ORIGINAL RESEARCH

Should Theories of Logical Validity Self-Apply?

Marco Grossi[1]

Abstract

Some philosophers argue that a theory of logical validity should not interpret its own language, because a Russellian argument shows that self-applicability is inconsistent with the ability to capture all the interpretations of its own language. First, I set up a formal system to examine the Russellian argument. I then defend the need for self-applicability. I argue that self-applicability seems to be implied by generality, and that the Russellian argument rests on a test for meaning that is biased against self-applicability. I propose an alternative test which is unproblematic for self-applicability. I conclude that self-applicability can be vindicated and the Russellian argument can be resisted.

1 Self-Applicability and Russellian Concerns

In this article, I focus on logical validity, rather than on a more generic notion of validity. A formula ϕ logically follows from a set of premises Γ exactly when, in all interpretations of the non-logical vocabulary, if all the premises in Γ are true so-and-so interpreted, so is ϕ. I take the standard model-theoretic account of logical validity to fit this definition.[1]

   To define logical validity we need a notion of satisfaction in an interpretation. Satisfaction in an interpretation is usually defined for a specific range of languages that share a common syntax (propositional languages, modal languages, higher-order languages etc). The scope of what a theory of logical validity can cover is determined by the class of languages for which satisfaction in an interpretation is defined. Sometimes a theory of logical validity is self-applicable: it is able to interpret

---

[1] Etchemendy ([1990](#), [2008](#)) notably argued against a purely interpretational account of models, claiming they represent possible worlds. Still, most will not dispense with interpretations altogether, and all I need to assume for my arguments is that preservation of satisfaction in an interpretation is necessary for logical validity. Also, structurally similar arguments can likely be put forward by switching interpretations for possible worlds.

✉ Marco Grossi
  marco.grossi@new.ox.ac.uk

[1] Faculty of Philosophy, University of Oxford, New College, Holywell st, OX13BN Oxford, UK

 Springer

languages of the same type as the meta-language. We can check this by looking at the class of languages for which satisfaction in an interpretation is defined.[2] Self-applicability implies that satisfaction in an interpretation is untyped, i.e. it applies to sentences where the predicate 'is satisfied in an interpretation' occurs. I call logical validity untyped if satisfaction in an interpretation is untyped. The two notions are different, since 'self-applicable' is more general than 'untyped'. Yet, they should be extensionally equivalent: if satisfaction is defined for languages of the same type as the meta-language then it is defined for the meta-language, too; and vice-versa if it is defined for the meta-language then the semantic machinery should be extendable to any language of the same kind of the meta-language.

Definition 1 (*Self-applicability*) T is self-applicable: satisfaction in an interpretation in T is defined for all languages of the same type as L_T.

Definition 2 (*Untyped Satisfaction and Logical Validity*) Satisfaction in an interpretation in T is untyped if it is defined for all formulae of T. Logical validity in T is untyped if satisfaction in an interpretation in T is untyped.

For the sake of this article, I take self-applicable and untyped to be extensionally equivalent.

One might ask what is the point of this discussion, since we know that logical validity self-applies. The text-book theory of logical validity is model theory, which is spelled out in first-order Zermelo-Fraenkel set-theory with choice (ZFC). It is a theory of logical validity for all first-order languages, in particular for the language of ZFC. In fact, classical logical validity for first-order languages is already definable in PA, as a direct consequence of soundness and Gödel's completeness theorem (Ketland, 2012). However, these considerations hold for logical consequence for first-order languages only. The interesting, open question is whether logical validity for any language self-applies, in particular for higher-order languages that are essentially incomplete.

Recently some philosophers (which I will call 'higher-orderists') have criticised self-applicability in higher-order languages (Rayo & Uzquiano, 1999; Rayo & Williamson, 2003; Rayo, 2006). According to them, when we interpret a language of order $n > 2$, we need order $n + 1$ resources. Higher-orderists claim that any self-applicable theory is not general and therefore not conceptually adequate. Their argument is based upon the notion of 'semantic openness', which we can informally introduce as follows:

---

[2] I am sure there are ways to abuse this definition. For example, one could set down recursive axioms for all formulae of L for each connective except for negation. The axioms would 'define' satisfaction in an interpretation for all formulae of L only on paper. Hopefully the definition is clear enough to be workable.

Definition 3 (*Semantic Openness*) A theory of validity *T* is semantically open if and only if, given any predicate *P* of the language we are interpreting and any ϕ of the language of *T*, there is an interpretation *I* where *P* means ϕ.

Semantic openness tests generality. Semantic openness is desirable: if the theory of validity is not semantically open, it is not able to capture all the intuitively acceptable interpretations of a given language, so it might deliver incorrect results: some formula might be true in all the interpretations the theory captures, but false in an unreachable one. In first-order logic these worries are set aside, since we can run a squeezing argument to show that the usual set-based definition of logical validity in ZFC is equivalent to 'intuitive logical validity' (Kreisel, 1967). Reflection principles are also provable in ZFC (Montague, 1961; Levy, 1960). Reflection principles ensure that if a formula is satisfied in some class-sized model, it is satisfiable in a set-sized model (Shapiro, 1987). When we venture into higher-order logic, however, things get more complicated. Higher-order logic is incomplete under the standard semantics, so no squeezing argument is available. Reflection principles are not provable either if our meta-theory is second-order ZFC. Thus, if our aim is to give a theory of logical validity for higher-order languages, a lack of semantic openness might be concerning.

In this article, I will focus on a Russellian argument that higher-orderists bring forward to argue for the incompatibility of self-applicability and semantic openness. The argument originated from Williamson (2003, 426). It was then referenced in a number of articles and books, with some minor modifications (Rayo & Williamson, 2003; Linnebo, 2006; Florio, 2014; Studd, 2019). To get to a more formal version of the argument, we first need to set down a simple formal theory to work on. This is the topic of the next section.

1.1 The Theory *S*

For my discussion I use a very weak base that I call *S*. This base is meant to be schematic, useful nonsense; it is just a way to talk about different theories of logical validity, which will mold the primitive terms as they see fit.

The signature L_S of *S* is $\{=, Sat, \langle\rangle\}$. This signature can be extended as needed when applied to different theories of validity. *Sat* is a ternary predicate letter, which is meant to apply to formulae, sequences and interpretations. $\langle\rangle$ is a primitive function for *n*-tuples. Where $v_1, \ldots, v_n$ are variables, $\langle v_1, \ldots, v_n \rangle$ is a sequential term. The results in this article do not rely on making $\langle\rangle$ a primitive term.[3]

An atomic formula is $Sat(v_1, \langle v_2, \ldots, v_n \rangle, v_{n+1})$ or $v_n = v_m$. Where ϕ and ψ are formulae of L_S, $\neg\phi \mid \phi \wedge \psi \mid \forall v_n \phi$ are formulae of L_S. Nothing else is a formula of L_S. The usual other quantifiers and connectives are meta-abbreviations. We also

---

[3] What I prove is already provable in Peano Arithmetic (PA) without using sequential terms, using the fact that there is a closed term for every object. Secondly, what I prove can be shown in a language like that of ZFC, with the additional axiom that substitution of identicals holds under satisfaction in an interpretation. The proofs are just lengthier and a bit more convoluted.

abbreviate strings of quantifiers $\forall x_1 \ldots \forall x_n$ as $\forall x_1, \ldots, x_n$. We abbreviate strings of terms $x_1, \ldots, x_n$ as $\underset{n}{\overset{1}{x}}$.

About the meta-theory S: all we need to assume is that S is strong enough to define n-tuples and coding. That is, we assume that S proves the following:

$$\underset{n}{\overset{1}{\forall x}}, \underset{n}{\overset{1}{y}} \left(\underset{n}{\overset{1}{\langle x \rangle}} = \underset{n}{\overset{1}{\langle y \rangle}} \leftrightarrow (x_1 = y_1 \wedge \cdots \wedge x_n = y_n) \right) \tag{Ext}$$

This is already possible in fragments of Peano Arithmetic (Hájek & Pudlák, 2016). We don't need to assume anything else to produce the Russellian argument, in particular we don't need any additional axiom for *Sat*.

Throughout the paper I help myself with the following notation: we use ϕ, ψ etc for formulae, and $\ulcorner \phi \urcorner$ for the code of ϕ. We use $\mathcal{I}$ as a meta-variable for interpretations (which will be defined in different ways in different semantic theories). We paraphrase $Sat(\ulcorner \phi \urcorner, \underset{n}{\overset{1}{\langle x \rangle}}, \mathcal{I})$ as ϕ is satisfied by $\underset{n}{\overset{1}{x}}$ in $\mathcal{I}$.

1.2 The Russellian Argument

We can now turn to the Russellian argument against self-applicability. Semantic openness is a pre-formal notion, meant to be something we gesture at rather than formalise. When it comes to the Russellian argument, however, we need a way to link the idiomatic 'P means to ϕ in Γ' to a formal requirement in S. This is implicitly done in the Russellian argument via the following translation $T1$:

Definition 4 (*T1*) An n-ary predicate P means to ϕ in an interpretation $\mathcal{I}$[4]:

$$\underset{n}{\overset{1}{\forall x}} \left(Sat(\ulcorner P(\underset{n}{\overset{1}{v}}) \urcorner, \underset{n}{\overset{1}{\langle x \rangle}}, \mathcal{I}) \leftrightarrow \phi(\underset{n}{\overset{1}{x}}) \right)$$

With T1 in place, semantic openness can be formalised as follows in S:

Definition 5 (*SO1*) Given any ϕ of L_S with n free variables and any n-ary P of the language we are interpreting:

$$\exists \mathcal{I} \underset{n}{\overset{1}{\forall x}} \left(Sat(\ulcorner P(\underset{n}{\overset{1}{v}}) \urcorner, \underset{n}{\overset{1}{\langle x \rangle}}, \mathcal{I}) \leftrightarrow \phi(\underset{n}{\overset{1}{x}}) \right)$$

The following is immediate:

Theorem 1 $S + SO1$ *is inconsistent.*

To see this, just instantiate ϕ with $\ulcorner \neg Sat(\ulcorner P(v) \urcorner, \langle u \rangle, u) \urcorner$ and x with $\mathcal{I}$.

---

[4] I am ignoring whether P applies to sequential terms or not, to keep the definition as simple as possible. When we apply T1 to satisfaction, we are slightly changing it to accommodate the fact that satisfaction applies to sequential terms and terms.

Some comments: one would uphold SO1 only if

a   They believe S should be semantically open;
b   They believe T1 is a fair translation of the idiomatic 'means that';
c   They take S to be self-applicable.

(c) is implicit in the fact that we take $\neg Sat(\ulcorner P(v) \urcorner, \langle u \rangle, u)$ to be an appropriate interpretation of P: this makes sense only if P is of the same type as Sat. Therefore, what the Russellian argument shows is that the pre-formal notion of semantic openness, T1, and self-applicability are inconsistent.

1.3 The Typing Escape

Higher-orderists escape the Russellian argument by typing the language, thereby restricting the range of available ϕs of the schema SO1. I use as example a system similar to that in Rayo and Uzquiano (1999), which can be adapted from S as follows. First, we type the variables in L_S: we now have first-order variables $v_1, \ldots, v_n$ and second-order variables $V_1^2, \ldots, V_n^2$. We update the definition of sequential terms to allow both first and second-order variables in them. We add $V^2(v)$ and $\forall V^2 \phi$ as well-formed. Note that sequences are at most of order 2 in the language. Finally, since satisfaction applies to second-order terms, its order is 3: we add $Sat^3(x, \langle \underset{n}{\overset{1}{x}} \rangle, Y)$ as well-formed. I call the language we obtain L_{S^+}.[5]

To S we add a standard axiomatisation of second-order logic, obtaining S^+. In particular, we add impredicative comprehension principles for any formula of L_{S^+}. An interpretation and an assignment can now be defined directly in second-order logic as special kinds of second-order relations.[6] S^+ does not contain any axiom for Sat^3.

So far, nothing is stopping us from making satisfaction untyped despite the language being typed, because sequences and variables are at most of order 2 in L_{S^+}, and formulae are objects of order 1, so the following is well-formed:

$$Sat^3(\ulcorner Sat^3(v, \langle V \rangle, U) \urcorner, \langle x, X^2, Y^2 \rangle, Z^2)$$

Thus, if we assume that in S^+ we have a way to interpret a predicate P^3 of the likes of Sat^3, we could re-run the same Russellian argument as before. When we rewrite SO1 in the typed language, it now implies the following schema:

$$\exists I^2 \forall X^2 (Sat^3(\ulcorner P^3(V^2) \urcorner, \langle X^2 \rangle, I^2) \leftrightarrow \phi(X^2))$$

---

[5] Alternatively, instead of having a primitive satisfaction predicate we can explicitly define satisfaction in full third-order logic (Rayo & Uzquiano, 1999).

[6] Sequences of second-order terms like $\langle X_1^2, \ldots, X_n^2 \rangle$ are second-order and not third-order objects, as they can be defined as special kinds of second-order relations. For details, see Linnebo and Rayo (2012) for type theory and Florio and Linnebo (2021, 263–265) for plural logic.

We can easily obtain a contradiction by applying a similar reasoning as before.

This means that, to avoid the paradox, the higher-orderist should also reject self-applicability. That is, they should insist that through Sat^3 we can only interpret languages at most of order 2, without any third-order predicate like Sat^3. Semantic Openness is therefore restricted as follows:

Definition 6 (*SO1-typed*) Given any ϕ of L_{S+} *of order 2* with n free variables and any n-ary P of order 2 of the language we are interpreting:

$$\exists \mathcal{I} \forall \underset{n}{x} \left(Sat(\ulcorner P(\underset{n}{\overset{1}{v}}) \urcorner, \langle \underset{n}{\overset{1}{x}} \rangle, \mathcal{I}) \leftrightarrow \phi(\underset{n}{\overset{1}{x}}) \right)$$

What if we wish to interpret a predicate P^3 of order 3? Higher-orderists in this case resort to further typing: they will construct a new satisfaction predicate Sat^4 of order 4, which again doesn't self-apply, which interprets languages of at most order 3. We obtain a hierarchy of typed predicates Sat^n, one for every order $n \geq 3$.

2 In Defence of Self-Applicability

Can we dispense with self-applicability? First off, it should be clear from the previous section that by typing the language, higher-orderists give up on simple Semantic Openness as formalised via SO1. Instead, they adopt SO1-typed, which results from restricting ϕ to formulae of the same order as P, which is never of the same type as satisfaction. So, from the perspective of the friend of self-applicability, failure of self-applicability restricts semantic openness, and self-applicability naturally follows from generality considerations. This point can be made independently of any natural language considerations, and goes directly against the higher-orderists' claim that there is a trade-off between semantic openness and self-applicability. On the contrary, the first naturally suggests the second.

Secondly, even if there is a trade-off between semantic openness and self-applicability, it is not obvious that one should give up the second to keep the first; it might be that self-applicability is just conceptually necessary to logical validity as much as semantic openness is. Self-applicability naturally follows from the desire to define a universal theory of logical validity. The relation between a theory of logical validity and its own language is like the relation between a list and a set of rules about how to write a list (which is itself written as a list): either it talks about itself, or it is missing some lists. Ideally we are after a theory of how to write a list, not a theory of how to write some lists but not some others.

2.1 Kripkean Worries

Self-applicability is needed to make sense of the way we talk about logical validity, for some seemingly meaningful pieces of reasoning about logical validity can be expressed only if logical validity is untyped and, as we argued above, untyped

and self-applicable should be extensionally equivalent. We can adapt some examples from Kripke (1975):

(K1) Each sentence Nixon uttered during the Supreme Court case against him is logically false;

(K2) What Nixon said during the Supreme Court case is contradictory.

K1 is clearly false and, since there were inconsistencies in what he told during the trial, K2 is true. Suppose that, during the trial, Nixon uttered K3:

(K3) Everything I say during my trial is tautological.

K3 looks false and thus meaningful, because something nonsensical cannot be false or true. K3 does not change our evaluation of K1 or K2: they are still false and true, respectively. Yet, for familiar reasons, we cannot express K1, K2 and K3 unless logical validity is untyped. Nixon might not have said K3 during the trial: in that case, K3 would be perfectly meaningful and also false. So, whether K3 is meaningful or not depends on whether it is said by Nixon.

We can give another example, already noted in Nicolai and Rossi (2018):

(K4) K5 is a logical falsity;

(K5) K4 is a logical truth.

Again, K4 and K5 require logical validity to be untyped. I understand what K4 and K5 mean, and I can reason with them. Clearly, (a) if $\ulcorner \phi \urcorner$ is a logical truth, then ϕ and (b) if $\ulcorner \phi \urcorner$ is a logical falsity, then $\neg \phi$. We reason as follows:

> Suppose K5: 'K4 is a logical truth'. By (a), K4, so K5 is a logical falsity, so not K5 via (b). Contradiction. Therefore not K5, and thus K4 is not a logical truth.

Contrary to Kripke's analogous example, we cannot infer from the fact that K4 is not a logical truth that K4 is false, so no contradiction follows. In fact, we know a contradiction doesn't follow because K4-K5 can be formalised in model-theory, which is a self-applicable theory. Also, ZFC proves the principles (a) and (b) I have employed, because reflection principles ensure that a formula holds in ZFC only if it is satisfiable in a set-based model of the Von Neumann hierarchy: if ϕ then $\ulcorner \phi \urcorner$ has a (set-based) model. By contraposition and given that models make true $\ulcorner \phi \urcorner$ exactly when they don't make true $\ulcorner \neg \phi \urcorner$, this implies both (a) and (b). So, unless ZFC is inconsistent, K4 and K5 are consistent.

2.2 Expressibility and Cross Order Universality

We can press our worries about expressibility of non-applicable theories of logical validity even further. The way we form and express general theories about logical validity in philosophical logic requires self-applicability and a type-free notion of logical validity. Consider the discussion between a monist and a

pluralist about logical consequence. While it is hard to pin down exactly where the core of their disagreement lies in a way that works for everybody, it is common to describe it as follows: the monist believes that there is a 'one true logic' that is correct, in some absolute sense of the term; the pluralist does not. When spelling out the consequences of their position, monists will likely quantify over some relevant cases. A case, in turn, will likely comprise of a context and a language. Roughly, they are claiming that there is a set of rules and axioms that apply to all cases whatsoever: in any context and language. Clearly, then, their claim must also cover their present context and the language they are speaking. It would be incoherent for them to hold that classical logic is the one true logic, while at the same time refusing to admit that 'Excluded middle is valid or not' is valid.

Pluralists are in a somewhat different position: they claim that what is valid changes with the case at hand. They will want to say that excluded middle is invalid in intuitionistic cases and valid in classical cases. From this claim the status of excluded middle in their specific case cannot be settled, so the lack of connection between what the theory should say and what it can say in the absence of self-applicability is less apparent. However, their claim is still meant to be universal: part of what they are saying is that, if their case is classical then 'Excluded middle is valid or not' is valid and, if their case is intuitionistic then 'Excluded middle is valid or not' is not valid. Again, this is not expressible unless logical validity self-applies.

The monist's position bears normative consequences, too, which do not follow unless logical validity self-applies. Particular care needs to be applied when we try to bridge from a theory of logical entailment to a theory of reasoning (Harman, 1984). Yet, many think that the monist's or pluralist's position will create normative constraints on one's beliefs. I take Field (2009) as an example. If one is unhappy with Field's proposal, they can take their favourite bridge principle instead – the argument will likely go through anyway, as it does not depend on the specifics of Field's theory. Field believes that logical validity constrains an agent's overall degree of belief. To model an agent's degree of belief, we can use probability theory; degrees of belief are represented by real numbers in the [0, 1] interval (Genin & Huber, 2021). Where $cr(A)$ is the degree of belief in A, the uncertainty of A is defined as $1 - cr(A)$. Field roughly proposes the following rule, which applies well to the monist:

Definition 7 (*Belief Rule*) If B follows from $A_1, \ldots, A_n$ in the light of an agent's logic, then their degree of belief ought to be such that: $u(B)$ is less or equal to the sum of $u(A_1),\ldots, u(A_n)$.

If Maria is a monist who believes in *modus ponens*, then she ought to regulate her overall degree of belief so that it tracks the Belief rule. If she is pondering on the shoes she should wear for the day, she should not be more certain of the fact that it rains and if it rains streets are wet than she is of the fact that streets are wet. For if she is a monist about *modus ponens*, she believes that the rule holds

in any language and context, and in particular in English and in the context of her pondering about shoes. If Maria is a pluralist, then the rule might be changed to arrange for shifts in cases: given the case at hand, her degree of belief ought to adjust to what is valid in that case. If the pondering-about-shoes case is classical, then what we just said applies again; if it is not, then she should react accordingly. What happens, though, if self-applicability fails? Then whatever theory of logical validity Maria has in mind can never apply to her case, and the normative constraints that we expect the theory to have are cut off. Both the monist and the pluralist are unable to correctly infer the normative consequences their positions bear in their context, and to regulate their subjective degree of belief accordingly.

2.3 Counterpoints

I now discuss some counterpoints. First, some philosophers have recently argued that truth-like paradoxes arise for validity intended as necessary truth preservation (Whittle, 2004; Shapiro, 2010; Beall & Murzi, 2013; Murzi & Rossi, 2017). Their point seems to clash with my claim that logical validity can safely self-apply without contradictions. I don't think that what I say clashes with what they are saying, because the notion of validity these philosophers have in mind is not the one discussed by higher-orderists, who are the interlocutor of this paper. For example, Beall and Murzi (2013) develop a Curry-style paradox for a primitive notion of validity '*Val*' between two sentences. The paradox stems from the following two rules:

$$\text{If } \phi \vdash \psi \text{ then } \vdash Val(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner) \tag{VP}$$

$$\phi, Val(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner) \vdash \psi \tag{VD}$$

If the notion of logical validity we are discussing preserves truth, VD is plausible. However, VP is not plausible for logical validity. When introducing VP, Murzi and Beall admit that they are assuming that 'validity claims are appropriately "necessary", so that validity claims are themselves valid if true' (2013, 10). Yet, this iteration rule simply doesn't hold for logical validity. As Ketland (2012) notes, it might be that we conclude ψ from ϕ using arithmetical theorems, in which case it wouldn't follow that ψ *logically* follows from ϕ. He therefore proposes to restrict VP to proofs with only purely logical steps. If so, then extending PA with a validity predicate yields a conservative extension of PA, whose consistency is as problematic as that of PA. Cook (2014) reaches a similar conclusion, and argues that VP should be restricted to steps that don't apply a validity rule (2014). Nicolai and Rossi (2018) also agree that 'object-linguistic treatments of logical consequence simply do not give rise to paradox.' The system they set up is one where validity can be iterated, but they are careful to distinguish this notion of validity from logical validity, where iteration just doesn't hold. Murzi and Rossi (2017), too, discuss a notion of *naïve validity* where the paradox might arise, which they carefully distinguish from 'logical validity'.

A simple way to distinguish between the two notions appeal to Substitutivity:

Definition 8 (*Substitutivity*) For any formulae ϕ_1, ϕ_2, primitive non-logical expression ψ, and (possibly complex) expression ξ of the same logical type as ψ, if the argument from ϕ_1 to ϕ_2 is valid, then the one from $\phi_1[\psi/\xi]$ to $\phi_2[\psi/\xi]$ is valid.

As Cook (2014) notes, substitutivity fails for a notion of validity where VP holds because *Val* is a predicate, so ϕ and ψ are mentioned and not used in the statement of VP. Clearly the notion of logical validity the higher-orderists are discussing satisfies Substitutivity, therefore the higher-orderists' notion of validity is not the notion Beall and Murzi (2013) are interested in.

A second counterpoint is that, in my discussion, there is an implicit step from the fact that something is not expressible in a typed setting to the fact that something is not meaningful. This is unfair to the higher-orderist, since they would claim that this is not a matter of meaning, but of the very grammar of a suitably regimented language. The problem with this argument is that the grammar of higher-order languages is not enough to ensure that validity is typed. As we noted in Sect. 1.3, it is perfectly consistent to argue that satisfaction is untyped despite L_{S^+} being a typed language, as sequences and variables are at most of order 2 and formulae are objects of order 1. Thus, it is not the grammar that forces the higher-orderist to deny that sentences like K3–K5 are not well-formed. Rather, it is their insistence that the system doesn't self-apply, so to interpret a language of order n, interpretations must be of order n (and therefore satisfaction must be of order $n + 1$). Only with this additional assumption in place, K3–K5 become inexpressible in any language, as they try to apply a predicate of order n (the satisfaction predicate for the meta-language) to a variable of order n (the meta-variable for interpretations for the meta-language). So, that K3–K5 are not well-formed is a matter of semantics, not of syntax.

The higher-orderist might respond that, even if this is not just a grammatical point, there is no need to appeal to an untyped notion of validity to accommodate the Kripkean examples above: rather we can simply explain away how K3–K5 look meaningful, by using the distinction between character and content.[7] To make an analogy: when we deal with context-sensitivity, we can reason with an expression even though the expression itself does not have a content at the present context. For example, 'This is that' and 'This is human' imply 'That is human', even though I have made no attempt to assign a referent to 'this' or 'that', so they have no content at the present context. In Kaplanian terms, we can distinguish between the character and the content of an expression: the character is the rule we follow when we fix the semantic content of an expression given a context; the content is the extension of an expression at any given context (Kaplan, 1989). 'This' and 'This is that' are content-less at the present context, yet, they possess a character, so they are meaningful. Similarly, the higher-orderist could argue that K3–K5 are meaningful even though they are content-less at every order of the hierarchy, because they retain a character.

To make this counterpoint the higher-orderist needs to explain what they mean by character and content in the context of typed validity. However, regardless, I claim that this distinction is available to the friend of self-applicability but not to

---

[7] I thank Tim Williamson for the suggestion.

the higher-orderist. We can make an analogy with truth: Kripke (1975) would insist that the liar sentence is ungrounded, and therefore it is neither true nor false. Yet we have some semantic grasp on it, we can reason with it and we can create meaningful, complex sentences out of it. This is an advantage of Kripke's theory over a strictly typed theory where the liar sentence is not even well-formed.[8] Similarly, the friend of self-applicability can agree that K3–K5 are content-less but meaningful, and that they can be neither true nor false, pretty much like the liar sentence in Kripke's theory. Higher-orderists cannot do the same, however, because of their insistence on the fact that to interpret a language of order n interpretations must be of order n. This assumption makes K3–K5 not even well-formed, so content-less but also meaningless.

3 Resisting the Russellian Argument

The Russellian argument stems from SO1. SO1 relies on self-applicability, semantic openness and the appropriateness of T1 as a suitable formalisation of 'means that'. Here, I would like to argue that self-applicability itself is a reason to limit T1, which makes the reliance on T1 in the Russellian argument against self-applicability suspicious. One should not find this surprising, since T1 is in essence a disquotation principle, and disquotation principles are known to be problematic in type-free theories, and are often restricted to avoid inconsistencies.

T1, far from being trivial, has substantial consequences for a type-free truth predicate. Say that our meta-language is E and our theory of logical validity is self-applicable. Suppose we wish to state the obvious: for any predicate P, P in E means to P. In other words, there is a homophonic interpretation of the predicate P. In particular, 'to be satisfied' in E means what it means, i.e. to be satisfied. By substituting $P(v_n^1)$ for $Sat(v_1, \langle v_1, v_2 \rangle, v_2)$ and ϕ for $Sat(x, \langle x, y \rangle, y)$ in T1, we obtain the following iteration principle:

$$\forall x, y \big(Sat(\ulcorner Sat(v_1, \langle v_1, v_2 \rangle, v_2) \urcorner, \langle x, y \rangle, E) \leftrightarrow Sat(x, \langle x, y \rangle, y) \big) \qquad \text{(Iter)}$$

Iter states that satisfaction in E can be iterated and discharged at will. This is highly non-trivial if truth is type-free. Usually satisfaction in an interpretation commutes with negation: it is bivalent and consistent.

Definition 9 Bivalence and Consistency

$$\forall \phi, \mathcal{I} \big(\neg Sat(\ulcorner \phi \urcorner, \langle \ulcorner \phi \urcorner, \mathcal{I} \rangle, \mathcal{I}) \rightarrow Sat(\ulcorner \neg \phi \urcorner, \langle \ulcorner \phi \urcorner, \mathcal{I} \rangle, \mathcal{I}) \big) \qquad \text{(Biv)}$$

---

[8] The theory in Tarski (1933) was as such, even though it is easy to change it to make the liar well-formed but just undecidable: just expand the axioms for connectives to all sentences of the meta-language, but don't add any T-schema for sentences with the truth-predicate (Halbach, 2014: 14).

$$\forall \phi, \mathcal{I}\big(Sat(\ulcorner \neg \phi \urcorner, \langle \ulcorner \phi \urcorner, \mathcal{I} \rangle, \mathcal{I}) \to \neg Sat(\ulcorner \phi \urcorner, \langle \ulcorner \phi \urcorner, \mathcal{I} \rangle, \mathcal{I})\big) \qquad \text{(Cons)}$$

Yet, Iteration is inconsistent with Bivalence and Consistency.[9]

Lemma 1 *Iter + Biv + Cons is inconsistent.*

Proof Consider $\delta = \neg Sat(v_1, \langle v_1, v_2 \rangle, v_2)$. We reason as follows:

$$\neg Sat(\ulcorner \delta \urcorner, \langle \ulcorner \delta \urcorner, E \rangle, E)$$
$$\leftrightarrow \neg Sat(\ulcorner \neg Sat(v_1, \langle v_1, v_2 \rangle, v_2) \urcorner, \langle \ulcorner \delta \urcorner, E \rangle, E) \qquad \text{def } \delta$$
$$\leftrightarrow \neg \neg Sat(\ulcorner Sat(v_1, \langle v_1, v_2 \rangle, v_2) \urcorner, \langle \ulcorner \delta \urcorner, E \rangle, E) \qquad \text{Biv, Cons}$$
$$\leftrightarrow Sat(\ulcorner Sat(v_1, \langle v_1, v_2 \rangle, v_2) \urcorner, \langle \ulcorner \delta \urcorner, E \rangle, E) \qquad \text{Logic}$$
$$\leftrightarrow Sat(\ulcorner \delta \urcorner, \langle \ulcorner \delta \urcorner, E \rangle, E) \qquad \text{Iter}$$

The first line is inconsistent with the last. □

So, according to T1, any classic self-applicable theory of logical validity where truth means truth is inconsistent. A popular truth theory of this kind is the Revision Theory (Gupta, 1982; Gupta et al., 1993). I don't imagine that someone who defends the Revision Theory would find T1 a particularly compelling reason to abandon their theory of truth. They know that it is perfectly consistent to hold that truth is classical and type-free and deny iteration; yet, it is just a contradiction in terms to hold that a predicate P in the language I am speaking does not mean P in the language I am speaking. However, embracing T1 makes the second follow directly from the first. The obvious move for the defender of the Revision Theory is to reject T1.

There are other issues. It looks like we can coherently imagine a language where predicates mean the opposite of what they actually mean: a Mirror-*E*. As we talk in *E* about Mirror-*E*, we will say that, for any predicate P, P in Mirror-*E* means to not P. According to T1, however, there cannot be any such language. For suppose there was; then 'to be satisfied' in Mirror-*E* would mean to not be satisfied. We can then apply T1 as follows, where E^* is Mirror-*E*:

$$\forall x, y \big(Sat(\ulcorner Sat(v_1, \langle v_1, v_2 \rangle, v_2) \urcorner, \langle x, y \rangle, E^*) \leftrightarrow \neg Sat(x, \langle x, y \rangle, y)\big)$$

We obtain a contradiction by instantiating x with $\ulcorner Sat(v_1, \langle v_1, v_2 \rangle, v_2) \urcorner$ and y with E^*. This seems too easy, which again makes T1 suspicious. There seems nothing wrong with Mirror-E *per se*, let alone something that warrants such a straightforward contradiction.

We can extend the worry a bit further. If P means to ϕ in $\mathcal{I}$ then $\neg P$ means to $\neg \phi$ in $\mathcal{I}$. Whatever intuition it is fueling T1, it will likely fuel the following, similar definition:

Definition 10 (*TN1*) For any n-ary predicate P, $\neg P$ means to $\neg \phi$ in $\mathcal{I}$:

---

[9] See also Lemma 14.3 in Halbach (2014, 146).

$$\forall \overset{1}{\underset{n}{x}}\left(Sat(\ulcorner\neg P(\overset{1}{\underset{n}{v}})\urcorner, \langle\overset{1}{\underset{n}{x}}\rangle, \mathcal{I}) \leftrightarrow \neg\phi(\overset{1}{\underset{n}{x}})\right)$$

TN1 is directly inconsistent with the obvious fact that, for any predicate P, $\neg P$ means to $\neg P$ in E, which via TN1 reads:

$$\forall x, y\left(Sat(\ulcorner\neg Sat(v_1, \langle v_1, v_2\rangle, v_2)\urcorner, \langle x, y\rangle, E) \leftrightarrow \neg Sat(x, \langle x, y\rangle, y)\right)$$

We obtain an inconsistency by instantiating x with $\ulcorner\neg Sat(v_1, \langle v_1, v_2\rangle, v_2)\urcorner$ and y with E.

3.1 An Alternative Definition

If I am right, T1 is not a good definition for 'means that' if the theory at hand is self-applicable. It is not true that if P means to ϕ in $\mathcal{I}$, then P is satisfied by any x in $\mathcal{I}$ exactly when x is ϕ. At most, P means to ϕ in $\mathcal{I}$ if P is satisfied by any x in $\mathcal{I}$ exactly when x is ϕ. From the standpoint of a self-applicable theory, the biconditional in T1 offers a sufficient condition for 'means that', not a necessary one. Consequently, SO1 too is sufficient but not necessary for semantic openness, because it relies on T1.

Can we give an alternative? That is, can we provide a condition that is necessary and sufficient for interpreting P by ϕ, which is suitable if the theory self-applies? Our semantics is extensional, and terms like 'means-to' are intensional, so we have to be clear about the limitations of our endeavour. Our interpreting will generally be blind to any non-extensional differences: to interpret P to mean 'v is human' or 'v is a featherless biped' will amount to the very same thing. In virtue of the extensional nature of our interpreting, we can use the notion of satisfaction in a language to track meaning. Imagine a situation where Maria is speaking in E and wishes to teach John a foreign language like Mirror-E. One way to explain this to John is by talking in E about Mirror-E, using truth and the E-equivalent of 'is human'. Maria can explain that, for any x, 'v is human' is satisfied in Mirror-E by x exactly when x is not human. We obtain the biconditional suggested by T1. As we saw, however, using this biconditional will lead Mary to contradict herself. Maria needs to be more careful. One way to be more careful is to talk about the semantic relations between E and Mirror-E, rather than attempt to use the E-equivalent of the predicate in Mirror-E. She can say that 'v is satisfied by v in Mirror-E' means in Mirror-E the opposite of what it *actually* means in E. I argue that in this way she will not contradict herself, and yet she clearly explained to John the meaning of predicates in Mirror-E, modulo John's understanding of E. As before, we can translate this semantic relation into a biconditional using truth: Maria can say (in E) that, for any x, 'is human' is satisfied by x in Mirror-E exactly when 'is not human' is satisfied by it, in E. We can generalise the test as follows:

Definition 11 (*T2*) Where E is the language we are speaking, an n-ary predicate P means to ϕ in an interpretation $\mathcal{I}$:

$$\forall^1_{n}x \left(Sat(\ulcorner P(\overset{1}{\underset{n}{v}})\urcorner, \langle \overset{1}{\underset{n}{x}} \rangle, \mathcal{I}) \leftrightarrow Sat(\ulcorner \phi(\overset{1}{\underset{n}{v}})\urcorner, \langle \overset{1}{\underset{n}{x}} \rangle, E) \right)$$

The test presupposes that the theory is able to specify an untyped theory of truth, otherwise a concept of truth of x in E would not be defined in E. From the perspective of T2, T1 suppresses the step from the sentence being true in E to its content being the case, which is harmless if the formula does not talk about interpreting and truth, but might become paradoxical if it does.

Semantic Openness through T2 looks like this:

Definition 12 (*SO2*) Given any ϕ of L_S with n free variables and any n-ary P of the language we are interpreting:

$$\exists \mathcal{I} \forall^1_{n}x \left(Sat(\ulcorner P(\overset{1}{\underset{n}{v}})\urcorner, \langle \overset{1}{\underset{n}{x}} \rangle, \mathcal{I}) \leftrightarrow Sat(\ulcorner \phi(\overset{1}{\underset{n}{v}})\urcorner, \langle \overset{1}{\underset{n}{x}} \rangle, E) \right)$$

SO2 is not obviously inconsistent, unlike SO1, and the Russellian instance is now harmless. There is no issue in interpreting P as '$\ulcorner P(v)\urcorner$ is not satisfied by u in u'. All we obtain is that, in such an interpretation $\mathcal{I}$, $\ulcorner P(v)\urcorner$ is satisfied by $\mathcal{I}$ in $\mathcal{I}$ exactly when $\ulcorner \neg Sat(\ulcorner P(v)\urcorner, \langle u \rangle, u)\urcorner$ is satisfied in E by $\mathcal{I}$. The Russellian argument doesn't go through unless a T-schema is available to discharge *Sat* in E, which is again unlikely because satisfaction is type-free.

T2 does not presuppose Iteration. In fact via T2, that any predicate P in E means to P is the obvious truism:

$$\forall^1_{n}x \left(Sat(\ulcorner P(\overset{1}{\underset{n}{v}})\urcorner, \langle \overset{1}{\underset{n}{x}} \rangle, E) \leftrightarrow Sat(\ulcorner P(\overset{1}{\underset{n}{v}})\urcorner, \langle \overset{1}{\underset{n}{x}} \rangle, E) \right)$$

Mirror-E now does not trivially entail a contradiction. All we obtain through T2 is the following, where $\psi = Sat(v_1, \langle v_1, v_2 \rangle, v_2)$:

$$Sat\left(\ulcorner \psi \urcorner, \langle \ulcorner \psi \urcorner, E^* \rangle, E^* \right) \leftrightarrow Sat\left(\ulcorner \neg \psi \urcorner, \langle \ulcorner \psi \urcorner, E^* \rangle, E \right)$$

This is inconsistent if a T-schema is available for ψ, which is unlikely precisely because *Sat* is type-free and therefore the T-schema needs to be restricted, accordingly for sentences where *Sat* occurs.

TN2 is not inconsistent, either, unlike TN1.

Definition 13 (*TN2*) For any n-ary predicate P, $\neg P$ means to $\neg \phi$ in $\mathcal{I}$:

$$\forall^1_{n}x \left(Sat(\ulcorner \neg P(\overset{1}{\underset{n}{v}})\urcorner, \langle \overset{1}{\underset{n}{x}} \rangle, \mathcal{I}) \leftrightarrow Sat(\ulcorner \neg \phi(\overset{1}{\underset{n}{v}})\urcorner, \langle \overset{1}{\underset{n}{x}} \rangle, E) \right)$$

Again, that $\neg Sat$ means $\neg Sat$ in E is just an obvious truism, via TN2.

3.2 Typed Bias

From the perspective of a self-applicable theory where T2 holds, we can show that T1 presupposed the availability of an unrestricted T-schema. That is, T1 has an implicit 'typed bias' built into it. To see this, let's first set the following, common definition, where ϕ is a sentence, i.e. a closed formula:

Definition 14 (*Tr*)

$$Tr(\ulcorner\phi\urcorner) \leftrightarrow_{df} \forall x Sat(\ulcorner\phi\urcorner, \langle x \rangle, E)$$

For the lemma, we only need two very weak assumptions: that *Tr* commutes with conjunction and that *Sat* in E satisfies a T-schema for self-identity statements:

$$\forall\phi, \psi \left(Tr(\ulcorner\phi \wedge \psi\urcorner) \leftrightarrow \left(Tr(\ulcorner\phi\urcorner) \wedge Tr(\ulcorner\psi\urcorner) \right) \right) \quad\quad (A^\wedge)$$

$$\forall x \left(x = x \leftrightarrow Sat(\ulcorner v = v\urcorner, \langle x \rangle, E) \right) \quad\quad (A^=)$$

Since $x = x$ is a theorem, from $A^=$ it follows that $\forall x Sat(\ulcorner v = v\urcorner, \langle x \rangle, E)$. We make extensive use of this in the following proof.

Lemma 2 (Typed bias) *Modulo T1 and T2, $S + A^= + A^\wedge$ implies the following unrestricted T-schema, where ϕ is any closed formula of S*:

$$Tr(\ulcorner\phi\urcorner) \leftrightarrow \phi$$

Proof We prove $\phi \vdash Tr(\ulcorner\phi\urcorner)$ and vice-versa $Tr(\ulcorner\phi\urcorner) \vdash \phi$. The lemma follows by the deduction theorem. The deduction theorem holds in standard higher-order proof systems of the likes higher-orderists assume, so its use is warranted here.

Suppose $Tr(\ulcorner\phi\urcorner)$. We reason as follows:

$$Tr(\ulcorner\phi\urcorner) \quad\quad (1)$$

$$\forall x Sat(\ulcorner\phi\urcorner, \langle x \rangle, E) \quad\quad (2)$$

$$\forall x \left(Sat(\ulcorner\phi\urcorner, \langle x \rangle, E) \wedge Sat(\ulcorner v = v\urcorner, \langle x \rangle, E) \right) \quad\quad (3)$$

$$\forall x \left(Sat(\ulcorner v = v\urcorner, \langle x \rangle, E) \leftrightarrow Sat(\ulcorner\phi \wedge v = v\urcorner, \langle x \rangle, E) \right) \quad\quad (4)$$

$$\ulcorner v = v\urcorner \text{ means to } \phi \wedge x = x \text{ in } E \quad\quad (5)$$

$$\forall x \left(Sat(\ulcorner v = v\urcorner, \langle x \rangle, E) \leftrightarrow (\phi \wedge x = x) \right) \quad\quad (6)$$

$$\phi \quad\quad (7)$$

(2) follows by definition of *Tr*. (3) follows since $\forall x Sat(\ulcorner v = v \urcorner, \langle x \rangle, E)$ by $A^=$. (4) follows by logic and by $A^\wedge$. (4) is an instance of T2, and it lets us conclude that (5): $v = v$ means being such that ϕ and $x = x$ in E. Note that being such that ϕ and $x = x$ is a perfectly legitimate reinterpretation of $x = x$, since it is a formula with x free.[10] (5) is a meta-conclusion, as it were, which is not carried out directly in S. Since T1 is in place as well, from (5) we can infer (6), which is just the instance of T1 when $v = v$ in E means being such that ϕ and $x = x$. (7) follows from (6) by $A^=$ and logic. The proof also works in reverse, with the step from (6) to (4) being that, via T1, $v = v$ in E means being such that ϕ and $x = x$, so we can apply T2. $\square$

Lemma 2 shows that, *modulo* the acceptance of T2 and some very minimal assumptions about the power of the semantic system, there is a direct way to link the discussion on T1 and the axioms of 'means-to' to the discussion on truth. The strong reasons we have for embracing a type-free theory of truth work against T1, since T1 forces a truth-predicate that satisfies an unrestricted T-schema. This would make the type-free theory of truth immediately inconsistent via Tarski's theorem, assuming that the theory itself contains some very basic arithmetic.

Since T1 has a typed bias and it is incompatible with any consistent theory of type-free truth, self-applicability itself is a strong reason to limit T1. How much should we limit it? The details are likely to depend on the theory of satisfaction in an interpretation in which we are working. However, any type-free theory of truth will at least verify a T-schema restricted to formulae where *Sat* does not occur. Putting T2 and this T-schema together, we obtain the following restricted version of T1:

Definition 15 (*T1**) When *Sat* does not occur in ϕ, an *n*-ary predicate P means to ϕ in an interpretation $\mathcal{I}$:

$$\forall \underset{n}{\overset{1}{x}} \left(Sat(\ulcorner P(\underset{n}{\overset{1}{v}}) \urcorner, \langle \underset{n}{\overset{1}{x}} \rangle, \mathcal{I}) \leftrightarrow \phi(\underset{n}{\overset{1}{x}}) \right)$$

A problem with T2 is that it does not provide 'worldly conditions' for when P is satisfied by some objects in an interpretation, it just ties the satisfaction of P in an interpretation to the satisfaction of some formula in our language. Instead of connecting language and world, it just refers you back to your language. Via T1*, we can be reassured that, in a very wide range of cases, we can trace the linguistic condition back to a 'worldly condition'. In fact, assume we have a typed theory T_1 of validity for a language L. T1 will hold for all ϕ without the *Sat* predicate. If we extend T_1 to a type-free theory T_2 where *Sat* is now untyped, we obtain T1* which will apply to exactly the same ϕs T1 applied to in T_1, so no generality is lost in terms of worldly-conditions by going from T_1 to T_2.

To conclude, advocates of an untyped theory of truth often point out the advantages of their theory over the classic Tarskian typed hierarchy. With an untyped

---

[10] If we assume that 'being such that ϕ' where ϕ is a sentence is a legitimate interpretation of a predicate, then the proof can be shortened and $A^\wedge$ can be dropped.

notion of truth at hand, we are able to reflect on the system we are using, and make generalisations that are impossible to state within a typed theory of truth. The same is true for logical validity: untyped logical validity is necessary to state generalisations across languages, generalisations that are essential to our discussions in philosophical logic. To suit our theoretical needs, we simply need self-applicability. I made the case that self-applicable validity might be possible, if we are careful not to inject into the system assumptions that presuppose that logical validity or truth are typed.

Declarations

References

Beall, J., & Murzi, J. (2013). Two flavors of Curry's paradox. *The Journal of Philosophy, 110*(3), 143–165.

Cook, R. T. (2014). There is no paradox of logical validity. *Logica Universalis, 8*, 447–467.

Etchemendy, J. (1990). *The concept of logical consequence*. CSLI Publications.

Etchemendy, J. (2008). Reflections on consequence. In D. Patterson (Ed.), *New essays on Tarski and philosophy* (pp. 263–299). Oxford: Oxford University Press.

Field, H. (2009). I—What is the normative role of logic? In *Aristotelian Society Supplementary Volume*, Vol. 83.1 (pp. 251–268). Wiley.

Florio, S. (2014). Untyped pluralism. *Mind, 123*(490), 317–337.

Florio, S., & Linnebo, Ø. (2021). *The many and the one: A philosophical study of plural logic*. Oxford University Press.

Genin, K., & Huber, F. (2021). Formal representations of belief. In Zalta, E. N. (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2021 ed.). Stanford University.

Gupta, A. (1982). Truth and paradox. *Journal of Philosophical Logic, 11*(1), 1–60.

Gupta, A., Gupta, A. K., Belnap, N. D., Belnap, N., et al. (1993). *The revision theory of truth*. Mit Press.

Hájek, P., & Pudlák, P. (2016). *Metamathematics of first-order arithmetic*. Cambridge University Press.

Halbach, V. (2014). *Axiomatic theories of truth*. Cambridge University Press.

Harman, G. (1984). Logic and reasoning. In H. Leblanc, E. Mendelson, & A. Orenstein (Eds.), *Foundations: Logic, language, and mathematics* (pp. 107–127). Springer.

Kaplan, D. (1989). Demonstratives: An essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). Oxford University Press.

Ketland, J. (2012). Validity as a primitive. *Analysis, 72*(3), 421–430.

Kreisel, G. (1967). Informal rigour and completeness proofs. In I. Lakatos (Ed.), *Studies in logic and the foundations of mathematics* (Vol. 47, pp. 138–186). Elsevier.

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

Levy, A., et al. (1960). Axiom schemata of strong infinity in axiomatic set theory. *Pacific Journal of Mathematics, 10*(1), 223–238.

Linnebo, Ø. (2006). *Sets, properties, and unrestricted quantification*. In A. Rayo, & G. Uzquiano (Eds.), *Absolute Generality* (pp. 149–178). Clarendon Press.

Linnebo, Ø., & Rayo, A. (2012). Hierarchies ontological and ideological. *Mind, 121*(482), 269–308.

Montague, R. (1961). Fraenkel's addition to the axioms of zermelo. In Y. Bar-Hillel, E. I. J. Poznanski, M. O. Rabin, & A. Robinson (Eds.), *Essays on the foundations of mathematics* (pp. 91–114). Magnes Press.

Murzi, J., & Rossi, L. (2017). Naïve validity. *Synthese*, pp. 1–23.

Nicolai, C., & Rossi, L. (2018). Principles for object-linguistic consequence: From logical to irreflexive. *Journal of Philosophical Logic, 47*, 549–577.

Rayo, A. (2006). Beyond plurals. In A. Rayo & G. Uzquiano (Eds.), *Absolute generality* (pp. 220–254). Oxford University Press.

Rayo, A., & Uzquiano, G. (1999). Toward a theory of second-order consequence. *Notre Dame Journal of Formal Logic, 40*(3), 315–325.

Rayo, A., & Williamson, T. (2003). A completeness theorem for unrestricted first-order languages. In J. Beall (Ed.), *Liars and heaps: New essays on paradox* (pp. 331–356). University of Hawaii Press.

Shapiro, L. (2010). Deflating logical consequence. *The Philosophical Quarterly, 61*(243), 320–342.

Shapiro, S. (1987). Principles of reflection and second-order logic. *Journal of Philosophical Logic, 16*(3), 309–333.

Studd, J. (2019). *Everything, more or less: A defence of generality relativism*. Oxford University Press.

Tarski, A. (1935). Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica*, *1*, 261–405. Translated by J. H. Woodger as "The concept of truth in formalized languages. In J. Corcoran (Ed.), *Logic, Semantics, Metamathematics*, 1983, Indianapolis: Hackett, second edn.

Whittle, B. (2004). Dialetheism, logical consequence and hierarchy. *Analysis, 64*(4), 318–326.

Williamson, T. (2003). Everything. *Philosophical Perspectives, 17*(1), 415–465.