Naive Truth and Naive Logical Properties*

Elia Zardini

LOGOS, Logic, Language and Cognition Research Group

Department of Logic, History and Philosophy of Science

University of Barcelona

Northern Institute of Philosophy

Department of Philosophy

University of Aberdeen

elia.zardini@ub.edu

January 7, 2014

Abstract

A unified answer is offered to two distinct fundamental questions: whether a non-classical solution to the semantic paradoxes should be extended to other apparently similar paradoxes (in particular, to the paradoxes of logical properties) and whether a non-classical logic should be expressed in a non-classical metalanguage. The

*Earlier versions of the material in this paper have been presented in 2011 at the 7th Barcelona Workshop on Issues in the Theory of Reference on *Paradoxes of Truth and Denotation* (University of Barcelona); in 2012, at the 4th PETAF Workshop on *Modality* (University of Stockholm) and at the Jowett Society (University of Oxford); in 2013, at the LOGOS Semantic Paradoxes and Vagueness Seminar (University of Barcelona), at a NIP research seminar (University of Aberdeen) and at a research seminar at the University of the Azores. I would like to thank all these audiences for very stimulating comments and discussions. Special thanks go to Andrew Bacon, Franz Berto, Cian Dorr, Hartry Field, Joost Joosten, Dan López de Sa, José Martínez, Julien Murzi, Peter Pagin, Josh Parsons, Matteo Plebani, Graham Priest, Steve Read, Dave Ripley, Sven Rosenkranz, Rui Sampaio da Silva, Tim Williamson, Crispin Wright and two anonymous referees. At different stages during the writing of the paper, I have benefitted from a UNAM Postdoctoral Research Fellowship, from an AHRC Postdoctoral Research Fellowship and from the FP7 Marie Curie Intra-European Research Fellowship 301493 on *A Non-Contractive Theory of Naive Semantic Properties: Logical Developments and Metaphysical Foundations* (NTNSP), as well as from partial funds from the project CONSOLIDER-INGENIO 2010 CSD2009-00056 of the Spanish Ministry of Science and Innovation on *Philosophy of Perspectival Thoughts and Facts* (PERSP), from the FP7 Marie Curie Initial Training Network 238128 on *Perspectival Thoughts and Facts* (PETAF), from the project FFI2011-25626 of the Spanish Ministry of Science and Innovation on *Reference, Self-Reference and Empirical Data* and from the project FFI2012-35026 of the Spanish Ministry of Economy and Competition on *The Makings of Truth: Nature, Extent, and Applications of Truthmaking.*

paper starts by reviewing a budget of paradoxes involving the logical properties of validity, inconsistency and compatibility. The author's favoured substructural approach to naive truth is then presented and it is explained how that approach can be extended in a very natural way so as to solve a certain paradox of validity. However, three individually decisive reasons are later provided for thinking that no approach adopting a classical metalanguage can adequately account for all the features involved in the paradoxes of logical properties. Consequently, the paper undertakes the task to do better, and, building on the system already developed, introduces a theory in a non-classical metalanguage that expresses an adequate logic of naive truth and of some naive logical properties.

1 Introduction and Overview

This paper brings together two distinct fundamental questions which both arise in the context of solutions to the *semantic paradoxes* requiring a weakening of *classical* logic (henceforth, simply '*non-classical solutions*').[1] The first question is this: *how far does the non-classicality emerging in the semantic paradoxes extend to other areas of discourse*? The question arises in view of the fact that the so-called 'semantic paradoxes'—i.e. paradoxes involving the notions of *truth*, *satisfaction*, *reference* etc., as exemplified for instance by the *Liar paradox* (see Tarski [1933]), the *heterologicality paradox* (see Grelling and Nelson [1908]), the various *paradoxes of reference* (see Russell [1908]) etc.—actually do not appear essentially to hinge on the use of obviously *semantic* notions (i.e. notions that connect words with world). It is well-known that apparently similar paradoxes can be generated using the notion of *set* (see Russell [1903]), which is not obviously semantic. And it is equally well-known, at least among philosophers, that apparently similar paradoxes can be generated using the notion of *knowledge* (see Kaplan and Montague [1960]), or *necessity* (see Montague [1963]), or *belief* (see Burge [1978]) etc., which, again, are not obviously semantic. But it is much less well-known, save for historians of logic and for some discerning contemporary philosophers of logic, that the first period of the history of Western philosophy and logic showing a general appreciation of, interest in and a flurry of competing theories about the semantic paradoxes—i.e. the Middle Ages—often presented the semantic paradoxes as being of the same ilk as paradoxes involving obviously *logical* (rather than obviously semantic) notions, such as the notions of *validity*, *inconsistency*, *compatibility* etc.

Here is a modified example from the Pseudo-Scotus (*In Librum Primum Priorum Analyticorum Aristotelis Quaestiones*, q. 10). Consider the argument $\mathfrak{A}$: "God exists. Therefore, π and the negation of κ are compatible", where 'π' and 'κ' refer respectively to the premise and conclusion of $\mathfrak{A}$. Suppose that $\mathfrak{A}$ is not valid. Then surely 'π and the

[1] There are interesting logics of truth that validate all classical laws and rules but not all classical metarules (for example, the *supervaluationist* logic of McGee [1991], which validates all classical laws and rules but does not validate reasoning by cases and some other related classical metarules). For these logics, there is a largely terminological question concerning their "classicality". For the purposes of this paper, it'll be more convenient to include them in the non-classical solutions.

negation of κ are compatible'—i.e. κ—is logically true, and that plausibly implies that it also follows from[2] any other logical truth like 'God exists'.[3] Therefore, κ in effect follows from π, and hence $\mathfrak{A}$ is valid after all. By *reductio*, $\mathfrak{A}$ is valid (and κ is a logical truth, since it follows from a logical truth). But that surely implies that 'π and the negation of κ are compatible'—i.e. κ—is inconsistent, while we've just established that it is a logical truth!

Faced with such a wide variety of apparently similar paradoxes, non-classical solutions are pressed to take a stance as to how far their favoured non-classical logic should be deployed in order to provide a structurally identical solution. Although maybe natural, the stance requiring that all apparently similar paradoxes be solved by deploying the same non-classical logic (famously defended by Graham Priest, see in particular Priest [2003]) is far from obvious and far from being universally accepted (see for example Field [2008] for a much more conservative stance). I should stress that I don't mean to imply that an analogous question never arises for classical solutions. For example, a classical solution *stratifying* the notion of truth will be pressed to take a stance as to how many of the notions giving rise to apparently similar paradoxes should be stratified in order to provide a structurally identical solution. Still, the specific question pressed on non-classical solutions, having to do with nothing less than *deviation from classical logic* in certain areas of discourse, possesses its own distinctive interest and gives rise to its own distinctive issues.

The second question is this: *how much of a non-classical logic is expressed*[4] *by a classical metalanguage*?[5] Typically, non-classical solutions criticise one of the most prominent classical ones—namely, that requiring a hierarchy of truth predicates—as imposing, as a matter of principle, an unacceptable gap between the object language $\mathscr{C}_0$ expressing, say, snow's whiteness and the metalanguage $\mathscr{C}_1$ expressing truth-in-$\mathscr{C}_0$. But, so one of these criticisms goes, cannot a language $\mathscr{L}$ express truth-in-$\mathscr{L}$? On the face of it, natural languages can: on the face of it, 'true in English' in at least one of its uses in English expresses truth-in-English, 'vero in italiano' in at least one of its uses in Italian expresses truth-in-Italian, 'melauak ipan Mazeualtlahtolli' in at least one of its uses in Nahuatl

[2]Throughout, I use '*follow from*' and its relatives to denote the relation of logical consequence (broadly understood as to encompass also the "logic of truth" and the "logic of logical properties"), '*entail*' and its relatives to denote the converse relation and '*implication*' and its relatives to denote the operation expressed by a conditional. '*Equivalence*' and its relatives denote two-way entailment.

[3]The rather strange claim that 'God exists' is a logical truth followed given most medievals' conceptions of what logical truth is (unfortunately, something along the lines of necessary truth) and of what is necessary (which included God's existence). Both conceptions are no longer dominant. If you don't share both, just substitute your favourite logical truth for 'God exists'. I'll propose a particularly good candidate in section 2.

[4]I wish to leave the notion of *expressing* at an intuitive level. I would provisionally gloss 'expressing property p' with 'being or containing a word that has as extension (intension) the extension (intension) of p', if the latter did not involve the semantic or property-theoretic notions of extension or intension, whose proper understanding is at issue in the debates we're interested in.

[5]Throughout, I take languages to be individuated semantically, i.e. to come with their expressions *endowed of meaning*, and I understand meaning in a broad enough way as to suffice to *determine a logic* governing the language.

expresses truth-in-Nahuatl etc.

Yet, one might object that non-classical solutions are still guilty of fundamentally the same sin when, at least as a matter of fact, impose a gap between the object language $\mathscr{N}_0$ governed by their favoured non-classical logic and the metalanguage $\mathscr{N}_1$ expressing validity-in-$\mathscr{N}_0$, which is typically one of the languages of classical mathematics or its like. It is true that some parts of classical mathematics can be encoded by some part of $\mathscr{N}_0$ which behaves classically and which might contain the vocabulary in which the $\mathscr{N}_1$-model-theoretic definition of validity-in-$\mathscr{N}_0$ is formulated. However, even in that case, in non-classical solutions the (typically) model-theoretic definition in $\mathscr{N}_1$ of validity-in-$\mathscr{N}_0$—precisely because it behaves classically—has to treat at least some of that vocabulary as having in $\mathscr{N}_0$ a meaning at least partly *different* from the meaning it is intended to have in $\mathscr{N}_1$,[6] and it is in this very natural sense that, even in that case, it is still true that there is a gap between $\mathscr{N}_0$ and $\mathscr{N}_1$. And, in a very natural sense too, that gap is not closed until validity in a language is expressed in that language in the *logical specificity* of the language—that is, in a way that crucially depends on the language being governed by the non-classical logic it is governed by. But cannot a language $\mathscr{L}$ express validity-in-$\mathscr{L}$—and moreover do so in its logical specificity? On the face of it, natural languages can: on the face of it, 'valid in English', in at least one of its uses in English, expresses validity-in-English—and moreover does so in English's logical specificity—'valido in italiano', in at least one of its uses in Italian, expresses validity-in-Italian—and moreover does so in Italian's logical specificity—'nelli ipan Mazeualtlahtolli', in at least one of its uses in Nahuatl, expresses validity-in-Nahuatl—and moreover does so in Nahuatl's logical specificity— etc. Just as a language $\mathscr{L}$ that expresses truth-in-$\mathscr{L}$ is usually called '*semantically closed*', a language $\mathscr{L}$ that, in its logical specificity, expresses validity-in-$\mathscr{L}$ could aptly be called '*logically closed*'. Typically, the object language of non-classical solutions is semantically closed, but it is not logically closed.

I will argue that our two questions meet: that *the paradoxes of logical properties force one wishing to account for all the features involved in them by deploying a non-classical logic to express such logic in a non-classical metalanguage.* I will trace the route to this conclusion, and further beyond it to a concrete proposal as to how to express in a non-classical metalanguage a non-classical logic immune to the paradoxes of logical properties, by assuming the non-classical solution originally developed in Zardini [2011] in a classical metalanguage. Keeping fixed the classical metalanguage, I will develop further the theory presented in that paper in order to show how it can deal with a certain paradox of validity

[6] One way to see this starts by reflecting that the (typically) model-theoretic definition involves, among other things, the construction of models of $\mathscr{N}_0$ in which some sentences are assigned a "designated value". Given the assumption that some parts of classical mathematics are encoded by some part of $\mathscr{N}_0$ which behaves classically and which contains the vocabulary in which the $\mathscr{N}_1$-model-theoretic definition of validity-in-$\mathscr{N}_0$ is formulated, under further minimal assumptions it follows that, for every model $\mathbb{M}$, there is an $\mathscr{N}_0$-sentence $\lambda_{\mathbb{M}}$ which in $\mathscr{N}_1$ means that $\lambda_{\mathbb{M}}$ is not assigned a designated value in $\mathbb{M}$ and which, given classical logic, is either assigned a designated value in $\mathbb{M}$ (in which case it is not treated by $\mathbb{M}$ as really meaning that $\lambda_{\mathbb{M}}$ is not assigned a designated value in $\mathbb{M}$) or not assigned a designated value in $\mathbb{M}$ (in which case again it is not treated by $\mathbb{M}$ as really meaning that $\lambda_{\mathbb{M}}$ is not assigned a designated value in $\mathbb{M}$).

(in a way which I deem superior to most other extant non-classical solutions), and then, shifting to a non-classical metalanguage, develop it even further (or better, revise it) in order to show how it can express its own logical-consequence relation and deal more generally with the paradoxes of logical properties.

The rest of this paper is organised as follows. After some setting up, section 2 offers a budget of paradoxes involving the logical properties of validity, inconsistency and compatibility. Section 3 presents my favoured substructural approach to naive truth (as developed in other work), highlighting in particular its advantages in matters of validity with respect to other non-classical solutions. Section 4 explains how my favoured approach to naive truth can be extended in a very natural way so as to solve a certain paradox of validity. Section 5 provides however three individually decisive reasons for thinking that no approach adopting a classical metalanguage can adequately account for all the features involved in the paradoxes of logical properties. Section 6 undertakes then the task to do better, and, building on the system already developed, introduces a theory in a non-classical metalanguage that expresses an adequate logic of naive truth and of some naive logical properties. Section 7 recapitulates the main moves of the paper and draws some conclusions from them.

2 A Budget of Paradoxes

Right at the outset, let me be explicit about some assumptions I'll make throughout concerning a relation of logical consequence. I'll assume that a relation of logical consequence is a relation between *multisets* of sentences. Roughly, a multiset is like a set, but sensitive to the *number of occurrences* of a member in it. The assumption might be questioned, and I encourage the reader to pursue the study of the topics of this paper by relaxing it (fn 19 and Zardini [2013c] provide two different examples of families of logics in which the assumption is relaxed by taking collections of sentences that are *finer-grained* than multisets, while Zardini [2013h] defends the claim that, in any event, the relevant collections should have *sentences* as their members). However, I think that, for our purposes, the assumption can be taken to be correct, and making it will greatly simplify the dialectic (after all, my ultimate aim in this paper is to develop a particular theory rather than to provide a general map). Moreover, I'll assume that the multisets can be *empty* in both places of the relation of logical consequence, and (denoting the empty multiset with '$\varnothing$' and the logical-consequence relation with '$\vdash$') [that φ is a *logical truth* iff $\varnothing \vdash [\varphi]$ holds][7] and [that Γ is *inconsistent* iff $\Gamma \vdash \varnothing$ holds] (see Moruzzi and Zardini [2007] for more

[7]Throughout, I use square brackets for two different purposes (with context making clear which use is intended): to disambiguate constituent structure in English and to refer to multisets by listing the occurrences of their members (so that '$[\varphi]$' denotes the multiset in which φ occurs once and everything else does not occur, whereas '$[\varphi, \varphi]$' denotes the multiset in which φ occurs twice and everything else does not occur). When no danger of confusion threatens, I'll henceforth omit square brackets in the second use for the singleton case (and hence for example write 'φ' for '$[\varphi]$'). Also, I'll use the comma to denote the natural operation of sum between multisets (so that 'Γ, Δ' denotes the multiset that has all the occurrences of Γ plus all the occurrences of Δ and nothing else).

relevant background on logical-consequence relations).

With so much setting up in place, we can move to some outstanding paradoxes of logical properties. First, a paradox of validity. Let's assume a standard first-order language with, for every well-formed formula (henceforth, 'wff') φ, a canonical name $\ulcorner\varphi\urcorner$ referring to φ. Let $\vdash$ be the relevant logical-consequence relation and let V express (single-premise, single-conclusion) $\vdash$-validity (with $\vdash$ including the logic of validity). Naively, the following metarules would seem correct for V:[8]

$$\frac{\Gamma_0 \ \vdash \ \Delta_0, \varphi \qquad \Gamma_1, \psi \ \vdash \ \Delta_1}{\Gamma_0, \Gamma_1, V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \ \vdash \ \Delta_0, \Delta_1} \, V\text{-L}^- \qquad\qquad \frac{\Gamma, \varphi \ \vdash \ \Delta, \psi}{\Gamma \ \vdash \ \Delta, V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)} \, V\text{-R}$$

with the restriction that in V-R all the members of Γ and Δ be *logical*^{V} (that is, be either a wff of the form $V(x, y)$ or be the negation, conjunction, disjunction or implication of logical^V wffs). Let's then call a predicate 'a *naive-validity* predicate' only if it exhibits the above logical behaviour.[9]

Unfortunately, intuitive as they are at first glance, V-L$^-$ and V-R—and thus naive validity—would seem to lead to catastrophe. We assume that the language contains a sentence $\digamma^-$ identical to $V(\ulcorner\digamma^-\urcorner, \ulcorner\varphi\urcorner)$, where φ is arbitrary. It would then seem that we could reason as follows. By reflexivity, both $\digamma^- \vdash \digamma^-$ and $\varphi \vdash \varphi$ hold, and hence, by V-L$^-$, $\digamma^-, V(\ulcorner\digamma^-\urcorner, \ulcorner\varphi\urcorner) \vdash \varphi$ holds. But, by definition of '$\digamma^-$', that is tantamount to $\digamma^-, \digamma^- \vdash \varphi$ holding, and hence, by contraction, $\digamma^- \vdash \varphi$ holds, and so, by V-R, $\varnothing \vdash V(\ulcorner\digamma^-\urcorner, \ulcorner\varphi\urcorner)$ holds. However, by definition of '$\digamma^-$', $\digamma^-, V(\ulcorner\digamma^-\urcorner, \ulcorner\varphi\urcorner) \vdash \varphi$ holding is also tantamount to $V(\ulcorner\digamma^-\urcorner, \ulcorner\varphi\urcorner), V(\ulcorner\digamma^-\urcorner, \ulcorner\varphi\urcorner) \vdash \varphi$ holding, and hence, by contraction, $V(\ulcorner\digamma^-\urcorner, \ulcorner\varphi\urcorner) \vdash \varphi$ holds. Thus, both $\varnothing \vdash V(\ulcorner\digamma^-\urcorner, \ulcorner\varphi\urcorner)$ and $V(\ulcorner\digamma^-\urcorner, \ulcorner\varphi\urcorner) \vdash \varphi$ hold, and hence, by transitivity, $\varnothing \vdash \varphi$ holds, and so, by monotonicity, $\vdash$ is trivial. (Let's label this paradox '(V$^-$)'.)

Second, a paradox of inconsistency. Let I express (single-sentence) $\vdash$-inconsistency (with $\vdash$ including the logic of inconsistency). Naively, the following metarules would seem correct for I:

$$\frac{\varphi \ \nvdash \ \varnothing}{I(\ulcorner\varphi\urcorner) \ \vdash \ \varnothing} \, I\text{-L} \qquad\qquad \frac{\varphi \ \vdash \ \varnothing}{\varnothing \ \vdash \ I(\ulcorner\varphi\urcorner)} \, I\text{-R}$$

Let's then call a predicate 'a *naive-inconsistency* predicate' only if it exhibits the above logical behaviour (with a comment analogous to that of fn 9 applying to this 'only if').

Unfortunately, intuitive as they are at first glance, I-L and I-R—and thus naive inconsistency—would seem to lead to catastrophe. We assume that the language contains a sentence ι identical to $I(\ulcorner\iota\urcorner)$. It would then seem that we could reason as follows.

[8] The reason for labelling the first metarule 'V-L$^-$' rather than 'V-L' will become clear in section 5.

[9] Only 'only if' because sections 4 and 5 will argue for a new necessary condition for counting as a naive-validity predicate.

Suppose that $\iota \nvdash \varnothing$ holds. Then, by I-L, $I(\ulcorner\iota\urcorner) \vdash \varnothing$ holds. But, by definition of 'ι', that is tantamount to $\iota \vdash \varnothing$ holding, and hence, by *reductio*, $\iota \vdash \varnothing$ holds. Therefore, by I-R, $\varnothing \vdash I(\ulcorner\iota\urcorner)$ holds. But, by definition of 'ι', that is tantamount to $\varnothing \vdash \iota$ holding. Thus, both $\varnothing \vdash \iota$ and $\iota \vdash \varnothing$ hold, and hence, by transitivity, $\varnothing \vdash \varnothing$ holds, and so, by monotonicity, $\vdash$ is trivial. (Let's label this paradox '(I)'.)

Third, a paradox of compatibility, making more precise the paradox briefly presented in section 1. Let C express $\vdash$-compatibility (with $\vdash$ including the logic of compatibility), and let $\neg$ be a negation operator. Naively, the following metarules would seem correct for C:

$$\frac{\varphi \vdash \psi}{C(\ulcorner\varphi\urcorner, \ulcorner\neg\psi\urcorner) \vdash \varnothing}\ C\text{-L} \qquad\qquad \frac{\varphi \nvdash \psi}{\varnothing \vdash C(\ulcorner\varphi\urcorner, \ulcorner\neg\psi\urcorner)}\ C\text{-R}$$

(at least understanding the compatibility of φ with ψ as, roughly, the *compossibility* of φ's *truth* and ψ's lack of *falsity*; see Zardini [2013e] for more details about this). Let's then call a predicate with the above logical behaviour 'a *naive-compatibility* predicate'.[10]

Unfortunately, intuitive as they are at first glance, C-L and C-R—and thus naive compatibility—would seem to lead to catastrophe. The point can be made probably at its sharpest by introducing in the language a 0ary *logical-truth* operator $\mathtt{t}$. Being 0ary, $\mathtt{t}$ is in effect a sentence, and its intended meaning is such that $\mathtt{t}$ may be thought of as expressing the *conjunction of all logical truths*, and hence such that the two-way metarule:

$$\frac{\varnothing \vdash \Delta}{\overline{\mathtt{t} \vdash \Delta}}\ \varnothing/\mathtt{t}$$

is compelling. We can now assume that the language contains a sentence κ identical to $C(\ulcorner\mathtt{t}\urcorner, \ulcorner\neg\kappa\urcorner)$. It would then seem that we could reason as follows. Suppose that $\mathtt{t} \nvdash \kappa$ holds. Then, by C-R, $\varnothing \vdash C(\ulcorner\mathtt{t}\urcorner, \ulcorner\neg\kappa\urcorner)$ holds. But, by definition of 'κ', that is tantamount to $\varnothing \vdash \kappa$ holding, and hence, by monotonicity (or $\varnothing/\mathtt{t}$), $\mathtt{t} \vdash \kappa$ holds, and so, by *reductio*, $\mathtt{t} \vdash \kappa$ holds (and, by $\varnothing/\mathtt{t}$, $\varnothing \vdash \kappa$ holds). Therefore, by C-L, $C(\ulcorner\mathtt{t}\urcorner, \ulcorner\neg\kappa\urcorner) \vdash \varnothing$ holds. But, by definition of 'κ', that is tantamount to $\kappa \vdash \varnothing$ holding. Thus, both $\varnothing \vdash \kappa$ and $\kappa \vdash \varnothing$ hold, and hence, by transitivity, $\varnothing \vdash \varnothing$ holds, and so, by monotonicity, $\vdash$ is trivial. (Let's label this paradox '(C)'.)

[10] I stress that, in this paper, my focus will be on validity and inconsistency rather than on compatibility, since, as will become apparent in section 6, my aim in this paper is to give an adequately strong theory of naive validity and inconsistency rather than of naive compatibility (as a consequence, for example, I'll rest content with this rough sketch of what it is to be a naive-compatibility predicate). I still include some discussion of naive compatibility and its paradoxes both in honour of the Pseudo-Scotus and in order to give a sense of one of the directions of future research (with the little pun at the end of the second paragraph of section 3 as an added bonus).

3 A Non-Contractive Theory of Naive Truth

These are then some of the most arresting paradoxes of logical properties. Their apparent similarity to the semantic paradoxes (as well as to the other paradoxes mentioned in section 1) is glaring. In what I take to be a rather superficial sense, it could be said that they are even more similar to the semantic paradoxes than, say, the set-theoretic paradoxes, as they similarly involve principles exhibiting some sort of *semantic ascent* or *descent* (although in the case of the paradoxes of logical properties at least the ascent is naturally understood as being in turn made within an implicit quotation environment; I delve more into a defence of the compellingness of the principles of naive logical properties in Zardini [2013e]). I mentioned in section 1 the disagreements among defenders of non-classical solutions as to how far their favoured non-classical logic should be deployed in order to provide a structurally identical solution to other apparently similar paradoxes. I take it that the paradoxes of logical properties are at least similar *enough* as to qualify as *apparently* similar (Priest [2003] for example has a detailed account of an interesting sufficient condition triggering the requirement that a paradox be solved by deploying the same non-classical logic applied to the semantic paradoxes, condition which is satisfied in this case),[11] and hence that there is space for an analogous disagreement about this particular case. The appealing if vague working hypothesis which—without further defence—I myself wish to assume in the rest of this paper is that a good non-classical solution accommodating for naive semantic properties should *also* be able to accommodate for naive logical properties—naive validity, inconsistency and compatibility should be saved just as much as naive truth. But just how?

In order to start the constructive part of this paper, it will be useful for the time being (and until section 5) to focus on paradox (V$^-$). One obvious way to start is to take a non-classical solution and simply add on top of it V-L$^-$ and V-R. Can this be done without falling afoul of paradox (V$^-$)? Inspection of the reasoning involved in paradox (V$^-$) shows

[11] Essentially, the condition is that the paradox in question can apparently be shown to have an "*inclosure*" structure with a set Z, properties F and G and a function dia such that:

(i) $Z = \{x : x \text{ is } F\}$ and Z is G;

(ii) If $X \subseteq Z$ and X is G:

 (iia) $\mathsf{dia}(X) \notin X$;

 (iib) $\mathsf{dia}(X) \in Z$

(a structure which straightforwardly yields a contradiction given that $Z \subseteq Z$). Taking for example paradox (V$^-$), let F be the property of being true, let G be the property of being definable using an expressible property (thus vindicating (i)) and, if $X \subseteq Z$ and is definable as $\{x : x \text{ is } P\}$, let $\mathsf{dia}(X)$ be the sentence F^-_X identical to $V(\ulcorner\ulcorner\mathsf{F}^-_X\urcorner \in \{x : x \text{ is } P\}\urcorner, \ulcorner\varphi\urcorner)$ (where φ is absurd). Then, using among other things naive principles for truth and V-L$^-$, we can apparently show that $\mathsf{dia}(X) \in X \vdash \varphi$ holds (very much following the first steps of paradox (V$^-$)), from which it follows that either φ holds or $\mathsf{dia}(X) \notin X$, and hence, since φ is absurd, that $\mathsf{dia}(X) \notin X$ (thus vindicating (iia)). Moreover, since we can apparently show that $\mathsf{dia}(X) \in X \vdash \varphi$ holds, essentially by V-R it follows that $\varnothing \vdash V(\ulcorner\ulcorner\mathsf{F}^-_X\urcorner \in \{x : x \text{ is } P\}\urcorner, \ulcorner\varphi\urcorner)$ holds (very much following the immediately further steps of paradox (V$^-$)), and hence, by naive principles for truth, that $\mathsf{dia}(X)$ is true, and so that $\mathsf{dia}(X) \in Z$ (thus vindicating (iib)).

that, for most non-classical solutions (such as e.g. Kripke [1975]; McGee [1991]; Gupta and Belnap [1993]; Brady [2006]; Priest [2006]; Field [2008]; Beall [2009]), this cannot be done, since they accept all the principles used in that reasoning in addition to V-L$^-$ and V-R: reflexivity, contraction,[12] transitivity and monotonicity. I regard the inability to accommodate for naive logical properties—as demonstrated by paradoxes (V$^-$), (I) and (C)—as a major flaw of these theories (with a little pun, one could call this the '(V$^-$)(I)(C)torious argument against non-substructural non-classical solutions').[13]

Fortunately, however, there are some theories that do reject some of those principles: Greenough [2001] can be read as implicitly rejecting reflexivity, Zardini [2011] rejects contraction and Ripley [2012] rejects the relevant version of transitivity.[14] While non-reflexive, non-transitive and non-monotonic approaches deserve further consideration (I make a start on a critical discussion of non-transitive approaches in Zardini [2013b], pp. 578–580; [2013f], and discuss non-transitive and non-monotonic approaches in particular in relation to the paradoxes of logical properties in Zardini [2013e]), in the rest of this paper I'll assume and develop further the theory of Zardini [2011] (see also Zardini [2013a]; [2013b]; [2013d]; [2013f]; [2013g] for developments of the theory in other direc-

[12]Very confusingly, 'contraction' is often used, especially in the literature on the semantic paradoxes, as a name not for the *structural metarule*, but for the *law* $\varphi \to (\varphi \to \psi) \to (\varphi \to \psi)$. Most non-classical solutions reject the law of contraction but accept the structural metarule of contraction. I have been using and will continue to use 'contraction' exclusively to refer to the structural metarule. Analogous comments apply to 'reflexivity', 'transitivity' and 'monotonicity'.

[13]Paradoxes (V$^-$) and (C) have been around for a while. While Pseudo-Scotus' version of paradox (C) dates back to the 14$^{\text{th}}$ century (see Read [2001], p. 184), versions of paradox (V$^-$)—which is basically the contrapositive of paradox (C)—are attested at least from the 16$^{\text{th}}$ century (see Ashworth [1974], p. 125). In modern times, both these paradoxes have been widely discussed in one version or another (first among historians of logic and then among analytic philosophers, with Mates [1965] and Read [1979] having served as bridges between the former community and the latter one). There is much to learn, and to debate, in these modern works, whose list is too long to be given here (somewhat invidiously, in addition to mentioning Priest [2010], pp. 127–128 as a recent place in which the principles of naive validity operative in paradox (V$^-$) are clearly isolated, sympathetically treated and interestingly brought to bear on issues of relevance for non-classical solutions, let me single out the two recent pieces Shapiro [2011] and Beall and Murzi [2013] as being in many respects very congenial to my own approach in this paper; I discuss some crucial differences between Shapiro's treatment of paradox (V$^-$) and mine in Zardini [2013b]). This paper agrees with those authors that use the paradoxes of logical properties to argue against the adequacy of certain non-classical solutions, and tries to push forward the debate by providing a provably consistent theory at least able to deal with paradox (V$^-$) (in sections 3 and 4), by assessing the full import of these paradoxes (in section 5) and by providing a provably consistent theory able to deal with that import (in section 6). Thanks to an anonymous referee for discussion of some of this literature.

[14]I myself am in general sympathetic to restricting transitivity, and have developed a proposal as to how to do so in the case of *vagueness* (see for example Zardini [2008b], which in fact defines the general family of logics to which Ripley's belongs). In the case of the semantic paradoxes, however, for reasons going beyond this paper and partially exposed in the three works to be immediately referenced in the text, I favour a non-contractive approach over a non-transitive one. I should also mention that, before both Ripley's and my work, Alan Weir already developed a non-transitive theory of truth (and sets) with possible application to vagueness too (see Weir [2005] for a recent presentation). However, Weir's theory does validate the instances of transitivity used in paradoxes (V$^-$), (I) and (C), since on that theory—contrary to both Ripley's and my theory—transitivity holds in the absence of side premises and conclusions.

tions). More precisely, since the real action—as far as the problems this paper engages with are concerned—is at the *sentential* level and since, given some distinctive features of that theory, extending what I'll be doing in this paper to the first order would require some fairly lengthy and distracting complications, I'll assume and develop further the quantifier-free fragment of the theory (which, for reasons that will shortly be obvious, I'll call '**IKT**'). I stress that focussing on the quantifier-free fragment is compatible with discerning at least the subsentential structure of *predication* (which is of course desirable in a discussion concerning truth and validity attributions!).

IKT can very naturally be presented in sequent-calculus style. The *background logic* is defined as the smallest logic containing as axiom the structural rule:

$$\frac{}{\varphi \vdash_{\mathbf{IKT}} \varphi}\ \mathrm{I}$$

and closed under the structural metarules:

$$\frac{\Gamma_0 \vdash_{\mathbf{IKT}} \Delta}{\Gamma_0, \Gamma_1 \vdash_{\mathbf{IKT}} \Delta}\ \text{K-L} \qquad \frac{\Gamma \vdash_{\mathbf{IKT}} \Delta_0}{\Gamma \vdash_{\mathbf{IKT}} \Delta_0, \Delta_1}\ \text{K-R}$$

$$\frac{\Gamma_0 \vdash_{\mathbf{IKT}} \Delta_0, \varphi \qquad \Gamma_1, \varphi \vdash_{\mathbf{IKT}} \Delta_1}{\Gamma_0, \Gamma_1 \vdash_{\mathbf{IKT}} \Delta_0, \Delta_1}\ \mathrm{S}$$

and under the operational metarules:[15]

$$\frac{\Gamma \vdash_{\mathbf{IKT}} \Delta, \varphi}{\Gamma, \neg\varphi \vdash_{\mathbf{IKT}} \Delta}\ \neg\text{-L} \qquad \frac{\Gamma, \varphi \vdash_{\mathbf{IKT}} \Delta}{\Gamma \vdash_{\mathbf{IKT}} \Delta, \neg\varphi}\ \neg\text{-R}$$

$$\frac{\Gamma, \varphi, \psi \vdash_{\mathbf{IKT}} \Delta}{\Gamma, \varphi \,\&\, \psi \vdash_{\mathbf{IKT}} \Delta}\ \&\text{-L} \qquad \frac{\Gamma_0 \vdash_{\mathbf{IKT}} \Delta_0, \varphi \qquad \Gamma_1 \vdash_{\mathbf{IKT}} \Delta_1, \psi}{\Gamma_0, \Gamma_1 \vdash_{\mathbf{IKT}} \Delta_0, \Delta_1, \varphi \,\&\, \psi}\ \&\text{-R}$$

$$\frac{\Gamma_0, \varphi \vdash_{\mathbf{IKT}} \Delta_0 \qquad \Gamma_1, \psi \vdash_{\mathbf{IKT}} \Delta_1}{\Gamma_0, \Gamma_1, \varphi \vee \psi \vdash_{\mathbf{IKT}} \Delta_0, \Delta_1}\ \vee\text{-L} \qquad \frac{\Gamma \vdash_{\mathbf{IKT}} \Delta, \varphi, \psi}{\Gamma \vdash_{\mathbf{IKT}} \Delta, \varphi \vee \psi}\ \vee\text{-R}$$

[15]Throughout, faithful to its origins as simply evocative of Latin 'vel' (and for lack of a better alternative), I understand $\vee$ as *neutrally* referring to our informal notion of (inclusive) disjunction (as it occurs for example in informal presentations of the semantic paradoxes), *without any prejudice to the details of its logical behaviour*. In particular, to forestall a natural association that the symbol might trigger for those familiar with lattices, I emphasise that I do *not* presuppose that disjunction has "join behaviour" (in fact, in **IKT** and its further developments in this paper, disjunction does fail to obey some basic lattice principles like for example subidempotency; see also fn 16). Thanks to an anonymous referee for advice about notation and for urging this clarification.

$$\frac{\Gamma_0 \vdash_{\mathbf{IKT}} \Delta_0, \varphi \qquad \Gamma_1, \psi \vdash_{\mathbf{IKT}} \Delta_1}{\Gamma_0, \Gamma_1, \varphi \to \psi \vdash_{\mathbf{IKT}} \Delta_0, \Delta_1} \to\text{-L} \qquad \frac{\Gamma, \varphi \vdash_{\mathbf{IKT}} \Delta, \psi}{\Gamma \vdash_{\mathbf{IKT}} \Delta, \varphi \to \psi} \to\text{-R}$$

where Γ and its likes are *finite* (in the strong sense of having finitely many members that occur finitely many times).

IKT can be extended to a *logic of truth* by adding the following metarules for a truth predicate T:

$$\frac{\Gamma, \varphi \vdash_{\mathbf{IKT}} \Delta}{\Gamma, T(\ulcorner \varphi \urcorner) \vdash_{\mathbf{IKT}} \Delta} T\text{-L} \qquad \frac{\Gamma \vdash_{\mathbf{IKT}} \Delta, \varphi}{\Gamma \vdash_{\mathbf{IKT}} \Delta, T(\ulcorner \varphi \urcorner)} T\text{-R}$$

In Zardini [2011], I go through in much detail the background-logical and truth-theoretic strength of **IKT** (while proving its consistency), especially in the surprisingly many respects of philosophically interesting strength in which it outperforms other non-classical solutions (indeed, I do all this for my favoured **IKT**'s extension to the first order). On the background-logical side, it's easy to see that **IKT** validates the law of excluded middle, the law of non-contradiction, certain restricted versions of the metarules of *reductio* and reasoning by cases (roughly, those very important restricted versions where the conclusion of the supposition is inconsistent), the De Morgan equivalences, double-negation introduction and elimination (so that disjunction is definable by conjunction and negation, and conjunction definable by disjunction and negation), the rules of simplification and adjunction for conjunction and of addition and abjunction ($\varphi \vee \psi \vdash_{\mathbf{IKT}} \varphi, \psi$) for disjunction,[16] the rule of *modus ponens* and the deduction theorem for implication (plus other principles that are characteristic of material implication, like the so-called 'paradoxes of material implications', so that the conditional is in effect definable by conjunction and negation

[16]The metarules given for & and $\vee$ are essentially the metarules for the "*multiplicative*" operators *tensor* and *par* of *linear* logics and the "*intensional*" operators *fusion* and *fission* of *relevant* logics. In both these kinds of logics, such operators are opposed to the "*additive*" or "*extensional*" operators, which are typically supposed to express *our informal notions of conjunction and disjunction* (as they occur for example in informal presentations of the semantic paradoxes). Against a likely misunderstanding, and expanding on an issue already touched on in fn 15, I cannot emphasise enough that, with **IKT**'s & and $\vee$, I intend to give a *theory of precisely our informal notions of conjunction and disjunction* (hence my use of the standard symbolism), and that I take this interpretation to be warranted by the fact the last four rules just mentioned in the text are valid: together, those rules amount to saying, a bit roughly, that *a conjunction is true iff both of its conjuncts are true*, and that *a disjunction is true iff either of its disjuncts is*, which I take to capture the core of our informal notions of conjunction and disjunction. The divergence of my interpretation with the interpretation typically given for linear and relevant logics is explainable by the fact that those logics *lack* K-L *and* K-R, which determines that at least some of the rules just mentioned in the text are not valid for their "multiplicative" or "intensional" operators (whence these logics' search for other operators that can represent more adequately our informal notions of conjunction and disjunction). I should also stress that **IKT**—as well as its further developments in this paper—can consistently be extended by adding conjunction-like and disjunction-like "additive" or "extensional" operators; it is just that, for the reasons developed in Zardini [2011]; [2013a]; [2013b]; [2013f]; [2013g], I wouldn't recommend taking these to represent our informal notions of conjunction and disjunction.

or by disjunction and negation). Moreover, it's also easy to see that contraction can be locally recovered for a set of sentences X by adding, for every $\varphi \in X$, $\varphi \to \varphi \mathbin{\&} \varphi$[17] as a further axiom to the theory. On the truth-theoretic side, it's easy to see that **IKT** validates the rules of T-introduction ($\varphi \vdash_{\mathbf{IKT}} T(\ulcorner\varphi\urcorner)$) and T-elimination ($T(\ulcorner\varphi\urcorner) \vdash_{\mathbf{IKT}} \varphi$), the Tarskian biconditionals, the full intersubstitutability of φ with $T(\ulcorner\varphi\urcorner)$ (henceforth, '*transparency*'),[18] the standard truth-functional laws and the traditional constraint of truth preservation on logical consequence.

The last point is actually very much germane to paradox (V^-). Very interestingly, it is a surprising feature of many non-classical solutions (for example, those of Priest [2006] and Field [2008]) that they validate *modus ponens* but cannot validate the law of *truth preservingness of modus ponens* $T(\ulcorner\varphi\urcorner) \mathbin{\&} T(\ulcorner\varphi \to \psi\urcorner) \to T(\ulcorner\psi\urcorner)$, in the sense that adding that law to them generates inconsistency (the point has forcefully been made by Field [2006], pp. 588–599 in the context of a wider investigation on classical and non-classical solutions and principles of truth preservation). The reason for this is relatively straightforward and, in its essence, has first been pointed out by Meyer et al. [1979]. Take a Curry sentence ρ identical to $T(\ulcorner\rho\urcorner) \to \varphi$, where φ is arbitrary, and instantiate the law of truth preservingness of *modus ponens* with $T(\ulcorner\rho\urcorner)$ for the antecedent and φ for the consequent. By definition of 'ρ' and transparency (which holds on many non-classical solutions), that instance is equivalent with $T(\ulcorner\rho\urcorner) \mathbin{\&} T(\ulcorner\rho\urcorner) \to \varphi$, which, on most non-classical solutions, is in turn equivalent with $T(\ulcorner\rho\urcorner) \to \varphi$ (*since, on most non-classical solutions, ψ is equivalent and intersubstitutable with $\psi \mathbin{\&} \psi$*). *And that in turn entails φ on most non-classical solutions.* I think that such inconsistency with the law of truth preservingness of *modus ponens* is extremely problematic, especially when coupled with acceptance of unrestricted *modus ponens* (I've sometimes heard in conversation sympathisers of these non-classical solutions using the previous argument to contend that logical consequence does not require truth preservation, rather than—more correctly, it seems to me—taking the argument to indicate a serious problem with their theories of truth).[19]

---

[17] To save on brackets, I will assume the usual scope hierarchy among the operators (with $\neg$ binding more strongly than $\&$ and $\vee$, and with these in turn binding more strongly than $\to$ and other conditionals) and right associativity for each 2ary operator (so that $\varphi_0 \star \varphi_1 \star \varphi_2 \ldots \star \varphi_i$ reads $\varphi_0 \star (\varphi_1 \star (\varphi_2 \ldots \star \varphi_i))) \ldots)$, with $\star$ being a 2ary operator).

[18] I myself am no supporter of unrestricted transparency (see Zardini [2008a]; [2013i]), but it will do no harm to assume it in this paper.

[19] Priest [2010], pp. 133–134 grants that what I've called 'the law of truth preservingness of *modus ponens*' fails on his theory, but, noting that $T(\ulcorner\varphi \to \psi\urcorner) \to T(\ulcorner\varphi\urcorner) \to T(\ulcorner\psi\urcorner)$ still holds, claims that that "still serves as a statement of the facts". The claim is problematic at several levels. First, it is far from clear that the facts stated by Priest's proposal would always be the facts of truth preservation *of modus ponens*. If τ_0 is identical to $T(\ulcorner\tau_0\urcorner)$ and τ_1 is identical to $T(\ulcorner\tau_1\urcorner)$, $T(\ulcorner\tau_0 \to \tau_1\urcorner) \to T(\ulcorner\tau_0\urcorner) \to T(\ulcorner\tau_1\urcorner)$ is really $T(\ulcorner T(\ulcorner\tau_0\urcorner) \to T(\ulcorner\tau_1\urcorner)\urcorner) \to T(\ulcorner\tau_0\urcorner) \to T(\ulcorner\tau_1\urcorner)$ (since $\tau_0 \to \tau_1$ is identical to $T(\ulcorner\tau_0\urcorner) \to T(\ulcorner\tau_1\urcorner)$, and hence $\ulcorner\tau_0 \to \tau_1\urcorner$ is identical to $\ulcorner T(\ulcorner\tau_0\urcorner) \to T(\ulcorner\tau_1\urcorner)\urcorner$), which seems to state no more than a certain instance of the disquotation principle for truth. Second, it is very unclear how Priest's proposal can be extended to cover arguments with *infinitely many* premises. Third, the general (and traditionally recognised) fact about logical consequence and truth preservation is that, if *every* premise of a valid argument is true, so is *some* conclusion of the argument: I don't know of any plausible way to understand this general fact so that Priest's proposal about *modus ponens* turns out to be an instance of it.

And it is not only the *semantic* law of truth-preservingness of *modus ponens* that fails on these non-classical solutions: its *non-semantic* analogue, i.e. the law $\varphi \,\&\, (\varphi \rightarrow \psi) \rightarrow \psi$, fails on these non-classical solutions for just about the same reasons (indeed, fewer principles about truth are needed in order to derive an arbitrary sentence from it). The law is sometimes very tendentiously called 'pseudo *modus ponens*', but that seems the invidious reaction of someone who's shot himself in the foot by endorsing a logic of truth inconsistent with it. The law is quite properly called 'the law of *modus ponens*' *tout*

Fourth, Priest's proposal carefully goes for $T(\ulcorner\varphi \rightarrow \psi\urcorner) \rightarrow T(\ulcorner\varphi\urcorner) \rightarrow T(\ulcorner\psi\urcorner)$ instead of the "permuted" $T(\ulcorner\varphi\urcorner) \rightarrow T(\ulcorner\varphi \rightarrow \psi\urcorner) \rightarrow T(\ulcorner\psi\urcorner)$. It has to, as, in Priest's general framework, the latter option would lead into trouble. However, if the two options are not equivalent, we're clearly owed an account of how, when stating the traditional constraint of truth preservation on logical consequence, the premises of an argument should be *ordered*—we cannot rest content with an *ad hoc* if judicious case-by-case choice (while to leave it at a disjunction of the two options would simply be to shun a glaring issue that the proposal has raised). Such account is still missing, and, given the apparent *commutativity* of premise combination, I myself think that it won't be easy to come by. Fifth, as Priest himself recognises, when extended to cover arguments other than *modus ponens*, in his general framework the proposal as stated does not work on either option for arguments like *adjunction*, for which Priest says that we should stick to the original conjunctive version of truth preservation. However, even granting that the disjunction mentioned in the previous point can be decided, this makes the proposal irremediably *disjunctive*, and objectionably so, given that there does not seem to be anything disjunctive about the truth preservation that logical consequence requires (it does not seem that *modus ponens* preserves truth in a sense different from the one in which adjunction does). Sixth, the proposal, however extended to multiple-premise arguments, does nothing to address the related problem for *single-premise* arguments. For example, the problem arises already for *modus ponens*, since its equally good rendition as $\varphi \,\&\, (\varphi \rightarrow \psi) \vdash \psi$ is valid in Priest's general framework, but, for essentially the same reason as the one given in the text, in the framework the truth-preservation conditional for that argument is inconsistent (an example involving the logic of truth is that $T(\ulcorner\rho\urcorner) \vdash \varphi$ is valid in Priest's general framework, but, for essentially the same reason as the one given in the text, in the framework the truth-preservation conditional for that argument is inconsistent). Priest [2013] is an interesting variation on Priest [2010], pp. 133–134, basically taking the bull by the horns and, roughly, proposing that an argument is valid iff it is a logical truth that it is truth preserving ('truth preserving' in the specific sense of Priest [2010], pp. 133–134). Granting that this proposal can somehow tackle the first two points I've made, the last four points represent themselves in the form that premise combination is now not only *non-contractive*, but also *non-commutative*, *non-associative* and *non-monotonic*, and that adjunction and *modus ponens* in its renditions as $\varphi \,\&\, (\varphi \rightarrow \psi) \vdash \psi$ and even as $\varphi, \varphi \rightarrow \psi \vdash \psi$ are now *not valid*. One can extract from some observations in Priest [2013] a further variation (different from the official view of Priest [2013] just discussed) which, roughly, would claim that an argument is valid iff the set of its premises can be transformed into a "bunch" (a very fine-grained kind of collection whose details needn't detain us here) that entails the conclusion ('entails' in the specific sense of Priest [2013]). This proposal would return to a more standard, non-substructural notion of validity: it would thus have the same *original* problem with *truth preservation* that affects non-substructural non-classical solutions (and hence, assuming the theory of truth preservation of Priest [2013], have at least the specific problems emerging with the last four points I've made), and also have the *additional* problem of *not validating modus ponens* in its rendition as $\varphi \,\&\, (\varphi \rightarrow \psi) \vdash \psi$. I've included this rather extended discussion of Priest's proposals not only because of the proposals' intrinsic interest, but also because this allows me to make a more general observation. For all I know, it is possible that, in the frameworks in which the law of truth preservingness of *modus ponens* fails, one could still define or introduce a premise-combining operator in terms of which to formulate alternative truth-preservation conditionals, and that, in the resulting theories, these could hold in many cases in which their corresponding arguments are valid. But, even though, for all I know, it is possible—if very unlikely—that one could do so in such a way as to avoid the first five points I've made, there is precious little that on such scheme could be done to avoid the last point. Thanks to two anonymous referees for encouraging this discussion.

court—it expresses the most natural *proposition* that corresponds to the *rule* of *modus ponens*, namely the proposition that φ and $\varphi \rightarrow \psi$ are jointly a sufficient condition for ψ.[20]

Given that just about every non-classical solution inconsistent with the *law* of truth preservingness of *modus ponens* and the *law* of *modus ponens* does accept the *rule* of *modus ponens* (even when its premises are conjoined, see fn 19), these non-classical solutions in effect *reject the deduction theorem* (which would allow to go from the rule of *modus ponens* to the law of *modus ponens*, while transparency would allow to go from the law of *modus ponens* to the law of truth-preservingness of *modus ponens*). This is just another instance of the crucial role played in these theories by the rejection of the deduction theorem (the paradigmatic instance being their treatment of the version of Curry's paradox using the deduction theorem). Now, I think it is arguable that failure of the deduction theorem is a major flaw of these theories, but I won't belabour the point here (see Zardini [2011]; [2013d]; [2013f] for some discussion of this). The important point to be made here is rather that, no matter whether failure of the deduction theorem can eventually be sustained for the conditional (at least for those that satisfy the rule of *modus ponens*), it would seem rather bizarre to think that its analogue—i.e. V-R—fails for a validity predicate: if φ does entail ψ, no matter whether the conditional $\varphi \rightarrow \psi$ is a logical truth, it would seem rather bizarre to think that $V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$, which is supposed to say just that, is not a logical truth (once logic is supposed to encompass the "logic of validity", see fn 2). In fact, simply accepting $V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ (no matter whether as a logical truth or not) would presumably be sufficient, by V-L$^-$, to generate paradox; however, if φ does entail ψ, it would seem even more bizarre not to accept $V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$, which is supposed to say just that. That, I take it, is by and large the distinctive punch of paradox (V$^-$).

Notice that the previous argument against the consistency of the law of truth preservingness of *modus ponens* and of the law of *modus ponens* fails for **IKT**, for the simple reason that, in **IKT**, *ψ is not equivalent with $\psi \,\&\, \psi$*, as the former does not entail the latter (which is the crucial direction employed in the previous argument; the converse direction is indeed valid in **IKT**, as it is just a particular case of simplification). Indeed, it's easy to see that both laws are valid in **IKT**. More generally, it's easy to see that, if $\Gamma, \varphi_0, \varphi_1, \varphi_2 \ldots, \varphi_i \vdash_{\mathbf{IKT}} \Delta, \psi_0, \psi_1, \psi_2 \ldots, \psi_j$ holds, both $\Gamma \vdash_{\mathbf{IKT}} \Delta, T(\ulcorner\varphi_0\urcorner) \,\&\, T(\ulcorner\varphi_1\urcorner) \,\&\, T(\ulcorner\varphi_2\urcorner) \ldots \,\&\, T(\ulcorner\varphi_i\urcorner) \rightarrow T(\ulcorner\psi_0\urcorner) \vee T(\ulcorner\psi_1\urcorner) \vee T(\ulcorner\psi_2\urcorner) \ldots \vee T(\ulcorner\psi_j\urcorner)$ and $\Gamma \vdash_{\mathbf{IKT}} \Delta, \varphi_0 \,\&\, \varphi_1 \,\&\, \varphi_2 \ldots \,\&\, \varphi_i \rightarrow \psi_0 \vee \psi_1 \vee \psi_2 \ldots \vee \psi_j$ hold: in **IKT**, not only *modus ponens*, but *every* valid rule requires the corresponding truth-preservation conditional as well as its non-semantic analogue (with this being so even under side premises and conclusions).[21] All these appealing validities have their root in the validity of the

[20] This might be what is intended with the other common label for the law of *modus ponens*, namely 'assertion': the law of *modus ponens* stands to *assertion of a proposition* as the rule of *modus ponens* stands to *inference according to a rule* (I discuss issues surrounding several principles of *modus ponens* further in Zardini [2013b]; [2013f]).

[21] More precisely, given our restriction to a quantifier-free language, what is the case in **IKT** is that every valid rule requires all the *instances* of the corresponding truth-preservation conditional as well as all the *instances* of its non-semantic analogue (with this being so even under side premises and conclusions).

deduction theorem for **IKT**, encapsulated in $\to$-R.

4 Towards a Non-Contractive Theory of Naive Validity

Now, it should already have been clear that one can add on top of **IKT** V-L$^-$ and V-R without at least running afoul of paradox (V$^-$)—after all, that paradox requires contraction, and **IKT** is not even closed under that metarule (as proved in Zardini [2011], pp. 524–532). But, as always, the interesting question is not so much whether a *particular* paradoxical reasoning goes through in a certain theory, but whether the theory is proof against *any* paradoxical reasoning whatsoever—roughly, whether the theory is consistent. As I've already mentioned, Zardini [2011] does prove that **IKT** is consistent, but it is not immediately clear whether adding on top of it V-L$^-$ and V-R preserves its consistency. Fortunately, armed with the previous considerations concerning the traditional constraint of truth preservation on logical consequence, we can easily see that there is a simple argument showing that such addition preserves consistency.

The argument relies on the observation that *predicates obeying analogues of* V-L$^-$ *and* V-R *are definable in* **IKT**. To take the simplest example, consider the 2ary predicate $T(x) \to T(y)$. It's easy to see that this predicate obeys the relevant analogues of V-L$^-$ and V-R: *modulo* transparency, the analogue of V-L$^-$ is given by $\to$-L and the analogue of V-R is given by $\to$-R. Given this, the consistency of the system resulting from adding V-L$^-$ and V-R on top of **IKT** virtually follows. (Alternatively, a simple extension of the proof given in Zardini [2011], pp. 524–532 as to encompass a system including V-L$^-$ and V-R will do.)

Thus, $T(x) \to T(y)$ obeys the relevant analogues of V-L$^-$ and V-R, but can it sensibly be regarded as a naive-validity predicate? That it obeys the relevant analogues of V-L$^-$ and V-R does not suffice to show that it can, for it may obey some other principle which is intuitively not sound for (naive) validity. Indeed, the extremely tight connection with $\to$ brings not only good news to the prospects of $T(x) \to T(y)$ for being regarded as a naive-validity predicate. It also brings bad news in those respects in which $\to$ has quite properly been designed to be an ordinary conditional, and hence, at least in some respects, *weaker than a validity predicate.* Given its presentation in sequent-calculus style, in **IKT** (as elsewhere) weaknesses quite generally emerge on the right, and hence the respects in which $\to$ is weaker than a validity predicate emerge with $\to$-R, and in particular with its instances *with (logically contingent) side premises or conclusions.* By transparency, these respects carry over to $T(x) \to T(y)$. To take a paradigmatic example, by I and K-L, $\varphi, \psi \vdash_{\mathbf{IKT}} \varphi$ holds, and hence, by $\to$-R (with side premises), $\varphi \vdash_{\mathbf{IKT}} \psi \to \varphi$ holds, and so, by transparency, $\varphi \vdash_{\mathbf{IKT}} T(\ulcorner\psi\urcorner) \to T(\ulcorner\varphi\urcorner)$ holds. But it does not look like $\varphi \vdash V(\ulcorner\psi\urcorner, \ulcorner\varphi\urcorner)$ should hold (at least when $\psi \vdash \varphi$ does not hold): it does not look like the argument from an arbitrary ψ to φ should be valid, and it does not look like

Thanks to an anonymous referee for urging this clarification.

the situation should change if it is further assumed that φ is true—it rather looks like the argument should remain just as not valid as before (although, under that further assumption, it is now guaranteed to be truth preserving).[22]

That might not be so worrying if we stick to adding V on top of **IKT** as a *primitive* naive-validity predicate obeying V-L$^-$ and V-R—in that case, the only point of $T(x) \rightarrow T(y)$ is to help to show that such addition preserves consistency. But it is worrying if we're instead aiming at *defining* V in **IKT**, for the previous example shows that the most obvious candidate—i.e. $T(x) \rightarrow T(y)$—won't do. As it'll prove extremely important for our purposes to have indeed a definition of V in terms of a conditional, the natural course of action is to add a new conditional $\Rightarrow$ on top of **IKT**, thereby obtaining a new logic $\mathbf{IKT}_\Rightarrow$ closed under the additional operational metarules:

$$\frac{\Gamma_0 \vdash_{\mathbf{IKT}_\Rightarrow} \Delta_0, \varphi \qquad \Gamma_1, \psi \vdash_{\mathbf{IKT}_\Rightarrow} \Delta_1}{\Gamma_0, \Gamma_1, \varphi \Rightarrow \psi \vdash_{\mathbf{IKT}_\Rightarrow} \Delta_0, \Delta_1} \Rightarrow\text{-L} \qquad \frac{\Gamma, \varphi \vdash_{\mathbf{IKT}_\Rightarrow} \Delta, \psi}{\Gamma \vdash_{\mathbf{IKT}_\Rightarrow} \Delta, \varphi \Rightarrow \psi} \Rightarrow\text{-R}$$

with the restriction that in $\Rightarrow$-R all the members of Γ and Δ be *logical*$^\Rightarrow$ (that is, be either a wff of the form $P \Rightarrow Q$ or be the negation, conjunction, disjunction or $\rightarrow$-implication of logical$^\Rightarrow$ wffs). Again, the consistency of $\mathbf{IKT}_\Rightarrow$ virtually follows from that of **IKT**. (Alternatively, a simple extension of the proof given in Zardini [2011], pp. 524–532 as to encompass a system including $\Rightarrow$-L and $\Rightarrow$-R will do.) Moreover, $\Rightarrow$ has explicitly been designed to avoid the problems of the weakness of $\rightarrow$ generated by $\rightarrow$-R, more specifically by its instances with logically contingent side premises or conclusions.[23] Thus, we can reasonably take $T(x) \Rightarrow T(y)$ as a better definition of V than $T(x) \rightarrow T(y)$, in the "*intersubstitutability sense*" that it could be shown that, if we had primitive V as well in the logic, while $T(\ulcorner\varphi\urcorner) \rightarrow T(\ulcorner\psi\urcorner)$ would not be intersubstitutable with $V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ (to go back to the above example, $\varphi \vdash_{\mathbf{IKT}} T(\ulcorner\psi\urcorner) \rightarrow T(\ulcorner\varphi\urcorner)$ holds whereas $\varphi \vdash V(\ulcorner\psi\urcorner, \ulcorner\varphi\urcorner)$ does not hold), $T(\ulcorner\varphi\urcorner) \Rightarrow T(\ulcorner\psi\urcorner)$ would (if we also count every logical^V wff as logical$^\Rightarrow$ and *vice versa*).[24]

[22]Notice that to give up K-L and thereby block the implicit step from $\varphi \vdash_{\mathbf{IKT}} \varphi$ holding to $\varphi, \psi \vdash_{\mathbf{IKT}} \varphi$ holding won't offer a fully general solution to this problem. For example, $\varphi, \varphi \rightarrow \psi \vdash_{\mathbf{IKT}} \psi$ holds independently of K-L (and K-R), but, by $\rightarrow$-R, that implies that $\varphi \vdash_{\mathbf{IKT}} (\varphi \rightarrow \psi) \rightarrow \psi$ holds, which generates the same problem.

[23]By I, both $\varphi \vdash_{\mathbf{IKT}_\Rightarrow} \varphi$ and $\psi \vdash_{\mathbf{IKT}_\Rightarrow} \psi$ hold, and hence, by $\Rightarrow$-L, $\varphi, \varphi \Rightarrow \psi \vdash_{\mathbf{IKT}_\Rightarrow} \psi$ holds, and so, by $\rightarrow$-R, $\varphi \Rightarrow \psi \vdash_{\mathbf{IKT}_\Rightarrow} \varphi \rightarrow \psi$ holds (so that $\Rightarrow$ is at least as strong as $\rightarrow$). However, the converse proof fails: although, by I and $\rightarrow$-L, $\varphi, \varphi \rightarrow \psi \vdash_{\mathbf{IKT}_\Rightarrow} \psi$ holds, because of the presence of a non-logical$^\Rightarrow$ side premise ($\varphi \rightarrow \psi$) one cannot apply $\Rightarrow$-R to derive that $\varphi \rightarrow \psi \vdash_{\mathbf{IKT}_\Rightarrow} \varphi \Rightarrow \psi$ holds. More generally, with an eye at the definition of V in terms of $\Rightarrow$ and T forthcoming in the text, the idea behind the restriction that in $\Rightarrow$-R all the members of Γ and Δ be logical$^\Rightarrow$ is that valid arguments are only those in which the premise entails the conclusion *either all by itself or at most together with what are assumed to be the logical facts* (the idea is broadly reminiscent of the treatment of necessity in Prawitz [1965], pp. 74–80).

[24]'Define' and its likes can be used in so many different technical and non-technical senses, and the text is rather vague in this respect (if harmlessly so). In this connection, I should like to add that, while the rules and metarules given in section 3 and in this section define in the strictest mathematical sense '**IKT**' and '$\mathbf{IKT}_\Rightarrow$' respectively, and while V can be defined in terms of $\Rightarrow$ and T at least in

Indeed, something more general is available. For the purposes of setting up paradox (V$^-$), it sufficed to assume a *restricted* naive-validity predicate, namely one that, only applying to pairs of sentence names, only expresses validity for *single-premise and single-conclusion* arguments. However, restriction to single-premise and single-conclusion arguments is not necessary in order to be able to define naively the logical behaviour of a validity predicate in the way this has been done with V. For each $i, j \geq 1$, let V^{ij} express (i-premise and j-conclusion) $\vdash$-validity (with $\vdash$ including the logic of validity). Naively, the following principles would seem correct for V^{ij}:

$$\frac{\Gamma_0^1 \ \vdash \ \Delta_0^1, \varphi_1 \ldots \qquad \Gamma_0^i \vdash \Delta_0^i, \varphi_i \qquad \Gamma_1^1, \psi_1 \ \vdash \ \Delta_1^1 \ldots \qquad \Gamma_1^j, \psi_j \ \vdash \ \Delta_1^j}{\Gamma_0^1 \ldots, \Gamma_0^i, \Gamma_1^1 \ldots, \Gamma_1^j, V^{ij}(\ulcorner \varphi_1 \urcorner \ldots, \ulcorner \varphi_i \urcorner, \ulcorner \psi_1 \urcorner \ldots, \ulcorner \psi_j \urcorner) \ \vdash \ \Delta_0^1 \ldots, \Delta_0^i, \Delta_1^1 \ldots, \Delta_1^j} \; V^{ij}\text{-L}^-$$

$$\frac{\Gamma, \varphi_1 \ldots, \varphi_i \ \vdash \ \Delta, \psi_1 \ldots, \psi_j}{\Gamma \ \vdash \ \Delta, V^{ij}(\ulcorner \varphi_1 \urcorner \ldots, \ulcorner \varphi_i \urcorner, \ulcorner \psi_1 \urcorner \ldots, \ulcorner \psi_j \urcorner)} \; V^{ij}\text{-R}$$

with the restriction that in V^{ij}-R all the members of Γ and Δ be *logical*$^{V\#\#}$ (that is, be either a wff, for some k, l, of the form $V^{kl}(x, y)$ or be the negation, conjunction, disjunction, $\rightarrow$-implication or $\Rightarrow$-implication of logical$^{V\#\#}$ wffs). (Notice that our original V is now in effect V^{11}.) Again, simply by considering the relevant truth-preservation $\rightarrow$-conditionals, we know that the addition of any and every such predicate to **IKT** is consistent; and, in the provably consistent theory of truth $\mathbf{IKT}_{\Rightarrow}$, we can in fact define any such predicate $V^{ij}(x_1, x_2, x_3 \ldots, x_i, y_1, y_2, y_3 \ldots, y_j)$ as $T(x_1) \,\&\, T(x_2) \,\&\, T(x_3) \ldots \,\&\, T(x_i) \ \Rightarrow \ T(y_1) \vee T(y_2) \vee T(y_3) \ldots \vee T(y_j)$, in the "*intersubstitutability sense*" that it could be shown that, if we had primitive V^{ij} as well in the logic, $T(\ulcorner \varphi_1 \urcorner) \,\&\, T(\ulcorner \varphi_2 \urcorner) \,\&\, T(\ulcorner \varphi_3 \urcorner) \ldots \,\&\, T(\ulcorner \varphi_i \urcorner) \Rightarrow T(\ulcorner \psi_1 \urcorner) \vee T(\ulcorner \psi_2 \urcorner) \vee T(\ulcorner \psi_3 \urcorner) \ldots \vee T(\ulcorner \psi_j \urcorner)$ would be intersubstitutable with $V^{ij}(\ulcorner \varphi_1 \urcorner, \ulcorner \varphi_2 \urcorner, \ulcorner \varphi_3 \urcorner \ldots, \ulcorner \varphi_i \urcorner, \ulcorner \psi_1 \urcorner, \ulcorner \psi_2 \urcorner, \ulcorner \psi_3 \urcorner \ldots, \ulcorner \psi_j \urcorner)$ (if we also count every logical$^{V\#\#}$ wff as logical$^{\Rightarrow}$ and *vice versa*).

It might be worried that, while this treatment allows to express the validity of *multiple*-premise and *multiple*-conclusion arguments, it still does not allow to express the validity of *no*-premise or *no*-conclusion arguments. These arguments can however be represented as *single*-premise or *single*-conclusion if we add the logical-truth operator t and the logical-falsehood operator f on top of $\mathbf{IKT}_{\Rightarrow}$, thereby obtaining a new logic $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ closed under the additional operational rules and metarules (which include what, in $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$, is in effect a reformulation of $\varnothing/\mathsf{t}$):

the intersubstitutability sense, this is all compatible with thinking that the primitive metarules for an operator $\star$ or predicate Φ do *not* define in any *philosophically foundational* sense $\star$ or Φ. This is actually my view, as a consequence of which I think that we have an *antecedent conception* of the meanings of the various operators and predicates and that it is *on the basis of that conception* (inchoate and rough as it may be) that we recognise the primitive metarules given in section 3 and in this section as sound (see Zardini [2013j] for a defence of the view and fn 33 for an application).

$$\frac{\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} \Delta}{\mathsf{t} \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} \Delta}\ \mathsf{t}\text{-L} \qquad \frac{}{\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} \mathsf{t}}\ \mathsf{t}\text{-R}$$

$$\frac{}{\mathsf{f} \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} \varnothing}\ \mathsf{f}\text{-L} \qquad \frac{\Gamma \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} \varnothing}{\Gamma \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} \mathsf{f}}\ \mathsf{f}\text{-R}$$

(our definitions of logicality^V, logicality$^{\Rightarrow}$ and logicality$^{V\#\#}$ should of course be extended as to encompass t and f). The consistency of $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ follows from a simple extension of the proof given in Zardini [2011], pp. 524–532 as to encompass $\Rightarrow$-L, $\Rightarrow$-R, t-L, t-R, f-L and f-R. We can then represent the no-premise argument $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} \Delta$ with the single-premise argument $\mathsf{t} \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} \Delta$, and the no-conclusion argument $\Gamma \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} \varnothing$ with the single-conclusion argument $\Gamma \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} \mathsf{f}$.

5 From Naive Logical Properties to Non-Classical Metalanguages

Thus, the addition of naive-validity predicates preserves consistency in **IKT** and such predicates can actually be defined in terms of $\Rightarrow$ and T in $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ (in the intersubstitutability sense explained in section 4): $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ is not only a provably consistent theory of naive truth, but also a provably consistent theory of naive validity. However, I think that the success here is only apparent. There are three major, independent but converging reasons for being dissatisfied with $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ as a theory of naive logical properties. The three reasons concern, respectively, the proper characterisation of what it takes to be a naive-validity predicate, the full extent of the paradoxes of logical properties and the nature of logical consequence. They all lead to the conclusion that *the non-classical logic deployed by a non-classical solution should be expressed in a non-classical metalanguage*, thus connecting the two leading questions of this paper.

I'll approach the first reason for being dissatisfied with $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ as a theory of naive logical properties by developing an apparently correct line of thought leading to surprising conclusions and then diagnose where and why it shatters.[25] Consider any finitely axiomatised consistent (T- and $\Rightarrow$-free) classical theory (presented as a single sentence θ) sufficiently strong for Gödel's incompleteness theorems to apply to it (for example, Robinson's arithmetic). The existence of such theories is very much a matter of routine in $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ and its likes, as in these systems classical logic for a certain theory can be recovered in a very natural way with the addition of certain axioms (see Zardini [2011], p. 519): we'll thus extend $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ to a system $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$ such that, for every needed axiom φ, $\varnothing \vdash_{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}} \varphi$ holds. Also, for standard choices of θ, we'll assume that we have

[25] I'm here heavily indebted to conversations with Hartry Field back in 2006, when I first aired my ideas concerning the paradoxes of logical properties.

complemented $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$ with a minimal logic of quantification and identity (without going into details, I only observe that such complementation is straightforward and does not present the complications mentioned in section 3). Now, since, by I, both $\theta \vdash_{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}} \theta$ and $\mathsf{f} \vdash_{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}} \mathsf{f}$ hold, by V-L^{-}, $\theta, V(\ulcorner\theta\urcorner, \ulcorner\mathsf{f}\urcorner) \vdash_{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}} \mathsf{f}$ holds, and hence, by $\neg$-R, f-L and S, $\theta \vdash_{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}} \neg V(\ulcorner\theta\urcorner, \ulcorner\mathsf{f}\urcorner)$ holds. However, (single-premise, single-conclusion) $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$-validity has been defined in a *classical* metalanguage, and hence, given θ's stipulated strength in $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$, we can assume that θ has the resources to define a predicate $\Lambda^{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}}$ which does express $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$-validity. *If* V also really expressed $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$-validity, we would expect that, relative to θ, it is intersubstitutable with $\Lambda^{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}}$, and hence we would expect $\theta \vdash_{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}} \neg V(\ulcorner\theta\urcorner, \ulcorner\mathsf{f}\urcorner)$ holding to imply $\theta \vdash_{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}} \neg\Lambda^{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}}(\ulcorner\theta\urcorner, \ulcorner\mathsf{f}\urcorner)$ holding. But then, given θ's strength in $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$, a version of Gödel's second incompleteness theorem applies, so that $\theta \vdash_{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}} \neg\Lambda^{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}}(\ulcorner\theta\urcorner, \ulcorner\mathsf{f}\urcorner)$ holding implies the inconsistency of θ in $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$. Thus, we would have the surprising conclusion that certain classically consistent (T- and $\Rightarrow$-free) theories are inconsistent in $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$, and, more generally, that some (T- and $\Rightarrow$-free) $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$-valid arguments are not classically valid: in some respects, $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$ would be *stronger* than classical logic (although, of course, in other respects—such as the lack of unrestricted contraction—it would be weaker).

However, as presented, this line of thought shatters because, even relative to θ, there is no antecedent reason to think that V is intersubstitutable with $\Lambda^{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}}$ in $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$ (and indeed, on many choices of θ, it could be shown that it isn't), contrary to what the step from $\theta \vdash_{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}} \neg V(\ulcorner\theta\urcorner, \ulcorner\mathsf{f}\urcorner)$ holding to $\theta \vdash_{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}} \neg\Lambda^{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}}(\ulcorner\theta\urcorner, \ulcorner\mathsf{f}\urcorner)$ holding presupposes. This is so for the simple reason that, otherwise, relative to θ, V would contract in $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$ (since, relative to θ, $\Lambda^{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}}$ behaves classically in $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$ and hence, relative to θ, $\Lambda^{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}}$ contracts in $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$). And if, relative to θ, V contracted, a version of paradox (V^{-}) would suffice to show that θ is inconsistent in $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$, which there is no antecedent reason to think it is (and indeed, on many choices of θ, it could be shown that it isn't). But, since by definition $\Lambda^{\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}}$ does express $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$-validity (for it is just a formal rewrite in θ of the informal definition that I gave in sections 3 and 4 in a classical metalanguage), there is no reasonable sense in which V could be said to do so as well, *in spite of* the fact that it obeys V-L^{-} and V-R.

Something even stronger can be said. For, in fact, it is *precisely because of* the fact that V obeys V-L^{-} and V-R—that is, that it obeys metarules that no predicate expressing a classical relation can obey on pain of triviality—that there is no reasonable sense in which it can be said to express the classical relation of $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$-validity. Since the idea that V does express $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$-validity is the only rationale behind the postulation that it obey V-L^{-} and V-R, that postulation itself is revealed to be *incoherent*, independently of the question of what logic $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$ consists in (and hence independently of the fact that V-L^{-} and V-R can consistently be added to $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$). And, obviously, these points hold not only for $\mathbf{IKT}^{\theta}_{\Rightarrow \mathsf{tf}}$-validity, but for any relation of validity which is defined in a classical metalanguage.[26]

[26]There is no analogous incoherence in postulating that an ordinary conditional like $\rightarrow$ obey both of the extremely plausible $\rightarrow$-L and $\rightarrow$-R, but Curry's paradox teaches us that non-substructural non-classical

The general conclusion to draw from all this is that obeying (the relevant analogues of) V-L$^-$ and V-R, while necessary, is not sufficient for expressing *naive validity* (since it is not sufficient to express *validity* in the first place) and hence that our conception of what it takes to be a naive-validity predicate should be strengthened as follows:

(NVP^{ij}) Given a logic $\vdash$, Φ^{ij} is a naive-validity predicate for $\vdash$ iff:

(i) Φ^{ij} obeys the relevant analogues of V^{ij}-L$^-$ and V^{ij}-R;[27]

(ii) If Ψ^{ij} is an object-language validity predicate for $\vdash$, Φ^{ij} is intersubstitutable with Ψ^{ij} in the relevant contexts.[28]

Correspondingly, our conception of what it takes for a predicate to express naive inconsistency should be strengthened as follows:

(NIP) Given a logic $\vdash$, Φ is a naive-inconsistency predicate for $\vdash$ iff:

(i) Φ obeys the relevant analogues of I-L and I-R;[29]

(ii) If Ψ is an object-language inconsistency predicate for $\vdash$, Φ is intersubstitutable with Ψ in the relevant contexts

(as mentioned in fn 10, in this paper I don't aim at giving a corresponding adequately strong theory of naive compatibility).

---

solutions cannot validate both of those metarules. This is one important respect in which, contrary to what seems to be an emerging trend in the most recent discussions on naive logical properties, I think that, for non-substructural non-classical solutions, Curry's paradox is more problematic than paradox (V$^-$).

[27] As I proceed to argue in my treatment in this section of the second reason for being dissatisfied with $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ as a theory of naive logical properties, my own view is actually that, taking for purely illustrative purposes the single-premise, single-conclusion case, obeying V-L$^-$ is *still too weak* a necessary condition for being a naive-validity predicate, and that obeying a certain other metarule (labelled 'V-L' below in the text) should be added. As this last twist is somehow tangential to the main dialectic of this paper, it'll be easier however to stick to this official definition of a naive-validity predicate, while making sure (as I'll do) that the theory presented in section 6 is such that its validity predicate also obeys V-L (and not just V-L$^-$).

[28] Notice that the qualification 'in the relevant contexts' is arguably needed if the language over which $\vdash$ is defined contains for example operators of propositional attitudes. Notice also that the intersubstitutability in question (as well as, more generally, the relevant entailments concerning Φ^{ij} and Ψ^{ij}) should of course be understood as being relative to the theories that contain Φ^{ij} and Ψ^{ij}. Notice finally that condition (ii) is strong enough as to rule out $T(x) \to T(y)$ as a naive-validity predicate (and correctly so, see section 4).

[29] As I argue in Zardini [2013e], my own view is actually that obeying I-L is *still too weak* a necessary condition for being a naive-inconsistency predicate, and that obeying the converse metarule should be added. As this last twist is completely tangential to the main dialectic of this paper, it'll be easier however to stick to this official definition of a naive-inconsistency predicate, while making sure (as I'll do) that the theory presented in section 6 is such that its inconsistency predicate also obeys the converse of I-L (and not just I-L).

Now, focussing again on the case in which $\vdash$ is expressed in a classical metalanguage (as $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ is), a predicate expressing validity in $\vdash$ (as $\Lambda^{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}}$ does) will behave classically, and hence, on pain of paradox (V^-), will not obey the relevant analogues of both V-L$^-$ and V-R. (For what is worth, a suitable theory θ containing some such predicate Λ will typically validate at least some suitably modified version of the relevant analogue of V-R—it will be true that, if $\varphi \vdash \psi$ holds, $\theta \vdash \Lambda(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ holds—while it will fail to validate the relevant analogue of V-L$^-$—it will not be true that, if both $\Gamma_0 \vdash \Delta_0, \varphi$ and $\Gamma_1, \psi \vdash \Delta_1$ hold, $\theta, \Gamma_0, \Gamma_1, \Lambda(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \vdash \Delta_0, \Delta_1$ holds; see Zardini [2013e] for more details about this.) But hence no predicate obeying the relevant analogues of those metarules will be intersubstitutable with it—that is, no predicate satisfying condition (i) of (NVP^{ij}) will also satisfy condition (ii) of (NVP^{ij}), and thus no predicate will be a *naive-validity* predicate (although some predicates, like $\Lambda^{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}}$, might be *validity* predicates). Therefore, *if* $\vdash$ *is expressed in a classical metalanguage*, a predicate expressing naive validity cannot be added to it (lest $\varnothing \vdash \varphi$ hold for every φ). As this runs counter my working hypothesis that a good non-classical solution accommodating for naive semantic properties should also be able to accommodate for naive logical properties, I conclude that the non-classical logic deployed by a non-classical solution should be expressed in a non-classical metalanguage.

The same conclusion can be reached by considering the second reason for being dissatisfied with $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ as a theory of naive logical properties. Independently of the fact that, as we've just seen, the $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$-definition of V in terms of T and $\Rightarrow$ does not really express $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$-validity, I and C, which seem equally central to a theory of naive logical properties, cannot be added to it. This is the time to refocus our attention from paradox (V^-) to paradoxes (I) and (C). Very interestingly, the reasoning of the latter paradoxes uses almost no principles *in the object language*—in addition to I-L, I-R, C-L and C-R, only transitivity, monotonicity and $\varnothing/\mathsf{t}$ (and the last one only in the case of paradox (C)). In particular, such reasoning *does not use contraction* and hence it would seem that no non-contractive logic of truth preserving transitivity and monotonicity can accommodate for the naive logical properties of inconsistency and compatibility (in fact, I show in Zardini [2013e] that even transitivity and monotonicity are dispensable for certain paradoxes of logical properties).

However, the reasoning of paradoxes (I) and (C) does use distinctively classical principles *in the metalanguage* (indeed, principles that non-classical solutions typically reject for the object language, such as e.g. *reductio*).[30] This is still enough to undermine the possibilities for just about every extant non-classical solution (*including* $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$) to accommodate for the naive logical properties of inconsistency and compatibility (as just about every extant non-classical solution presupposes a classical metalanguage), but it does let in the little ray of hope constituted by the idea of shifting to a non-classical metalanguage: *via* another route, we have thus reached again the conclusion that the non-classical logic

[30]This is interestingly not so for the reasoning of paradox (V^-), which in the metalanguage only uses universal instantiation and *modus ponens* (in applying a principle to a particular case), intersubstitutability of identicals (in the use of the definition of '$\digamma^-$') and adjunction (in putting together the two pieces of the reasoning), all principles that are usually accepted by non-classical solutions for the object language as well as for the metalanguage.

deployed by a non-classical solution should be expressed in a non-classical metalanguage.

Let me hasten to add, however, that the space of manoeuvre here is *much less* than what may seem from the official reasoning of paradoxes (I) and (C). While I've presented those paradoxes in the smoothest and most straightforward way as using *very classical* principles, they can more cumbersomely be presented as using principles that most non-classical solutions are bound to accept *even if the metalanguage is supposed to be governed by their favoured non-classical logic*. For example, focussing on paradox (I), it is hard to see how anyone accepting I-R and the existence of ι as defined could not accept the conditionals:

(1) If $\iota \vdash \varnothing$ holds, then, by I-R, $\varnothing \vdash I(\ulcorner \iota \urcorner)$ holds;

(2) If $\varnothing \vdash I(\ulcorner \iota \urcorner)$ holds, then, by definition of 'ι', $\varnothing \vdash \iota$ holds,

which, on most non-classical solutions, entail the conditional:

(3) If $\iota \vdash \varnothing$ holds, then [$\varnothing \vdash \iota$ holds and $\iota \vdash \varnothing$ holds],

with the inference in question being that from 'If φ, then ψ' and 'If ψ, then χ' to 'If φ, then χ and φ' (valid on most non-classical solutions). But it is also hard to see how anyone accepting transitivity could not accept the conditional:

(4) If $\varnothing \vdash \iota$ holds and $\iota \vdash \varnothing$ holds, then, by transitivity, $\varnothing \vdash \varnothing$ holds,

which, together with (3), on most non-classical solutions entails the conditional:

(5) If $\iota \vdash \varnothing$ holds, then $\varnothing \vdash \varnothing$ holds,

with the inference in question being that from 'If φ, then ψ' and 'If ψ, then χ' to 'If φ, then χ' (valid on most non-classical solutions). But, on most non-classical solutions, at least some conditional validating the previous claims also contraposes, and hence we get:

(6) If $\varnothing \vdash \varnothing$ does not hold, then $\iota \vdash \varnothing$ does not hold.

Since virtually all non-classical solutions accept the antecedent of (6) and *modus ponens*, they are stuck with $\iota \vdash \varnothing$ not holding, which the first leg of the reasoning of paradox (I) shows to lead to contradiction by steps that everyone accepting I-L and the existence of ι as defined must accept (and, for good measure, one half of the contradiction—that $\iota \vdash \varnothing$ holds—is shown by the second leg of the reasoning of paradox (I) to lead to triviality by steps that everyone accepting I-R, the existence of ι as defined, transitivity and monotonicity must accept, so it won't do just to swallow the pill and accept the contradiction that $\iota \vdash \varnothing$ both holds and does not hold).

I will pursue to some length the little ray of hope constituted by the idea of shifting to a non-classical metalanguage in section 6. In the meanwhile, I should like to emphasise that I don't think that the difference just discussed between paradox (V^-) and paradoxes (I) and (C) reflects *any* interesting difference between naive validity and naive inconsistency and compatibility—it only reflects the objectionable *weakness* of the logic of naive validity codified by V-L^- and V-R. In particular, the weakness resides in V-L^-, which should be obvious given the fact that an analogue of V-L^-, $\rightarrow$-L, is satisfied by an operator, $\rightarrow$, that, as we've seen in section 4, can lay no claim to help to express validity. V-L^- should thus be strengthened, and the obvious strengthening suggested by V-R is:[31]

$$\frac{\varphi \nvdash \psi}{V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \vdash \varnothing}\ V\text{-L}$$

It may not be completely obvious why V-L is a strengthening of V-L^-, so let me explain it. Suppose that $\Gamma_0 \vdash \Delta_0, \varphi$ and $\Gamma_1, \psi \vdash \Delta_1$ (the premises of V-L^-) hold. Now, either $\varphi \vdash \psi$ holds or it does not. Suppose that $\varphi \vdash \psi$ holds. Then, by two instances of S, $\Gamma_0, \Gamma_1 \vdash \Delta_0, \Delta_1$ holds, and hence, by K-L, $\Gamma_0, \Gamma_1, V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \vdash \Delta_0, \Delta_1$ (the conclusion of V-L^-) holds. Suppose that $\varphi \nvdash \psi$ holds. Then, by V-L, $V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \vdash \varnothing$ holds, and hence, by K-L and K-R, $\Gamma_0, \Gamma_1, V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \vdash \Delta_0, \Delta_1$ holds. Therefore, reasoning by cases, $\Gamma_0, \Gamma_1, V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \vdash \Delta_0, \Delta_1$ holds, and hence it would seem that the validity of V-L implies that of V-L^-.

On reflection, this argument is actually rather puzzling. Taking the particular case in which the conclusion of V-L^- is $\varphi, V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \vdash \psi$ (the analogue for V of the rule of *modus ponens*, derivable by V-L^- from $\varphi \vdash \varphi$ and $\psi \vdash \psi$, in turn derivable by I), it seems to show that such a rule is in a natural sense *redundant*, since for each of its instances there is a stronger valid argument (either $\varphi \vdash \psi$ or $V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \vdash \varnothing$, depending on whether $\varphi \vdash \psi$ holds or not) from which it can be derived by monotonicity. Moreover, an analogous situation would hold for any predicate or operator strong enough as to obey the relevant analogue of V-L (as we would expect $\Rightarrow$ to be). All this would appear to be a grave offence to the absolutely crucial role traditionally assigned to *modus ponens*. I'll briefly comment on how the theory offered in section 6 avoids committing the offence (while validating V-L).

[31] In the sense that, subject to a qualification, V-L is the straightforward converse of V-R: just as V-R amounts to $\vdash$ "*proving*" $V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ whenever $\varphi \vdash \psi$ holds (and hence, given that $V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ is supposed to say just that, whenever $V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ is true), V-L amounts to $\vdash$ "*refuting*" $V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ whenever $\varphi \nvdash \psi$ holds (and hence, given that $V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ is supposed to say just the contradictory of that, whenever $V(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ is not true). The qualification is that V-R is actually stronger than just stated, as it also encompasses cases *with logical^V side premises and conclusions*. An analogous version of V-L with logical^V side premises and conclusions is just as viable. However, because of the *negative* character of V-L's premise, given monotonicity the version of V-L with logical^V side premises and conclusions is in fact implied by the version of V-L without logical^V side premises and conclusions (whereas the analogous implication does not hold for V-R because of the *positive* character of its premise). Given that monotonicity remains unquestioned in this paper (with the exception of fn 38), for simplicity I'll stick to the version of V-L without logical^V side premises and conclusions.

What is now important to observe is that, unfortunately, intuitive as they are at first glance, V-L and V-R would seem to lead to catastrophe *exactly under the same assumptions required by paradox* (I). We assume that the language contains a sentence $\mathcal{F}$ identical to $V(\ulcorner\mathcal{F}\urcorner, \ulcorner\varphi\urcorner)$, where φ is arbitrary. It would then seem that we could reason as follows. Suppose that $\mathcal{F} \nvdash \varphi$ holds. Then, by V-L, $V(\ulcorner\mathcal{F}\urcorner, \ulcorner\varphi\urcorner) \vdash \varnothing$ holds, and hence, by monotonicity, $V(\ulcorner\mathcal{F}\urcorner, \ulcorner\varphi\urcorner) \vdash \varphi$ holds. But, by definition of '$\mathcal{F}$', that is tantamount to $\mathcal{F} \vdash \varphi$ holding, and hence, by *reductio*, $\mathcal{F} \vdash \varphi$ holds. Therefore, by V-R, $\varnothing \vdash V(\ulcorner\mathcal{F}\urcorner, \ulcorner\varphi\urcorner)$ holds. But, by definition of '$\mathcal{F}$', $\mathcal{F} \vdash \varphi$ holding is tantamount to $V(\ulcorner\mathcal{F}\urcorner, \ulcorner\varphi\urcorner) \vdash \varphi$ holding. Thus, both $\varnothing \vdash V(\ulcorner\mathcal{F}\urcorner, \ulcorner\varphi\urcorner)$ and $V(\ulcorner\mathcal{F}\urcorner, \ulcorner\varphi\urcorner) \vdash \varphi$ hold, and hence, by transitivity, $\varnothing \vdash \varphi$ holds, and so, by monotonicity, $\vdash$ is trivial. (Let's label this paradox '(V)'.)

Again, very interestingly, the reasoning of paradox (V) uses almost no principles *in the object language*—in addition to V-L and V-R, only transitivity and monotonicity. In particular, such reasoning *does not use contraction* and hence it would seem that no non-contractive logic of truth preserving transitivity and monotonicity can accommodate not only for the naive logical properties of inconsistency and compatibility, but also, after all, for the naive logical property of validity. (Again, I hasten to add that the more sophisticated reasoning, mentioned in the case of paradoxes (I) and (C), doing away with distinctively classical principles could be deployed here as well to strengthen even more also the force of paradox (V).)

Deeper consideration of paradox (V^-) and consideration of paradoxes (I), (C) and (V) thus converge on the conclusion that *only a non-classical logic expressed in a non-classical metalanguage can accommodate for the naive logical properties of validity, inconsistency and compatibility*. Some defenders of non-classical solutions might take this finding to offer a conclusive argument for regarding naive logical properties beyond redemption, in contrast with naive truth whose redemption can be ensured by a non-classical logic expressed in a classical metalanguage. My working hypothesis that a good non-classical solution accommodating for naive semantic properties should also be able to accommodate for naive logical properties pulls me however towards the opposite direction, looking for a non-classical logic expressed in a non-classical metalanguage.

This stance is reinforced by the third reason, available at least from the perspective of non-classical solutions to the semantic paradoxes, for being dissatisfied with $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ as a theory of naive logical properties, which is really a reason for being dissatisfied with $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ (just as with any other logic expressed in a classical metalanguage) *as a logic in the first place*. Such reason is disarmingly simple and relies on the traditional doctrine, already partially emerged in section 3, according to which *logical consequence just is a certain form of truth preservation*, so that, for example, the validity of the argument from φ to ψ just is the fact that, if φ is true, so is ψ (given a suitably strong understanding of the conditional).[32] If that doctrine is on the right track, and if a non-classical solution is too, then, given that implications involving truth are rife with semantic paradoxes (as shown for example by Curry's paradox), it just follows that logical consequence is a non-

[32]The doctrine would require a lot of qualifications that have nothing to do with the semantic paradoxes, so I'll leave them aside (see Zardini [2012] for discussion of some the relevant issues).

classical relation, and hence that it should be expressed in a non-classical metalanguage. It's as simple as that.

Deeper consideration of paradox (V^-), consideration of paradoxes (I), (C) and (V) and consideration of the nature of logical consequence thus converge on the conclusion that *the non-classical logic deployed by a non-classical solution should be expressed in a non-classical metalanguage*. In fact, although $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ in itself does not constitute such a logic, we'll see that it still can be very usefully employed on our way to it. And doing so will allow us to see a very simple, natural and systematic way in which a non-classical logic can be expressed in a non-classical metalanguage, thereby assuaging at least some of the *horror vacui* that pulls one to consider naive logical properties beyond redemption.

6 A Non-Contractive Semantic Theory of Some Naive Logical Properties

Contrary to a fashionable contemporary trend I've already alluded to in section 3, and developing further one of the final themes of section 5, I take it to be as plausible as ever that *logical consequence just is a certain form of truth preservation*. And, given this, I take it to be plausible that the conditional expressing this certain form of truth preservation is something along the lines of a certain understanding of $\Rightarrow$.[33] If this is correct, the right reaction in the presence of the problems emerged in section 5 for one wishing to accommodate for naive logical properties is to shift to a non-classical metalanguage (so much is forced by those problems) which contains, among other things, both something along the lines of a truth predicate and something along the lines of $\Rightarrow$ (so much is forced by the conception of validity just sketched).

Before making a specific proposal to implement this plan, let me enter a crucial clarification. While my proposal in this paper is that, when theorising about validity, we should talk in a certain non-classical metalanguage and adopt a certain theory of validity in that language, the language that in turn I'll use in defining that language and theory, and in exploring some of their properties, remains classical. Thus, I'll in effect be studying that language and theory using a *classical metametalanguage*. I'll climb up the ladder.

My specific proposal in this paper is then, naturally enough, to implement the plan sketched in the second last paragraph by building on what has been achieved with $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$. In particular, I propose that we shift to a non-classical metalanguage $\mathscr{N}$ containing, beyond other innocent things, the operators and the truth predicate of $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$. For simplicity, we also assume that the object language of $\mathscr{N}$ is $\mathscr{N}$ itself. *In* $\mathscr{N}$, we can express a new logic $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$ by saying that an argument with premises $[\varphi_0, \varphi_1, \varphi_2 \ldots, \varphi_i]$ and con-

[33]To give at least a hint of my broader picture here, I should add that, as a consequence of the general view mentioned in fn 33, I don't think that our understanding of $\Rightarrow$ is given by $\Rightarrow$-L and $\Rightarrow$-R. I rather think that the relevant understanding is given by our conception of something *being in itself a sufficient condition* for something and I think that it is *on the basis of that conception* (inchoate and rough as it may be) that we recognise $\Rightarrow$-L and $\Rightarrow$-R as sound.

clusions $[\psi_0, \psi_1, \psi_2 \ldots, \psi_j]$ is $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$-*valid* (equivalently, that $\varphi_0, \varphi_1, \varphi_2 \ldots, \varphi_i \vdash_{\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}} \psi_0, \psi_1, \psi_2 \ldots, \psi_j$ holds) iff $T(\ulcorner\varphi_0\urcorner) \,\&\, T(\ulcorner\varphi_1\urcorner) \,\&\, T(\ulcorner\varphi_2\urcorner) \ldots \,\&\, T(\ulcorner\varphi_i\urcorner) \;\Rightarrow\; T(\ulcorner\psi_0\urcorner) \vee T(\ulcorner\psi_1\urcorner) \vee T(\ulcorner\psi_2\urcorner) \ldots \vee T(\ulcorner\psi_j\urcorner)$ is true (we maintain the method explained in section 4 for representing no-premise arguments and no-conclusion arguments).[34] We can then identify validity with $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$-validity. This beautifully simple and intuitive identification is thus the first part of our new theory of validity in $\mathscr{N}$ (theory which I'll call '$\mathcal{N}$'): validity is nothing fancy (as it would be if it were identified with certain sophisticated model-theoretic or proof-theoretic relations) or *sui generis* (as it would be if it were taken to be non-reducible)—it simply consists in the truth of certain suitably strong conditionals whose antecedents say that certain premises are true and whose consequents say that certain conclusions are true. Validity is preservation of truth.

In a good sense, then, according to $\mathcal{N}$ the *property* of being valid is the property of preserving truth in a suitably strong way. But, moving on to the *extension* of this property, which specific truth-preservation $\Rightarrow$-conditionals are true? Or, more precisely, which specific truth-preservation $\Rightarrow$-conditionals should be included in $\mathcal{N}$?[35] This is exactly the point at which the logic of truth $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$, which is defined in a classical metalanguage, still comes in handy. We can take it as a *sufficient* condition for a truth-preservation $\Rightarrow$-conditional to be true that it be a logical truth in $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$,[36] and we can take it as a *sufficient* condition for one such conditional to be false that it be a logical falsehood in $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$. Thus, for a start, we can in effect take $\mathcal{N}$ to contain the claim [that $\varphi_0, \varphi_1, \varphi_2 \ldots, \varphi_i \vdash_{\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}} \psi_0, \psi_1, \psi_2 \ldots, \psi_j$ holds] *if* $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} T(\ulcorner\varphi_0\urcorner) \,\&\, T(\ulcorner\varphi_1\urcorner) \,\&\, T(\ulcorner\varphi_2\urcorner) \ldots \,\&\, T(\ulcorner\varphi_i\urcorner) \;\Rightarrow\; T(\ulcorner\psi_0\urcorner) \vee T(\ulcorner\psi_1\urcorner) \vee T(\ulcorner\psi_2\urcorner) \ldots \vee T(\ulcorner\psi_j\urcorner)$ holds (which, given transparency, amounts to $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \varphi_0 \,\&\, \varphi_1 \,\&\, \varphi_2 \ldots \,\&\, \varphi_i \;\Rightarrow\; \psi_0 \vee \psi_1 \vee \psi_2 \ldots \vee \psi_j$ holding), and we can take $\mathcal{N}$ to contain the claim [that $\varphi_0, \varphi_1, \varphi_2 \ldots, \varphi_i \vdash_{\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}} \psi_0, \psi_1, \psi_2 \ldots, \psi_j$ does not hold] *if* $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \neg(T(\ulcorner\varphi_0\urcorner) \,\&\, T(\ulcorner\varphi_1\urcorner) \,\&\, T(\ulcorner\varphi_2\urcorner) \ldots \,\&\, T(\ulcorner\varphi_i\urcorner) \;\Rightarrow\; T(\ulcorner\psi_0\urcorner) \vee T(\ulcorner\psi_1\urcorner) \vee T(\ulcorner\psi_2\urcorner) \ldots \vee T(\ulcorner\psi_j\urcorner))$ holds (which, given transparency, amounts to $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \neg(\varphi_0 \,\&\, \varphi_1 \,\&\, \varphi_2 \ldots \,\&\, \varphi_i \Rightarrow \psi_0 \vee \psi_1 \vee \psi_2 \ldots \vee \psi_j)$ holding).

However, the idea behind taking it as a sufficient condition for a conditional of the form $T(x) \Rightarrow T(y)$ to be true that it be a logical truth in $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ is the much more

[34] More precisely, given that $\mathscr{N}$ is going to be a quantifier-free language just as the language of $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ is, whenever I informally speak of $\mathcal{N}$ containing certain *general* claims what I really mean is $\mathcal{N}$ containing all the *instances* of those claims (see also fn 21). Thanks to an anonymous referee for urging this clarification.

[35] $\mathcal{N}$ is a theory that, on the one hand, contains the identification of validity with $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$-validity and, on the other hand, contains claims about what is $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$-valid (and what is not so). I'll show below that, in this second respect, $\mathcal{N}$ is *essentially incomplete*, in the sense that there is a truth-preservation $\Rightarrow$-conditional (which, as such, is a claim about what is $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$-valid) such that, under plausible assumptions, neither it nor its negation can consistently be added to $\mathcal{N}$. Assuming that either that truth-preservation $\Rightarrow$-conditional or its negation is true, some true claim about what is $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$-valid (or what is not so) cannot possibly be added to $\mathcal{N}$. Hence the distinction between the two questions in the text.

[36] Of course, we shouldn't also take that as a *necessary* condition, given that the way it has been defined in sections 3 and 4 the predicate 'is a logical truth in $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$' behaves classically. For doing so will in effect make 'is valid' behave classically, while the whole point of our enterprise in this section is to give a non-classical theory of validity.

general idea that $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ is *sound at least "lawwise"* (in the sense that every logical truth of $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ is in fact a logical truth, and hence true). It thus makes perfect sense to strengthen $\mathcal{N}$ by also including in it *all* logical truths of $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$, whether or not they are truth-preservation $\Rightarrow$-conditionals (or negations of truth-preservation $\Rightarrow$-conditionals), and even whether or not they have $\Rightarrow$ as main operator (or $\neg$ as main operator governing $\Rightarrow$). Indeed, given the properties of $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$, already doing so merely for the truth-preservation $\Rightarrow$-conditionals commits us also to the idea that $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ is *sound at least "rulewise"* (in the sense that every $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$-valid argument is in fact valid, and hence preserves truth), since, by &-L, $\vee$-R, $\Rightarrow$-R and transparency, $\varphi_0, \varphi_1, \varphi_2 \ldots, \varphi_i \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} \psi_0, \psi_1, \psi_2 \ldots, \psi_j$ holds only if $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} T(\ulcorner \varphi_0 \urcorner) \,\&\, T(\ulcorner \varphi_1 \urcorner) \,\&\, T(\ulcorner \varphi_2 \urcorner) \ldots \,\&\, T(\ulcorner \varphi_i \urcorner) \Rightarrow T(\ulcorner \psi_0 \urcorner) \vee T(\ulcorner \psi_1 \urcorner) \vee T(\ulcorner \psi_2 \urcorner) \ldots \vee T(\ulcorner \psi_j \urcorner)$ holds.

Once we have adopted $\mathcal{N}$ as our theory of validity, adding a naive-validity predicate to the object language of $\mathcal{N}$ is easy. Indeed, just as we thought it was the case for $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$, there is no need to *add it* since *it is already there*. In fact, given the identity of metalanguage and object language ($\mathcal{N}$), it is already there, as it were, in its full splendour, in the sense that it is already there as exactly the same family of predicates that we've just used to define what validity is according to $\mathcal{N}$, i.e. the family of predicates of the form $T(x_1) \,\&\, T(x_2) \,\&\, T(x_3) \ldots \,\&\, T(x_i) \Rightarrow T(y_1) \vee T(y_2) \vee T(y_3) \ldots \vee T(y_j)$ (which for brevity I'll henceforth also write as $V^{ij}(x_1, x_2, x_3 \ldots, x_i, y_1, y_2, y_3 \ldots y_j)$: from now on, the latter predicates should always be understood as defined by the former).

We can show that, according to $\mathcal{N}$, V^{ij} as so defined obeys V^{ij}-L$^-$ and V^{ij}-R even when $\vdash$ is $\vdash_{\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}}$. On the one hand, V^{ij}-L$^-$ basically turns on $\Rightarrow$-L and in particular on the fact that, given some routine fiddling:

$$\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} (\bigwedge(\Gamma^1_0) \Rightarrow \bigvee(\Delta^1_0, \varphi_1)) \ldots \,\&\, (\bigwedge(\Gamma^i_0) \Rightarrow \bigvee(\Delta^i_0, \varphi_i)) \,\&\, (\bigwedge(\Gamma^1_1, \psi_1) \Rightarrow \bigvee(\Delta^1_1)) \ldots \,\&\, (\bigwedge(\Gamma^j_1, \psi_j) \Rightarrow \bigvee(\Delta^j_1)) \Rightarrow \bigwedge(\Gamma^1_0 \ldots, \Gamma^i_0, \Gamma^1_1 \ldots, \Gamma^j_1, \varphi_1 \ldots \,\&\, \varphi_i \Rightarrow \psi_1 \ldots \vee \psi_j) \Rightarrow \bigvee(\Delta^1_0 \ldots, \Delta^i_0, \Delta^1_1 \ldots, \Delta^j_1)$$

holds (where $\bigwedge(\Gamma)$ is the conjunction of all the occurrences of Γ and $\bigvee(\Delta)$ the disjunction of all the occurrences of Δ), which, by transparency, means that, according to $\mathcal{N}$, the premises of V^{ij}-L$^-$ entail in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$ its conclusion. In order to get a feel for the meaning of the long wff displayed on the right-hand side of '$\vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}}$', consider that it is, roughly and *modulo* transparency, the result of "*flattening*" the presentation of V^{ij}-L$^-$ in a classical metalanguage ("*If* all of $\Gamma^1_0 \vdash \Delta^1_0, \varphi_1 \ldots$ and $\Gamma^i_0 \vdash \Delta^i_0, \varphi_i$ and $\Gamma^1_1, \psi_1 \vdash \Delta^1_1 \ldots$ and $\Gamma^j_1, \psi_j \vdash \Delta^j_1$ hold, *then* $\Gamma^1_0 \ldots, \Gamma^i_0, \Gamma^1_1 \ldots, \Gamma^j_1, V^{ij}(\ulcorner \varphi_1 \urcorner \ldots, \ulcorner \varphi_i \urcorner, \ulcorner \psi_1 \urcorner \ldots, \ulcorner \psi_j \urcorner) \vdash \Delta^1_0 \ldots, \Delta^i_0, \Delta^1_1 \ldots, \Delta^j_1$ holds") by substituting the relevant truth-preservation $\Rightarrow$-conditionals for *both* the *metalinguistic conditional* 'If…, then…' and the *metalinguistic validity predicate* $\vdash$ (while, to repeat, the *object-language validity predicate* V^{ij} is already understood as defined by the relevant truth-preservation $\Rightarrow$-conditional). The fact that the result of such a three-level flattening is a logical truth in $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ is very significant: it means that, according to $\mathcal{N}$, the premises of V^{ij}-L$^-$ not only *imply* its conclusion (which, on the traditional way of talking about logics, is strangely enough all is required to vindicate the *validity* of a certain metarule), but indeed *entail* it in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$ (recall from fn 2 the contrast between implication and entailment).[37] This circumstance only befits what is, according to $\mathcal{N}$, the logically closed language $\mathscr{N}$ of $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$: according to $\mathcal{N}$, among the sentences of $\mathscr{N}$ that can occur as premises or conclusions in valid (i.e. $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$-valid) arguments are sentences about valid (i.e. $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$-valid) arguments, and the logic is sensitive enough to the features of the object-language (and metalanguage) notion of validity (i.e. truth-preservation $\Rightarrow$-conditionals) as to be apt to generate the relevant entailments among those sentences.

On the other hand, V^{ij}-R basically turns on $\Rightarrow$-R and in particular on the fact that, given some routine fiddling:

$$\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} (\bigwedge(\Gamma, \varphi_1, \varphi_2, \varphi_3 \ldots, \varphi_i) \Rightarrow \bigvee(\Delta, \psi_1, \psi_2, \psi_3 \ldots, \psi_j)) \Rightarrow \bigwedge(\Gamma) \Rightarrow \bigvee(\Delta, \varphi_1 \mathbin{\&} \varphi_2 \mathbin{\&} \varphi_3 \ldots \mathbin{\&} \varphi_i \Rightarrow \psi_1 \vee \psi_2 \vee \psi_3 \ldots \vee \psi_j)$$

holds (with the restriction that all the members of Γ and Δ be logical$^{\Rightarrow}$), which, by transparency, means that, according to $\mathcal{N}$, the premise of V^{ij}-R entails in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ its conclusion. Thus, according to $\mathcal{N}$, V^{ij} meets condition (i) of (NVP^{ij}) for being a naive-validity predicate for $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$. Moreover, according to $\mathcal{N}$, V^{ij} is trivially fully intersubstitutable with an object-language validity predicate for $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ (namely, itself!). Thus, according to $\mathcal{N}$, V^{ij} also meets condition (ii) of (NVP^{ij}) for being a naive-validity predicate for $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$, and hence, according to $\mathcal{N}$, $V^{ij}(x_1, x_2, x_3 \ldots, x_i, y_1, y_2, y_3 \ldots, y_j)$ (i.e. $T(x_1) \mathbin{\&} T(x_2) \mathbin{\&} T(x_3) \ldots \mathbin{\&} T(x_i) \Rightarrow T(y_1) \vee T(y_2) \vee T(y_3) \ldots \vee T(y_j)$) is a naive-validity predicate for $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$.

Thus, $\mathcal{N}$ is not only a theory of validity consisting in a set of *categorical* claims about what is valid (and what is not so)—given the presence of an object-language validity predicate, $\mathcal{N}$ is also *ipso facto* a far more general theory of validity containing non-categorical claims about validity. In particular, as the discussion in the last paragraph of the *metarules* V^{ij}-L$^-$ and V^{ij}-R makes vivid, the theory contains *hypothetical* claims about validity (both hypothetical ordinary $\rightarrow$-conditional claims and hypothetical $\Rightarrow$-conditional claims).[38]

[37]This property is of course lacked by a classical language: although the premises of a classically valid metarule imply its conclusion, they do not entail it in classical logic. For example, classical logic can only see in the premises and conclusion of a two-premise metarule something of the form $R(x, y)$, $R(z, w)$ and $R(v, t)$ (with $\vdash$ being the predicate taking the relevant multisets as arguments), and the argument from $R(x, y)$, $R(z, w)$ to $R(v, t)$ is not classically valid. (That argument is valid relative to a mathematical theory that adequately defines $\vdash$ for classical logic, but any such theory is of course far stronger than classical logic itself: the contrast remains that the premises of a classically valid metarule do not classically entail its conclusion, but only imply it, while, according to $\mathcal{N}$, the premises of the $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$-valid V^{ij}-L$^-$ not only imply its conclusion, but also entail it in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$.)

[38]In this connection, I think that, at least in some respects, $\mathcal{N}$ is a good theory of *counterlogicals*. For example, according to $\mathcal{N}$, the validity of the argument from φ to $\psi \mathbin{\&} \chi$ implies and entails the validity of the argument from φ to ψ (since both $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} (\varphi \Rightarrow \psi \mathbin{\&} \chi) \rightarrow (\varphi \Rightarrow \psi)$ and $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} (\varphi \Rightarrow \psi \mathbin{\&} \chi) \Rightarrow \varphi \Rightarrow \psi$ hold), which seem like good counterlogical judgements. I hasten to add however that, as it stands, $\mathcal{N}$ would in some other respects be regarded as too strong at least by theorists having certain intuitions. Consider any argument that, according to $\mathcal{N}$, is valid (for example, the argument from $\mathtt{t}$ to $\varphi \vee \neg\varphi$). According to $\mathcal{N}$, both the $\rightarrow$-conditional and the $\Rightarrow$-conditional whose antecedents deny the validity of that argument and whose consequent is arbitrary are true (indeed, logically true), contrary to some robust intuitions (for example, the conditionals 'If the law of excluded middle is (were) not valid, Brouwer is (would be) wrong' seem clearly false). Such conditionals

Moving on to some general claims that $\mathcal{N}$ contains about the properties of validity, it's easy to see that $\mathcal{N}$ contains the claims:

(TP) An argument is valid iff it is *truth preserving*;

(CR) Validity is *reflexive*, *monotonic* and *transitive*.[39]

Indeed, $\mathcal{N}$ declares all the instances of these claims to be logical truths in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$.[40] Claim (TP) characterises the *nature* of validity; claim (CR) some of its most prominent *formal properties*. Some traditional fundamental doctrines about validity are thus preserved in the otherwise radically new setting constituted by $\mathcal{N}$.

Moreover, it's easy to see that it is not the case that, according to $\mathcal{N}$, contraction is valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$, for that would require $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} (\bigwedge(\Gamma, \varphi, \varphi) \Rightarrow \bigvee(\Delta)) \rightarrow (\bigwedge(\Gamma, \varphi) \Rightarrow \bigvee(\Delta))$ to hold, which it doesn't (since, if it did, $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ would be subject to the Liar paradox, which it provably isn't). More strongly, not only does $\mathcal{N}$ not contain all instances of contraction, it in fact contains:

(NW) Validity is *non-contractive*,

in the natural sense of containing the negations of some conjunctions of instances of contraction (for example, the negation of the conjunction of the claim that contraction holds for the Liar sentence and the claim that contraction holds for the negation of the Liar sentence). As far as I can tell, as it stands $\mathcal{N}$ does *not* however contain *any particular instance of failure of contraction*. In this respect, $\mathcal{N}$ would be a *weaker* theory than a theory $\mathcal{C}$ endorsing $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ in a classical metalanguage (the theory in effect developed in sections 3 and 4), since the latter does contain the claim that, for example, contraction

are (logically) true according to $\mathcal{N}$ because, letting the target $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$-valid argument be the argument from $\varphi_1, \varphi_2, \varphi_3 \ldots, \varphi_i$ to $\psi_1, \psi_2, \psi_3 \ldots, \psi_j$, $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \varphi_1 \mathbin{\&} \varphi_2 \mathbin{\&} \varphi_3 \ldots \mathbin{\&} \varphi_i \Rightarrow \psi_1 \vee \psi_2 \vee \psi_3 \ldots \vee \psi_j$ holds, and hence, by $\neg$-L, $\neg(\varphi_1 \mathbin{\&} \varphi_2 \mathbin{\&} \varphi_3 \ldots \mathbin{\&} \varphi_i \Rightarrow \psi_1 \vee \psi_2 \vee \psi_3 \ldots \vee \psi_j) \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \varnothing$ holds, and so, by K-R, $\neg(\varphi_1 \mathbin{\&} \varphi_2 \mathbin{\&} \varphi_3 \ldots \mathbin{\&} \varphi_i \Rightarrow \psi_1 \vee \psi_2 \vee \psi_3 \ldots \vee \psi_j) \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \chi$ holds for arbitrary χ, whence, by $\rightarrow$-R, $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \neg(\varphi_1 \mathbin{\&} \varphi_2 \mathbin{\&} \varphi_3 \ldots \mathbin{\&} \varphi_i \Rightarrow \psi_1 \vee \psi_2 \vee \psi_3 \ldots \vee \psi_j) \rightarrow \chi$ holds (and, by $\Rightarrow$-R, $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \neg(\varphi_1 \mathbin{\&} \varphi_2 \mathbin{\&} \varphi_3 \ldots \mathbin{\&} \varphi_i \Rightarrow \psi_1 \vee \psi_2 \vee \psi_3 \ldots \vee \psi_j) \Rightarrow \chi$ holds). (Summing this up roughly, $\mathcal{N}$ as it stands is a good theory of counterlogicals whose antecedents envision a *strengthening* of $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$, but not a good theory of counterlogicals whose antecedents envision a *weakening* of $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$. I conjecture that this is an accident due to the fact that, as I'll shortly indicate in the text, $\mathcal{N}$ as it stands is not a good theory of what is *not* valid.) The proof given of why $\mathcal{N}$ is in conflict with certain intuitions about counterlogicals at the same time suggests the unsurprising and natural solution of the conflict: restrict K-R (and K-L). I won't investigate further in this paper how this suggestion could best be implemented.

[39]Validity is thus one of the *simple consequence relations* introduced by Avron [1991], in turn a generalisation of the *consequence relations* introduced by Scott [1974], in turn inspired by the *closure operations* introduced by Tarski [1930]. Thanks to an anonymous referee for comments that led to a revision of this fn.

[40]Even setting aside (TP) (which, involving $\Rightarrow$-implication and truth, in a classical framework is very hard to get to hold *even as a matter of fact* in either direction), this property is again lacked by a classical language: in classical logic, even the status of the instances of (CR) is at best that of *sophisticated mathematical* truths rather than that of *plain logical* truths (see also fn 37).

fails for the Liar sentence (and, for good measure, the claim that contraction fails for the negation of the Liar sentence). In this paper, I won't pursue further the very important but also very difficult question whether, within the non-contractive camp, one should prefer the stance of $\mathcal{N}$ or the stance of $\mathcal{C}$ with respect to the extent of failure of contraction.[41] What is important to observe here is that even the weaker sense in which contraction fails according to $\mathcal{N}$ is sufficient to show in a very straightforward way that $\mathcal{N}$ contains the claim:

(NC) Validity is *non-classical*,

in the natural sense of containing the negations of some conjunctions of instances of classically valid principles (for example, again, the negation of the conjunction of the claim that contraction holds for the Liar sentence and the claim that contraction holds for the negation of the Liar sentence).[42]

[41] To be clear, $\mathcal{N}$ is weaker than $\mathcal{C}$ in the sense that $\mathcal{C}$ contains a *non-validity* claim while $\mathcal{N}$ contains neither that claim nor its negation (pictorially, $\mathcal{C}$ takes a view on a case on which $\mathcal{N}$ takes no view). Of course, if we take instead the strength ordering in the sense in which, for example, classical logic is stronger than intuitionist logic, then the case under consideration shows that $\mathcal{N}$ is stronger than $\mathcal{C}$, at least in the sense that the latter *rules out as valid* an instance of a metarule that the former does not. I set out a more exhaustive if still brief comparison between $\mathcal{N}$ and $\mathcal{C}$ below in the text.

[42] Given a logical-consequence relation $\vdash$, let $\Vdash$ (the *external consequence relation* of $\vdash$) be such that $\Gamma \Vdash \Delta$ holds iff, roughly, the addition, for every φ member of Γ, of $\varnothing \vdash \varphi$ implies that $\varnothing \vdash \Delta$ holds (see Avron [1988]). The connections between a logical-consequence relation and its external consequence relation are of great interest. Presumably, a logical-consequence relation should be ("rulewise") *included* in its external consequence relation: for how can an argument be valid if it may be disobeyed by nothing less than the laws of logic themselves? (As I discuss in Zardini [2013b], the inclusion objectionably fails in the non-transitive theory of Ripley [2012] mentioned in section 3—indeed, it fails for arguments whose premises are in fact logical truths of the system.) The *converse* direction is however much more problematic: it is natural to think that, because of *special features* of *logical truth* as opposed to *mere truth*, an argument may preserve logical truth but not mere truth. And this is indeed what happens already in very familiar cases: for example, in the normal modal logic **K** as well as in many of its extensions, $\varphi \Vdash_{\mathbf{K}} \Box\varphi$ holds but $\varphi \vdash_{\mathbf{K}} \Box\varphi$ doesn't. It should thus be neither surprising nor objectionable that the same kind of divergence occurs with $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$: for example, $\varphi \Vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \varphi \,\&\, \varphi$ holds but, as observed in section 3, $\varphi \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \varphi \,\&\, \varphi$ doesn't. (Naturally, when this kind of divergence occurs, and so logical-truth preservation does not guarantee mere-truth preservation (and hence the fact that $\varphi \Vdash \psi$ holds does not imply that φ is in any reasonable sense a sufficient condition for ψ), one of the first causalties is the *deduction theorem*: for example, $\varphi \Vdash_{\mathbf{K}} \Box\varphi$ holds but $\varnothing \Vdash_{\mathbf{K}} \varphi \rightarrow \Box\varphi$ doesn't, and $\varphi \Vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \varphi \,\&\, \varphi$ holds but $\varnothing \Vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \varphi \rightarrow \varphi \,\&\, \varphi$ doesn't. Thus, while the *rules* of a logical-consequence relation may be included in the rules of its external consequence relation, this does not imply that also the *metarules* of the former are included in the metarules of the latter.) As the **K**-example makes clear, this kind of divergence can arise because of the presence of *modal* vocabulary that is *somehow sensitive to logical truths*: an analogous example for $\Rightarrow$ is that $\varphi \Vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \psi \Rightarrow \varphi$ holds but, as observed in section 4, $\varphi \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \psi \Rightarrow \varphi$ doesn't. While it is thus not at all reasonable to expect a logical-consequence relation to include its external consequence relation *over the full language with* $\Rightarrow$, it is very intriguing to note that, *over the restricted language without* $\Rightarrow$, while $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ still does not include its external consequence relation (as shown by our first example), under a natural understanding of what, according to $\mathcal{N}$, $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$'s external consequence relation is, it's easy to see that, according to $\mathcal{N}$, $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ does include its external consequence relation (a very pleasing result that is ultimately due to $\mathcal{N}$'s specific conception of what validity is). Since, according to $\mathcal{N}$, the converse direction also holds (as usual), *according to* $\mathcal{N}$

Moving on to the other two naive logical properties mentioned in this paper, inconsistency and compatibility, it is worth remarking that $\mathcal{N}$ can smoothly be extended to a theory of inconsistency and to an at least partial theory of compatibility. We can let $\mathcal{N}$ identify the *inconsistency* of φ with the truth of $T(\ulcorner\varphi\urcorner) \Rightarrow T(\ulcorner\mathsf{f}\urcorner)$ (which is of course just what it takes for $\varphi \vdash_{\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}} \varnothing$ to hold), and we can let $\mathcal{N}$ identify the *compatibility* of φ and ψ with the truth of $\neg(T(\ulcorner\varphi\urcorner) \Rightarrow T(\ulcorner\neg\psi\urcorner))$ (which is of course just what it takes for $\varphi \vdash_{\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}} \neg\psi$ not to hold).[43] Our previous conditions on what $\mathcal{N}$ contains suffice then to yield not only a theory of validity, but also a theory of inconsistency, and an at least partial theory of compatibility. Following up on fn 10, I stress the partiality of the theory of compatibility, given that $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ is manifestly too weak a theory for $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \neg(T(\ulcorner\varphi\urcorner) \Rightarrow T(\ulcorner\neg\psi\urcorner))$ to hold in a reasonable number of cases in which φ and ψ are naively compatible. It is an interesting project, if one going beyond the scope of this paper, to strengthen $\mathcal{N}$ to an adequate theory not only of validity and inconsistency, but also of compatibility (while preserving $\mathcal{N}$'s consistency), and it will suffice here to have flagged this direction of future research. Relatedly, I also note that, obviously, there are notions of inconsistency and compatibility that apply to multisets of sentences (rather than only to sentences or pairs of sentences), and that there also are logical notions beyond validity, inconsistency and compatibility. In many cases, the treatment of these notions in the current framework is completely straightforward, but, again, for the rest of this paper I will rest content instead with consolidating what I've done by giving some more details about $\mathcal{N}$ as a theory of validity and inconsistency.

Similarly to what was the case for validity, once we have adopted $\mathcal{N}$ as our theory of inconsistency, adding a naive-inconsistency predicate to the object language of $\mathcal{N}$ is easy. Indeed, just as it was the case for a naive-validity predicate, there is no need to *add it* since *it is already there*. In fact, it is already there, as it were, in its full splendour, in the sense that it is already there as exactly the same predicate that we've just used to define what inconsistency is according to $\mathcal{N}$, i.e. the predicate $T(x) \Rightarrow T(\ulcorner\mathsf{f}\urcorner)$ (which for brevity I'll henceforth also write as $I(x)$: from now on, the latter predicate should always be understood as defined by the former).

We can show that, according to $\mathcal{N}$, I as so defined obeys I-L and I-R when $\vdash$ is $\vdash_{\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}}$. On the one hand, I-L basically turns on the law of non-contradiction for $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ (i.e. $\varphi, \neg\varphi \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \varnothing$) and in particular on the fact that, given some routine fiddling:

$\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ *coincides with its external consequence relation*, as is the case for many *non-substructural* logics and contrary to what is the case for many *substructural* logics.

[43] $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \neg(\varphi \to \neg\psi) \Rightarrow \varphi \,\&\, \psi$ holds. If $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \neg(\varphi \Rightarrow \neg\psi) \Rightarrow \varphi \,\&\, \psi$ also held, that would be disastrous for the proposed notion of compatibility, since the argument from 'φ' and 'ψ' are compatible' to 'φ' is true and 'ψ' is true' would be valid, which it clearly isn't for any ordinary notion of compatibility. Fortunately, $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \neg(\varphi \Rightarrow \neg\psi) \Rightarrow \varphi \,\&\, \psi$ does not hold. It is instructive to analyse how one of the most natural arguments establishing that $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \neg(\varphi \to \neg\psi) \Rightarrow \varphi \,\&\, \psi$ holds fails in the case of $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \neg(\varphi \Rightarrow \neg\psi) \Rightarrow \varphi \,\&\, \psi$. Basically, the argument proceeds by establishing that $\neg(\varphi \,\&\, \psi) \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \varphi \to \neg\psi$ holds, and does this in turn by applying $\to$-R to the fact that $\neg(\varphi \,\&\, \psi), \varphi \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \neg\psi$ (the rule of *modus ponendo tollens*) holds. The relevant instance of $\to$-R is however with a (possibly) non-logical$^{\Rightarrow}$ side premise ($\neg(\varphi \,\&\, \psi)$), and hence $\Rightarrow$-R cannot analogously be applied. Similar considerations hold by substituting $\to$ for $\Rightarrow$ as main operator in the wffs mentioned at the beginning of this fn.

$\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} \neg(\varphi \Rightarrow \mathsf{f}) \Rightarrow (\varphi \Rightarrow \mathsf{f}) \Rightarrow \mathsf{f}$

holds, which, by transparency, means that, according to $\mathcal{N}$, the premise of I-L entails in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$ its conclusion.[44] On the other hand, I-R basically turns on $\Rightarrow$-R and in particular on the fact that, given some routine fiddling:

$\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} (\varphi \Rightarrow \mathsf{f}) \Rightarrow \mathsf{t} \Rightarrow \varphi \Rightarrow \mathsf{f}$

holds, which, by transparency, means that, according to $\mathcal{N}$, the premise of I-R entails in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$ its conclusion. Thus, according to $\mathcal{N}$, I meets condition (i) of (NIP) for being a naive-inconsistency predicate for $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$. Moreover, according to $\mathcal{N}$, I is trivially fully intersubstitutable with an object-language inconsistency predicate for $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$ (namely, itself!). Thus, according to $\mathcal{N}$, I also meets condition (ii) of (NIP) for being a naive-inconsistency predicate for $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$, and hence, according to $\mathcal{N}$, $I(x)$ (i.e. $T(x) \Rightarrow T(\ulcorner \mathsf{f} \urcorner)$) is a naive-inconsistency predicate for $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$.

It's very interesting to observe that the fact that, according to $\mathcal{N}$, in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$ $T(x) \Rightarrow T(\ulcorner \mathsf{f} \urcorner)$ obeys the relevant analogue of I-L follows from the more general fact that, according to $\mathcal{N}$, in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$ $\Rightarrow$ does finally obey the analogue for it of V-L (i.e. if $\varphi \nvdash_{\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}} \psi$ holds, then $\varphi \Rightarrow \psi \vdash_{\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}} \varnothing$ holds). That analogue also basically turns on the law of non-contradiction for $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ and in particular on the fact that, given some routine fiddling:

$\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} \neg(\varphi \Rightarrow \psi) \Rightarrow (\varphi \Rightarrow \psi) \Rightarrow \mathsf{f}$

holds, which, by transparency, means that, according to $\mathcal{N}$, the premise of V-L entails in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$ the relevant $\Rightarrow$-analogue of its conclusion. However, according to $\mathcal{C}$, in $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ $\Rightarrow$ does not obey that analogue of V-L (since if it did $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ would be subject to paradox (V), which it provably isn't). In this respect, $\mathcal{N}$ is a *stronger* theory than $\mathcal{C}$, in the sense that it contains a *validity* claim (of a metarule with a negative premise such as V-L) while $\mathcal{C}$ does not contain that claim (indeed contains the opposite claim that certain specific instances of the metarule in question fail).

Although, according to $\mathcal{N}$, I-L, I-R, V-L and V-R are valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$, neither the official reasoning of paradox (I) nor that of paradox (V) go through any more once $\mathcal{N}$ is adopted as a theory of validity, since those reasonings rely, among other things, on *reductio* being valid in the logic of the metalanguage, whereas it is not the case that, according to $\mathcal{N}$, *reductio* is valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$ (which, according to $\mathcal{N}$, is the logic of the metalanguage $\mathscr{N}$). This is so because the fact that, according to $\mathcal{N}$, *reductio* is valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$ would require that $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} (\varphi \Rightarrow \neg\varphi) \rightarrow (\mathsf{t} \Rightarrow \neg\varphi)$ held, which it doesn't (since if it did $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ would be subject to the Liar paradox, which it provably isn't). The more sophisticated reasoning I offered in section 5 on behalf of paradoxes (I) and (V) does not go through either, since, as I stressed in that section, it relies, among other

[44] To follow up on an issue briefly emerged in fn 29, the converse direction also holds, basically turning on the law of non-contradiction for $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ as well as on the law of excluded middle for $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ (i.e. $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} \varphi, \neg\varphi$) and in particular on the fact that, given some routine fiddling:
$\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow \mathsf{tf}}} ((\varphi \Rightarrow \mathsf{f}) \Rightarrow \mathsf{f}) \Rightarrow \neg(\varphi \Rightarrow \mathsf{f})$
holds, which, by transparency, means that, according to $\mathcal{N}$, the conclusion of I-L entails in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$ its premise.

things, on $\varphi \to \psi, \psi \to \chi \vdash \varphi \to \chi \,\&\, \varphi$ holding in the logic $\vdash$ of the metalanguage (in the step from (1) and (2) to (3)), whereas it is not the case that, according to $\mathcal{N}$, that argument is valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ (which, according to $\mathcal{N}$, is the logic of the metalanguage $\mathscr{N}$).[45] This is so because the fact that, according to $\mathcal{N}$, that argument is valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ would require that $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} (\varphi \to \psi) \,\&\, (\psi \to \chi) \Rightarrow (\varphi \to \chi \,\&\, \varphi)$ held, which it doesn't (since if it did $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ would be subject to the Liar paradox, which it provably isn't). A deeper analysis of why the argument in question fails would decompose it in two parts: one part is the argument $\varphi \to \psi, \psi \to \chi \vdash \varphi \to \chi$, which expresses the transitivity of $\to$ and which, according to $\mathcal{N}$, is in fact valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$; the other part is the argument $\varphi \to \chi \vdash \varphi \to \chi \,\&\, \varphi$, whose validity would require φ to contract and which is not such as to be, according to $\mathcal{N}$, valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$.

Similarly, the reasoning of the argument for the redundancy of *modus ponens* presented in section 5 does not go through any more once $\mathcal{N}$ is adopted as a theory of validity, since that reasoning relies, among other things, on reasoning by cases being valid in the logic of the metalanguage, whereas it is not the case that, according to $\mathcal{N}$, reasoning by cases is valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ (which, according to $\mathcal{N}$, is the logic of the metalanguage $\mathscr{N}$). This is so because the fact that, according to $\mathcal{N}$, reasoning by cases is valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ would require that $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} (\varphi \Rightarrow \chi) \,\&\, (\psi \Rightarrow \chi) \to (\varphi \vee \psi \Rightarrow \chi)$ held, which it doesn't (since if it did $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ would be subject to the Liar paradox, which it provably isn't). Thus, heterodox as it may be in many other respects, $\mathcal{N}$ manages to validate the relevant $\Rightarrow$-analogue of V-L (needed to secure the strength required from $\Rightarrow$ in order for it to help to express naive validity and inconsistency) without thereby committing the grave offence of declaring *modus ponens* for $\Rightarrow$ redundant. Somewhat surprisingly in a non-contractive framework, $\mathcal{N}$ both has its cake and eats it.

$\mathcal{N}$ thus offers, in my view, a very appealing package. It remains to observe that, from the perspective of our classical metametalanguage, it has various consistency properties which are a trivial consequence of the corresponding consistency properties for $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ and of the fact that $\mathcal{N}$ simply consists of the set of logical truths of $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ (plus its identification of validity with $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$-validity). So, for example, we know that $\mathcal{N}$ is *non-trivial* (i.e. does not contain every sentence) because we know that it is not the case that every sentence is a logical truth of $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ (see Zardini [2011], pp. 524–532 for a consistency proof that can easily be extended to a proof of this claim, as already mentioned in section 4). And, for example, we know that $\mathcal{N}$ is *non-contradictory* (i.e. does not contain both a sentence and its negation) because we know that it is not the case that both a sentence and its negation are a logical truth of $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ (again, see Zardini [2011], pp. 524–532 for a consistency proof that can easily be extended to a proof of this claim, as already mentioned in section 4). And it's easy to see that, according to $\mathcal{N}$, validity itself (i.e. $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$) is non-trivial and non-contradictory (i.e. according to $\mathcal{N}$, not every sentence is a logical truth of $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ and in particular no sentence and its negation are both logical truths of $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$). Moreover, just as, according to $\mathcal{C}$, $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ proves that

[45] An analogous reasoning would substitute the $\Rightarrow$-conditional for the $\to$-conditional. All the points I'm making here about the reasoning with the $\to$-conditional would apply just as well to the analogous reasoning with the $\Rightarrow$-conditional.

each of its valid arguments preserves truth, according to $\mathcal{N}$ $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ proves that each of its valid arguments preserves truth: thus, in a rather minimal sense, according to $\mathcal{N}$ $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ proves its own soundness (still, a sense in which, as we've seen in section 3, most non-classical solutions cannot prove their own soundness). Whether $\mathcal{N}$ and $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ can be extended in such a way that, according to $\mathcal{N}$, $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ proves its own soundness in a more interesting sense, will depend on the details of its extension to a theory of quantification and to some elementary theory of syntax, topics that I'm not addressing in this paper.

With all these facts in view, we can conclude with a brief comparison of $\mathcal{N}$ (the theory of validity expressed in a non-classical metalanguage presented in this section) and $\mathcal{C}$ (the theory of validity expressed in a classical metalanguage presented in sections 3 and 4). We've already seen in section 5 the weighty philosophical reasons to shift from $\mathcal{C}$ to $\mathcal{N}$. We can here start by briefly comparing the two logics they recommend ($\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ and $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ respectively). It's easy to see that the two logics agree on *logical truths* and *rules*: according to $\mathcal{N}$, $\Gamma \vdash_{\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}} \Delta$ holds iff, according to $\mathcal{C}$, $\Gamma \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \Delta$ holds. Their difference only consists in the attitudes they take towards certain *metarules*. On the one hand, as for the metarule of contraction (which has a positive premise), according to $\mathcal{C}$ certain specific instances are not valid in $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$, while it is not the case that, according to $\mathcal{N}$, such instances are not valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ (although, as we've seen in the discussion of (NW), nor is it that case that, according to $\mathcal{N}$, such instances are valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$, and although, according to $\mathcal{N}$, not all such instances are valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$). On the other hand, as for the metarule V-L (which has a negative premise), according to $\mathcal{N}$ it is valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$, while, according to $\mathcal{C}$, certain specific instances of it fail in $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$. These are both very interesting differences and should also help to decide whether to adopt $\mathcal{N}$ or $\mathcal{C}$ (while remaining neutral about the first difference just mentioned, I should like to record a naivety-driven sympathy for a theory that validates V-L). Both differences are clearly due to the fact that $\mathcal{C}$ is a stronger theory of what rules are *not valid* for $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ than $\mathcal{N}$ is of what rules are *not valid* for $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ (indeed, as I've already stressed, $\mathcal{N}$ as it stands is insufficiently strong in this respect, but, although it is a topic for another paper, we can reasonably expect that it can be strengthened while preserving these two differences with $\mathcal{C}$).

Beyond these differences in metarules, a more general difference between $\mathcal{N}$ and $\mathcal{C}$ should also be recorded (indeed, between $\mathcal{N}$ and any standard theory of validity expressed in a classical metalanguage). While, at least together with the underlying mathematical facts (in particular, the mathematical fact of which is the smallest set containing all instances of I and closed under the metarules given in section 3), $\mathcal{C}$ determines for any rule or metarule whether it is valid in $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ or not, the whole point of defining $\mathcal{N}$ the way I did was exactly to avoid this *completeness* with respect to its target logic $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$. For some rules and metarules, $\mathcal{N}$ neither contains the claim that they are valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ nor the claim that they are not valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ (although it contains the claim that they are either valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ or not valid in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$). Not only is this a fact about $\mathcal{N}$ *as defined so far*, but also any acceptable *extension* of $\mathcal{N}$ has to do so. For $\mathcal{N}$ as defined so far contains a claim φ only if it also contains its self-conjunction $\varphi \,\&\, \varphi$ (since $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \varphi$ holds only if $\varnothing \vdash_{\mathbf{IKT}_{\Rightarrow\mathsf{tf}}} \varphi \,\&\, \varphi$ holds)—and we may assume that any acceptable extension of $\mathcal{N}$ will preserve this feature—but some claims stating the validity of rules or metarules in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$ have the peculiar status that both their self-conjunction and the self-conjunction of their negation are inconsistent in $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$. One example is the sentence ι identical to $T(\ulcorner\iota\urcorner) \Rightarrow T(\ulcorner\mathsf{f}\urcorner)$: it's easy to see that, according to $\mathcal{N}$, both $\iota \,\&\, \iota \vdash_{\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}} \varnothing$ and $\neg\iota \,\&\, \neg\iota \vdash_{\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}} \varnothing$ hold. Now, it's easy to see that $\mathcal{N}$ as defined so far is closed under what, according to $\mathcal{N}$, is $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$-validity—and we may assume that any acceptable extension of $\mathcal{N}$ will preserve this feature. If so, not only does $\mathcal{N}$ as defined so far neither contain ι nor $\neg\iota$, but also any acceptable extension of $\mathcal{N}$ is as a matter of principle precluded from containing ι or $\neg\iota$—that is, as a matter of principle precluded from containing the claim that $\iota \vdash_{\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}} \mathsf{f}$ holds or the claim that $\iota \vdash_{\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}} \mathsf{f}$ does not hold. Contrary to $\mathcal{C}$, $\mathcal{N}$ is an *essentially incomplete* theory of its target logic.

As far as this paper is concerned, we've thus reached the end of the ladder. As the classical metametalanguage has served us so well in assuring us of the many virtues of $\mathcal{N}$, it'd really be a shame to throw it away. Do we have to? If we had used a non-classical metalanguage, say, to provide a *model theory* for the target non-classical logic, and used a classical metametalanguage to assure us, again in a *model-theoretic* fashion, of the many virtues of the non-classical model theory, there would indeed be pressure to throw away such ladder. Generally, the adoption of a non-classical model theory is presumably driven by the idea that *model-theoretic objects really behave non-classically*, and do so to the extent that the assumption that they behave classically would lead to an objectionably inadequate representation of the target non-classical logic; but, if so, surely the same assumption would also lead to an objectionably inadequate representation of the non-classical model theory (even more so if the logic of the non-classical model theory coincides with the target non-classical logic). More specifically, there would surely be as much reason to *add predicates of naive logical properties* to a full-blown non-classical model theory as there is to add them to a streamlined non-classical logic (even more so if the logic of the non-classical model theory coincides with the target non-classical logic), but the addition of such predicates to the relevant object language has been recognised in section 5 to require a non-classical metalanguage.

However, I think that our construction is, in the relevant respects, crucially different from the just envisaged model-theoretic one, and under no similar pressure to throw away its ladder. For our construction only regards as non-classical the behaviour of *truth* (and consequently does not use a classical language in talking about truth); the classical language is, however, only used to talk about *the theory $\mathcal{N}$ and its properties* (as the behaviour of these is regarded as classical). Now, since $\mathcal{N}$ is supposed to be the theory *we should accept*, it might still be worried that talk about $\mathcal{N}$ is after all a concealed way of talking about truth or something equivalent with it (and hence with an equally non-classical behaviour), for it might be thought that *we should accept all and only true sentences*. But, as I've been at pains to stress in this section, that is a *completeness ideal on what we should accept* that should be rejected in this area. $\mathcal{N}$ has been constructed with this essential incompleteness in view, and this is why it has not been identified with the set of all and only true sentences, but with the set of all and only logical truths of $\mathbf{IKT}_{\Rightarrow\mathsf{tf}}$ (plus the identification of validity with $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow\mathsf{tf}}$-validity). The latter set is indeed classical, and hence the use of a classical language in talking about it, far from being objectionably inadequate, is optimally adequate (indeed, it might be argued that any other choice should exhibit the same feature, since it might be argued that what we accept should be a classical matter). Thus, $\mathcal{N}$ and its properties behave classically even if what $\mathcal{N}$ talks about behaves non-classically, and that is why it is optimally adequate that our metametalanguage (which only talks about $\mathcal{N}$ and its properties) is classical while our metalanguage (which talks about truth, validity and truth preservation) is non-classical. Here just as well as in common life, the ladder should be let to stand.[46]

7 Conclusion

I have argued that a proper appreciation of the paradoxes of logical properties forces one wishing to accommodate for naive logical properties to shift to a non-classical metalanguage in expressing one's favoured non-classical logic. I have shown, within certain limitations, how this can be done in a very simple, natural and systematic way, using as a ladder an interesting logic of naive truth ($\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$) expressed by a theory ($\mathcal{C}$) in a classical metalanguage (a logic which is probably at least very close to the best one can do in a classical metalanguage to accommodate for naive logical properties). At the end of the ladder, there stands an equally interesting logic of naive truth and naive validity ($\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$)—indeed, a logic where naive validity just is a certain form of naive-truth preservation—expressed by a theory ($\mathcal{N}$) in a non-classical metalanguage. I've explored in some detail some of the salient properties of $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$ and of $\mathcal{N}$ itself. I think that the results have been pleasing enough to discourage the thought that naive logical properties lie beyond redemption, and maybe even so pleasing to encourage some to shift from $\mathbf{IKT}_{\Rightarrow \mathsf{tf}}$ to $\mathbf{IKT}^{\circlearrowleft}_{\Rightarrow \mathsf{tf}}$ as the correct logic of truth and validity.

References

Earline Ashworth. *Language and Logic in the Post-Medieval Period.* Reidel, Dordrecht, 1974.

Arnon Avron. The semantics and proof theory of linear logic. *Theoretical Computer Science*, 57:161–184, 1988.

Arnon Avron. Simple consequence relations. *Information and Computation*, 92:105–140, 1991.

JC Beall. *Spandrels of Truth.* Oxford University Press, Oxford, 2009.

JC Beall and Julien Murzi. Two flavors of Curry's paradox. *The Journal of Philosophy*, 110:143–165, 2013.

[46] A previous version of this paper did foolishly throw away the ladder. Thanks to an anonymous referee for comments that led to its recovery.

Ross Brady. *Universal Logic*. CSLI Publications, Stanford, 2006.

Tyler Burge. Buridan and epistemic paradox. *Philosophical Studies*, 34:21–35, 1978.

Hartry Field. Truth and the unprovability of consistency. *Mind*, 115:567–605, 2006.

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

Patrick Greenough. Free assumptions and the liar paradox. *American Philosophical Quarterly*, 38:115–135, 2001.

Kurt Grelling and Leonard Nelson. Bemerkungen zu den Paradoxien von Russell und Burali-Forti. *Abhandlungen der Fries'schen Schule*, 2:301–334, 1908.

Anil Gupta and Nuel Belnap. *The Revision Theory of Truth*. MIT Press, Cambridge MA, 1993.

David Kaplan and Richard Montague. A paradox regained. *Notre Dame Journal of Formal Logic*, 1:79–90, 1960.

Saul Kripke. Outline of a theory of truth. *The Journal of Philosophy*, 72:690–716, 1975.

Benson Mates. Pseudo-Scotus on the soundness of *consequentiae*. In Anna-Teresa Tymieniecka, editor, *Contributions to Logic and Methodology in Honor of J.M. Bocheński*, pages 132–141. North-Holland, Amsterdam, 1965.

Vann McGee. *Truth, Vagueness, and Paradox*. Hackett, Indianapolis, 1991.

Robert Meyer, Richard Routley, and Michael Dunn. Curry's paradox. *Analysis*, 39: 124–128, 1979.

Richard Montague. Syntactic treatments of modality, with corollaries on reflection principles and finite axiomatizability. *Acta Philosophica Fennica*, 16:153–167, 1963.

Sebastiano Moruzzi and Elia Zardini. Conseguenza logica. In Annalisa Coliva, editor, *Filosofia analitica*, pages 157–194. Carocci, Roma, 2007.

Dag Prawitz. *Natural Deduction*. Almqvist och Wiksell, Stockholm, 1965.

Graham Priest. *Beyond the Limits of Thought*. Oxford University Press, Oxford, 2nd edition, 2003.

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

Graham Priest. Hopes fade for saving truth. *Philosophy*, 85:109–140, 2010.

Graham Priest. Fusion and confusion. *Topoi*, 2013. Forthcoming.

Stephen Read. Self-reference and validity. *Synthese*, 42:265–274, 1979.

Stephen Read. Self-reference and validity revisited. In Mikko Yrjönsuuri, editor, *Medieval Formal Logic*, pages 183–196. Kluwer, Dordrecht, 2001.

David Ripley. Conservatively extending classical logic with transparent truth. *The Review of Symbolic Logic*, 5:354–378, 2012.

Bertrand Russell. *The Principles of Mathematics*. Cambridge University Press, Cambridge, 1903.

Bertrand Russell. Mathematical logic as based on the theory of types. *American Journal of Mathematics*, 30:222–62, 1908.

Dana Scott. Completeness and axiomatizability in many-valued logic. In Leon Henkin, editor, *Proceedings of the Tarski Symposium*, pages 411–435. American Mathematical Society, Providence RI, 1974.

Lionel Shapiro. Deflating logical consequence. *The Philosophical Quarterly*, 61:320–342, 2011.

Alfred Tarski. Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften. I. *Monatshefte für Mathematik und Physik*, 37:361–404, 1930.

Alfred Tarski. *Pojęcie prawdy w językach nauk dedukcyjnych*. Nakładem Towarzystwa Naukowego Warszawskiego, Warsaw, 1933.

Alan Weir. Naive truth and sophisticated logic. In Bradley Armour-Garb and JC Beall, editors, *Deflationism and Paradox*, pages 218–249. Oxford University Press, Oxford, 2005.

Elia Zardini. Truth and what is said. *Philosophical Perspectives*, 22:545–574, 2008a.

Elia Zardini. A model of tolerance. *Studia Logica*, 90:337–368, 2008b.

Elia Zardini. Truth without contra(di)ction. *The Review of Symbolic Logic*, 4:498–535, 2011.

Elia Zardini. Truth preservation in context and in its place. In Catarina Dutilh-Novaes and Ole Hjortland, editors, *Insolubles and Consequences*, pages 249–271. College Publications, London, 2012.

Elia Zardini. 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*, 1:309–319, 2013a.

Elia Zardini. Naive *modus ponens*. *Journal of Philosophical Logic*, 42:575–593, 2013b.

Elia Zardini. Context and consequence. An intercontextual substructural logic. *Synthese*, 2013c. Forthcoming.

Elia Zardini. Getting one for two, or the contractors' bad deal. Towards a unified solution to the semantic paradoxes. In Theodora Achourioti, Kentaro Fujimoto, Henri Galinon, and José Martínez, editors, *Unifying the Philosophy of Truth*. Springer, Berlin, 2013d. Forthcoming.

Elia Zardini. Naive logical properties and structural properties. *The Journal of Philosophy*, 2013e. Forthcoming.

Elia Zardini. Restriction by non-contraction. *Notre Dame Journal of Formal Logic*, 2013f. Forthcoming.

Elia Zardini. $\forall$ and ω. ms, 2013g.

Elia Zardini. The bearers of logical consequence. ms, 2013h.

Elia Zardini. The opacity of truth. ms, 2013i.

Elia Zardini. The underdetermination of the meaning of logical words by rules of inference. ms, 2013j.