories (including theories that restrict SContr and theories that restrict Cut) are committed to validating principles about paradoxicality and unparadoxicality under which these theories can only be closed on pain of triviality. In a nutshell, the basic idea is this. Non-classical theories restrict the application of classical principles to unparadoxical sentences. However, for each such non-classical theory, there exist finitely many principles $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ such that a sentence φ satisfies $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ if and only if it satisfies all the principles of classical logic. This allows one to characterise *paradoxical* sentences as the ones that only satisfy each of $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ on pain of triviality and *unparadoxical* sentences as the ones that satisfy each of $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$. Such a characterisation cannot be explicitly articulated in the theories in question, though, on pain of paradox-induced triviality. In the next section, we apply the recipe to TSTV and its extensions.

Remark We should stress that our distinction between paradoxical and unparadoxical sentences (in a given theory T) is not assumed to be exhaustive or exclusive. As Lucas Rosenblatt [30] points out, paracomplete Kripkean theories of truth, for instance, can feature sentences – such as the truth-teller τ, identical to its own truth predication $\mathsf{Tr}(\ulcorner\tau\urcorner)$ – which do *not* entail $\bot$ if reasoned with classically, and which yet fail to satisfy all the classically valid principles. In our framework, then, such sentences are neither paradoxical nor unparadoxical. Rosenblatt [30] uses arguments in the style of [18] to show that *every* classically valid principle $\mathfrak{P}$ has sets of instances which are mutually inconsistent – essentially, because any consistent set of instances of $\mathfrak{P}$ can be expanded in incompatible ways, e.g. by adding to it sentences such as τ, or their negations. These arguments, however, shouldn't be seen as showing that our notions of paradoxicality and unparadoxicality for a theory T are unacceptable. Let $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ be T's classical recapturing principles and suppose τ does not satisfy, in T, all the principles of classical logic. Now let's further assume that T is closed under principles to the effect that φ is paradoxical if and only if it satisfies $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$ only on pain of triviality and *unparadoxical* if and only if it satisfies each of $\mathfrak{P}_1, \ldots, \mathfrak{P}_n$. Then, T doesn't prove that τ is unparadoxical and, since τ doesn't entail $\bot$ in T, it also doesn't prove that it is paradoxical. That is, T is *silent* about τ. But, by our lights, this need not be a problem: we shouldn't expect sentences such as τ and 'Salzburg is East of Munich' to be assigned the same semantic status, if one behaves non-classically and the other doesn't.[16]

3 Revenge

Non-classical theories such as TSTV all share a common feature: despite their non-classicality, they have *fully classical* fragments. That is, they limit their restrictions to classical logic to *some* sentences. This is not only a basic fact about those theories; it also allows one to apply those theories to mathematics and science more generally. As it is sometimes said, non-classical theories can *recapture* classical reasoning when needed.[17]

To see how non-classical theories recapture classical theories, our starting point is a particularly simple way of characterising the classical fragment of non-classical theories such as TSTV and their extensions. Such theories all enjoy the following informal property:

(Classicality Principles) There are finitely many classically valid principles such that a sentence satisfies such principles only if it satisfies all classical principles.

We can then say that a theory recaptures classical logic if it is closed under weaker versions of classical rules which, whenever some extra conditions are satisfied,

[16] We thank an anonymous referee for helping us appreciate this point.
[17] See e.g. ([24] p. 221), [7], ([1] pp. 111-2), [34].

reduce to their classical counterparts. In the case of TSTV, the weaker rule is WSRef and the extra condition is given by the following axiom:

$$(\mathsf{Id})\ \vdash \varphi \rightarrow \varphi$$

The following definition formally captures these ideas.

Definition 2 (Classical recapture) Let S be a non-trivial theory. Then, S enjoys a classical recapturing property if it is $\mathfrak{P}$-classical recapturing for some principle $\mathfrak{P}$ invalid in S. In particular, S is Id-*classical recapturing* if it is closed under the rules of CPL, where SRef is replaced by WSRef.

It's immediate to see why TSTV is Id-classical recapturing: if φ satisfies Id, then, courtesy of WSRef, φ also satisfies $\varphi \vdash \varphi$, whence by definition of CPL it also satisfies all the principles of classical logic.

Now, since $\vdash \varphi \rightarrow \varphi$ is classical recapturing in TSTV, a sentence φ is paradoxical if the assumption that it satisfies $\vdash \varphi \rightarrow \varphi$ leads to absurdity. Conversely, the claims that φ is paradoxical and that it satisfies $\vdash \varphi \rightarrow \varphi$ must be jointly inconsistent. We can articulate these intuitive thoughts by means of the following higher-order rules [31]:[18]

$$\Box\ \frac{}{\Gamma \vdash \varphi \rightarrow \varphi}\ n$$

$$\vdots$$

$$\frac{\Gamma \vdash \bot}{\Gamma \vdash \mathsf{Par}(\ulcorner \varphi \urcorner)}\ \mathsf{Id\text{-}Par\text{-}I},\ n \qquad \frac{\Gamma \vdash \mathsf{Par}(\ulcorner \varphi \urcorner) \qquad \Delta \vdash \varphi \rightarrow \varphi}{\Gamma, \Delta \vdash \bot}\ \mathsf{Id\text{-}Par\text{-}E}$$

We employ a 'higher-order' rule *à la* Schroeder-Heister [see 31], since some non-reflexive logics that have actually been employed in the literature have no inferences, but only meta-inferences. For this reason, we need to be able to assume, and discharge, *sequents* in order to determine whether a sentence is paradoxical in our sense. Let's now explicitly include these rules in our theory of naïve validity and truth.

Definition 3 (TSTVP) TSTVP is the result of formulating TSTV in a language featuring a primitive unary predicate Par (call it $\mathcal{L}_{V,P}$), and closing it under Id-Par-I and Id-Par-E.

We can now show that TSTVP is trivial.[19]

[18] Following [33], the box here indicates that the assumption $\Gamma \vdash \varphi \rightarrow \varphi$ may not be discharged vacuously.

[19] As an anonymous referee points out, the following result could also be proven via a context-free version of our rule Id-Par-I, thus still weakening our assumptions about revenge-prone theories.

Proposition 4 TSTVP *is trivial, and so is the closure under* Id-Par-I *and* Id-Par-E *of any theory extending* TSTV.

To prove this, we first prove the following Lemma.

Lemma 5 *Let* ρ *be a sentence identical to* $\mathsf{Tr}(\ulcorner\rho\urcorner) \rightarrow \mathsf{Par}(\ulcorner\rho\urcorner)$. *Then, for any theory as strong as* TSTVP, *the following holds: if* $\vdash \rho \rightarrow \rho$, *then* $\vdash \mathsf{Tr}(\ulcorner\rho\urcorner) \rightarrow \mathsf{Tr}(\ulcorner\rho\urcorner)$.

Proof Let ρ be a sentence identical to $\mathsf{Tr}(\ulcorner\rho\urcorner) \rightarrow \mathsf{Par}(\ulcorner\rho\urcorner)$. We now reason in TSTVP, as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\vdash \rho \rightarrow \rho}{\rho \vdash \rho}\,\textsf{WSRef}}{\rho \vdash \mathsf{Tr}(\ulcorner\rho\urcorner)}\,\textsf{Tr-I} \quad \dfrac{\dfrac{\vdash \rho \rightarrow \rho}{\rho \vdash \rho}\,\textsf{WSRef}}{\rho \vdash \mathsf{Tr}(\ulcorner\rho\urcorner) \rightarrow \mathsf{Par}(\ulcorner\rho\urcorner)}\,\textsf{Def. of } \rho}{\rho, \rho \vdash \mathsf{Par}(\ulcorner\rho\urcorner)}\,\rightarrow\textsf{-E}}{\rho \vdash \mathsf{Par}(\ulcorner\rho\urcorner)}\,\textsf{SContr} \quad \vdash \rho \rightarrow \rho}{\vdash \mathsf{Tr}(\ulcorner\rho\urcorner) \rightarrow \mathsf{Tr}(\ulcorner\rho\urcorner)}\,\textsf{Id-Par-E}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\rho \vdash \bot}{\vdash \neg\rho}\,\neg\textsf{-I}}{\vdash \neg\mathsf{Tr}(\ulcorner\rho\urcorner)}\,\neg\textsf{Tr-I}}{\neg\mathsf{Tr}(\ulcorner\rho\urcorner) \vdash \neg\mathsf{Tr}(\ulcorner\rho\urcorner)}\,\textsf{SWeak}}{\vdash \neg\mathsf{Tr}(\ulcorner\rho\urcorner) \rightarrow \neg\mathsf{Tr}(\ulcorner\rho\urcorner)}\,\rightarrow\textsf{-I}}{\vdash \mathsf{Tr}(\ulcorner\rho\urcorner) \rightarrow \mathsf{Tr}(\ulcorner\rho\urcorner)}\,\textsf{Contr}$$

□

Making use of the Lemma, can now prove Proposition 5.

Proof Again, we reason in TSTVP. We begin by deriving $\mathsf{Tr}(\ulcorner\rho\urcorner) \rightarrow \mathsf{Par}(\ulcorner\rho\urcorner)$ from $\vdash \rho \rightarrow \rho$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\vdash \rho \rightarrow \rho}{\vdash \mathsf{Tr}(\ulcorner\rho\urcorner) \rightarrow \mathsf{Tr}(\ulcorner\rho\urcorner)}\,\textsf{Lemma 2}}{\mathsf{Tr}(\ulcorner\rho\urcorner) \vdash \mathsf{Tr}(\ulcorner\rho\urcorner)}\,\textsf{WSRef}}{\mathsf{Tr}(\ulcorner\rho\urcorner) \vdash \rho}\,\textsf{Tr-E}}{\mathsf{Tr}(\ulcorner\rho\urcorner) \vdash \mathsf{Tr}(\ulcorner\rho\urcorner) \rightarrow \mathsf{Par}(\ulcorner\rho\urcorner)}\,\textsf{Def. of } \rho \quad \dfrac{\dfrac{\vdash \rho \rightarrow \rho}{\vdash \mathsf{Tr}(\ulcorner\rho\urcorner) \rightarrow \mathsf{Tr}(\ulcorner\rho\urcorner)}\,\textsf{Lemma 2}}{\mathsf{Tr}(\ulcorner\rho\urcorner) \vdash \mathsf{Tr}(\ulcorner\rho\urcorner)}\,\textsf{WSRef}}{\mathsf{Tr}(\ulcorner\rho\urcorner), \mathsf{Tr}(\ulcorner\rho\urcorner) \vdash \mathsf{Par}(\ulcorner\rho\urcorner)}\,\rightarrow\textsf{-E}}{\mathsf{Tr}(\ulcorner\rho\urcorner) \vdash \mathsf{Par}(\ulcorner\rho\urcorner)}\,\textsf{SContr}}{\vdash \mathsf{Tr}(\ulcorner\rho\urcorner) \rightarrow \mathsf{Par}(\ulcorner\rho\urcorner)}\,\rightarrow\textsf{-I}$$

Call this derivation from the open assumption $\vdash \rho \rightarrow \rho$ (taken twice), $\mathcal{D}_6$. We now use two copies of $\mathcal{D}_6$ to prove that ρ is paradoxical:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\vdash \rho \to \rho}\,1}{\mathcal{D}_6}}{\vdash \mathsf{Tr}(\ulcorner \rho \urcorner) \to \mathsf{Par}(\ulcorner \rho \urcorner)}\,\text{Def. of } \rho}{\dfrac{\vdash \rho}{\vdash \mathsf{Tr}(\ulcorner \rho \urcorner)}\,\mathsf{Tr\text{-}I}} \qquad \dfrac{\dfrac{\dfrac{}{\vdash \rho \to \rho}\,1}{\mathcal{D}_6}}{\vdash \mathsf{Tr}(\ulcorner \rho \urcorner) \to \mathsf{Par}(\ulcorner \rho \urcorner)}}{\vdash \mathsf{Par}(\ulcorner \rho \urcorner)}\,{\to}\text{-E} \qquad \dfrac{}{\vdash \rho \to \rho}\,1}{\dfrac{\vdash \bot}{\vdash \mathsf{Par}(\ulcorner \rho \urcorner)}\,\text{Id-Par-I, 1}}\,\text{Id-Par-E}$$

Call this second derivation $\mathcal{D}_7$. We now use two copies of $\mathcal{D}_7$ to derive $\bot$ from the empty set of assumptions:

$$\dfrac{\dfrac{\mathcal{D}_7}{\vdash \mathsf{Par}(\ulcorner \rho \urcorner)} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\mathcal{D}_7}{\vdash \mathsf{Par}(\ulcorner \rho \urcorner)}\,\mathsf{SWeak}}{\mathsf{Tr}(\ulcorner \rho \urcorner) \vdash \mathsf{Par}(\ulcorner \rho \urcorner))}\,{\to}\text{-I}}{\vdash \mathsf{Tr}(\ulcorner \rho \urcorner) \to \mathsf{Par}(\ulcorner \rho \urcorner)}\,\text{Def. of } \rho}{\vdash \rho}\,\mathsf{SWeak}}{\rho \vdash \rho}\,{\to}\text{-I}}{\vdash \rho \to \rho}\,\text{Id-Par-E}}{\vdash \bot}$$

□

4 Discussion

It may be objected that $\mathcal{D}_7$ is unacceptable for someone accepting TSTV or some of its extensions. After all, one might argue, one cannot in general start a derivation from the assumption $\vdash \rho \to \rho$ in any theory whose underlying logic is TS, for the simple reason that this would be tantamount to relying on the validity of SRef, which is however TS-invalid. To be sure, if we know that a sentence is unparadoxical, then we can use the corresponding instance of SRef. But that's not the case with ρ: here *we don't yet know* ρ's semantic status; whence we cannot reason classically about it – in particular, disallowing SRef means that we can't start a derivation from $\rho \vdash \rho$, or, from that matters, from its TS∪WSRef-equivalent, $\vdash \rho \to \rho$. Or so one might argue.

Let's take this thought seriously for a moment. As we observed in Section 2, whether a sentence is paradoxical may depend on features of the context rather than on the sentence itself. This means, though, that in general we don't know if a sentence is unparadoxical. We might perhaps know that the sentences of the language of arithmetic, or set theory, are not paradoxical. But we certainly lack such a knowledge for sentences of languages that are rich enough to express contingent Liars – including potentially harmless sentences such as 'What Jane said is not true'. Thus, if we block derivations like $\mathcal{D}_7$ on the grounds that they may involve contingent Liars, then we simply cannot in general assume, and hence reason about, sentences such as 'What Jane said is not true'. However, disallowing the assumptions of such

potentially harmless sentences would not only cripple – as Field would put it – ordinary reasoning involving the truth predicate. It would also make little sense from a revisionary perspective.

Even if the logic is non-classical, one should be able to assume that a given sentence satisfies all the principles of classical logic, and see what follows from this. For instance, let our background logic be $\mathsf{TS} \cup \mathsf{WSRef}$ and suppose we assume that the Liar sentence λ satisfies $\vdash \lambda \to \lambda$, and hence all the principles of classical logic. We can now proceed to derive $\vdash \bot$ via familiar Liar-like reasoning – call his derivation $\mathcal{D}_8$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\vdash \lambda \to \lambda}{\lambda \vdash \lambda}\,\mathsf{WSRef}}{\lambda \vdash \mathsf{Tr}(\ulcorner \lambda \urcorner)}\,\mathsf{Tr\text{-}I} \qquad \dfrac{\dfrac{\vdash \lambda \to \lambda}{\lambda \vdash \lambda}\,\mathsf{WSRef}}{\lambda \vdash \neg\mathsf{Tr}(\ulcorner \lambda \urcorner)}\,\mathsf{Def.\ of}\ \lambda}{\neg\text{-}\mathsf{E}}}{\dfrac{\lambda, \lambda \vdash \bot}{\lambda \vdash \bot}\,\mathsf{SContr}}\,\neg\text{-}\mathsf{I}}{\vdash \neg\lambda}}{\vdash \neg\mathsf{Tr}(\ulcorner \lambda \urcorner)}\,\neg\mathsf{Tr\text{-}I} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\vdash \lambda \to \lambda}{\lambda \vdash \lambda}\,\mathsf{WSRef}}{\lambda \vdash \mathsf{Tr}(\ulcorner \lambda \urcorner)}\,\mathsf{Tr\text{-}I} \qquad \dfrac{\dfrac{\vdash \lambda \to \lambda}{\lambda \vdash \lambda}\,\mathsf{WSRef}}{\lambda \vdash \neg\mathsf{Tr}(\ulcorner \lambda \urcorner)}\,\mathsf{Def.\ of}\ \lambda}{\neg\text{-}\mathsf{E}}}{\dfrac{\lambda, \lambda \vdash \bot}{\lambda \vdash \bot}\,\mathsf{SContr}}\,\neg\text{-}\mathsf{I}}{\vdash \neg\lambda}}{\vdash \neg\mathsf{Tr}(\ulcorner \lambda \urcorner)}\,\neg\mathsf{Tr\text{-}I}}{\vdash \lambda}\,\mathsf{Def.\ of}\ \lambda}{\vdash \mathsf{Tr}(\ulcorner \lambda \urcorner)}\,\mathsf{Tr\text{-}I}}{\vdash \bot}\,\neg\text{-}\mathsf{E}$$

Of course, the natural conclusion of such a reasoning is that λ may not satisfy all the principles of classical logic after all. That is, λ is paradoxical. We can capture this intuitive, and sound, diagnosis by means of our rule Id-Par-I:

$$\dfrac{\underbrace{\dfrac{}{\vdash \lambda \to \lambda}\,1 \qquad \dfrac{}{\vdash \lambda \to \lambda}\,1 \qquad \dfrac{}{\vdash \lambda \to \lambda}\,1 \qquad \dfrac{}{\vdash \lambda \to \lambda}\,1}_{\mathcal{D}_8}}{\vdash \bot}}{\vdash \mathsf{Par}(\ulcorner \lambda \urcorner)}\,\mathsf{Id\text{-}Par\text{-}I},\ 1$$

Here, assuming $\vdash \lambda \to \lambda$ has caused us no harm. On the contrary, it has allowed us, courtesy of Id-Par-I, to establish that λ is paradoxical, and hence that it cannot be reasoned with classically. We now *know* that we shouldn't trust derivations which rely on undischarged copies of the sequent $\vdash \lambda \to \lambda$. However, this doesn't mean that such a sequent can never be used. It can be *assumed* and later discharged, as in the above derivation, in a proof of $\vdash \mathsf{Par}(\ulcorner \lambda \urcorner)$.

This diagnosis can be extended to models of naïve truth and validity. For example, one can capture the extension of the intended notion of paradoxicality (according to which φ is paradoxical just in case it satisfies the classical recapturing principles, and hence full classical logic, only on pain of triviality) within I_Ψ as follows:

$$\text{Paradoxical in } \mathcal{L}_V := \{\varphi \in \mathcal{L}_V \mid \text{ if } \vdash \varphi \to \varphi \in \mathsf{I}_\Psi, \text{ then } \vdash \bot \in \mathsf{I}_\Psi\}$$

Intuitively paradoxical sentences, such as λ and π, are clearly paradoxical in $\mathcal{L}_V$ in the above sense. To be sure, the above notion of paradoxicality is unproblematic because it is limited to $\mathcal{L}_V$ (and does not apply to the full language $\mathcal{L}_{V,P}$). That is, it doesn't apply to sentences which themselves contain the paradoxicality predicate.

However, by our lights, this nonetheless establishes that the motivation behind the paradoxicality predicate is intuitively coherent.

But why shouldn't such a predicate be self-applicable, i.e. why shouldn't Par apply to sentences in which it itself figures? Just like typed notions of truth are usually disregarded as excessively restricted, especially by revisionary theorists, the same goes with paradoxicality. Consider again $\mathcal{D}_7$. Here, too, just like in $\mathcal{D}_8$, we assume $\vdash \rho \rightarrow \rho$, derive $\vdash \bot$, and conclude that ρ must be paradoxical after all, i.e. $\vdash \mathsf{Par}(\ulcorner \rho \urcorner)$. Thus, we contend, the uses of SRef in derivations such as $\mathcal{D}_7$ are motivated if the uses of SRef in $\mathcal{D}_8$ are. To be sure, our revenge paradox, i.e. our proof of Proposition 5, shows that some of these seemingly natural uses of Par actually lead to triviality. Thus, the paradox points to some unavoidable limits of non-reflexive approaches. Indeed, since similar paradoxes can be produced for virtually any non-classical theory satisfying CLASSICALITY PRINCIPLES, the paradoxes of paradoxicality and unparadoxicality show the limits of revisionary approaches in general.

It might again be objected that to assume the sequent $\vdash \rho \rightarrow \rho$ is effectively to make use of SRef in the metatheory. That is, where $\models$ is the consequence relation of our metatheory, it might be argued that to assume the sequent $\vdash \rho \rightarrow \rho$ is to assume the metatheoretic sequent

$$\vdash \rho \rightarrow \rho \models \vdash \rho \rightarrow \rho,$$

which is however just *another* instance of SRef. And, one might reason, if SRef fails in general, it must also fail in the metatheory. There's two problems with this rejoinder, however. To begin with, we haven't yet been given a reason to think that the metatheory M of a theory such as TSTV must be non-classical. Second, if it *is* non-classical, and one cannot in general make assumptions in M, there will be very little, if anything, any such metatheory will be able to prove.

Could either Id-Par-I or Id-Par-E be faulted instead? From a revisionary perspective, we see two problems with this. The first is that these principles are seemingly justified by reasonings such as $\mathcal{D}_8$, just like naïve principles for naïve truth and validity are seemingly justified by reasonings such as AGREEMENT and AGREEMENT*. The second problem problem is that there simply are, in theories whose logic is TS, sentences which satisfy $\varphi \rightarrow \varphi$, and hence all the principles of classical logic, and sentences which only satisfy $\varphi \rightarrow \varphi$ (and hence the principles of classical logic) on pain of triviality. Now, it seems to us, it's entirely legitimate to *assign labels* to these two kinds of sentences. For want of a better word, we've called sentences of the first kind unparadoxical and sentences of the second kind paradoxical. Given this terminological choice, for a sentence φ to be paradoxical in a theory such as TSTV *is* to entail $\bot$ if reasoned with classically. And since we know that φ satisfies $\varphi \rightarrow \varphi$ in TSTV if and only if it satisfies all the principles of classical logic, we also know that φ is paradoxical if and only if the assumption that it satisfies $\varphi \rightarrow \varphi$ entails $\bot$. As far as we can see, Id-Par-I or Id-Par-E simply articulate this fact about TSTV. Thus, a theory's failure to justify these principles should be interpreted as a serious expressive limitation, of the same kind as the ones revealed by the Liar Paradox, Curry's Paradox, and the semantic paradoxes more generally.

Acknowledgements We wish to thank an anonymous reviewer for helpful comments on a previous draft as well as the FWF (project number 29716-G24) for generous financial support.

Funding Open access funding provided by Paris Lodron University of Salzburg.

References

1. Beall, J. (2009). *Spandrels of truth*. Oxford: Oxford University Press.
2. Beall, J., & Murzi, J. (2013). Two flavors of Curry's Paradox. *The Journal of Philosophy*, *CX*(3), 143–65.
3. Beall, J., & Ripley, D. (2014). Non-classical theories of truth. In M. Glanzberg (Ed.) *The Oxford Handbook of Truth*. Oxford: Oxford University Press.
4. Cobreros, P., Egré, P., Ripley, D., & van Rooij, R. (2012). Tolerant, classical, strict. *Journal of Philosophical Logic*, *41*(2), 347–85.
5. Cobreros, P., Egré, P., Ripley, D., & van Rooij, R. (2013). Reaching transparent truth. *Mind*, *122*, 841–866.
6. Curry, H. (1942). The inconsistency of certain formal logics. *Journal of Symbolic Logic*, *7*, 115–7.
7. Field, H. (2008). *Saving truth from paradox*. Oxford: Oxford University Press.
8. Field, H. (2017). Disarming a paradox of validity. *Notre Dame Journal of Formal Logic*, *58*(1), 1–19.
9. Fjellstad, A. (2017). Non-classical elegance for sequent calculus enthusiasts. *Studia Logica*, *105*(1), 93–119.
10. French, R. (2016). Structural reflexivity and the paradoxes of self-reference. *Ergo, an Open Access Journal of Philosophy*, 3.
11. Halbach, V., & Horsten, L. (2006). Axiomatizing Kripke's theory of truth. *Journal of Symbolic Logic*, *71*, 677–712.
12. Heck, Jr. R. G. (2007). Self-reference and the languages of arithmetic. *Philosophia Mathematica*, *15*, 1–29.
13. Horsten, L. (2009). Levity. *Mind*, *118*(471), 555–581.
14. Horsten, L. (2012). *The Tarskian Turn. Deflationism and axiomatic truth*. Cambridge: MIT Press.
15. Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, *72*, 690–716.
16. Mares, E., & Paoli, F. (2014). Logical consequence and the paradoxes. *Journal of Philosophical Logic*, *43*, 439–469.
17. McGee, V. (1991). *Truth, vagueness and paradox*. Indianapolis: Hackett Publishing Company.
18. McGee, V. (1992). Maximal consistent sets of instances of Tarski's schema (T). *Journal of Philosophical Logic*, *21*(3), 235–241.
19. Meadows, T. (2014). Fixed points for consequence relations. *Logique et Analyse*, 333–357.
20. Moschovakis, Y. (1974). Elementary induction on abstract structures. North-Holland.
21. Murzi, J., & Rossi, L. (2021). Generalised revenge. *Australasian Journal of Philosophy*, *98*(1), 153–177.
22. Murzi, J., & Rossi, L. (2018). Naïve validity, Synthese. Online first: https://doi.org/10.1007/s11229-017-1541-6.
23. Nicolai, C., & Rossi, L. (2018). Principles for object-linguistic validity: from logical to irreflexive. *Journal of Philosophical Logic*, *47*, 549–577.
24. Priest, G. (2006). *Doubt Truth to be a Liar*. Oxford: Oxford University Press.
25. Restall, G. (2000). *An introduction to substructural logics*. New York: Routledge.

26. Ripley, D. (2012). Conservatively extending classical logic with transparent truth. *Review of Symbolic Logic*, 354–78.
27. Ripley, D. (2013). Paradoxes and failures of cut. *Australasian Journal of Philosophy*, *91*(1), 139–64.
28. Ripley, D. (2013). Revising up. *Philosophers' Imprint, 13*(5).
29. Ripley, D. (2015). Comparing substructural theories of truth. *Ergo*, *2*(13), 299–328.
30. Rosenblatt, L. (2020). Maximal non-trivial sets of instances of your least favorite logical principle. *The Journal of Philosophy*, *117*(1), 30–54.
31. Schroeder-Heister, P. (1984). A natural extension of natural deduction. *Journal of Symbolic Logic*, *49*, 1284–1299.
32. Soames, S. (1999). *Understanding truth*. Oxford: Oxford University Press.
33. Tennant, N. (2012). Cut for core logic. *Review of Symbolic Logic*, *5*(3), 450–479.
34. Zardini, E. (2011). Truth without contra(di)ction. *Review of Symbolic Logic*, *4*, 498–535.
35. Zardini, E. (2013). It is not the case that [p and 'it is not the case that p' is true] nor is it the case that [p and 'p' is not true]. *Thought*, *1*(4), 309–19.
36. Zardini, E. (2013). Näive logical properties and structural properties. *The Journal of Philosophy*, *110*(11), 633–44.
37. Zardini, E. (2014). Naïve truth and naïve logical properties. *Review of Symbolic Logic*, *7*(2), 351–384.
38. Zardini, E. (2015). Getting one for two, or the contractors' bad deal. Towards a unified solution to the semantic paradoxes. In T. Achourioti, K. Fujimoto, H. Galinon, & J. Martinez-Fernandez (Eds.) *Unifying the Philosophy of Truth* (pp. 461–93). Springer.