How to adopt a logic

Daniel Cohnitz[1] & Carlo Nicolai[2]

[1]Utrecht University

[2]King's College London

Abstract

What is commonly referred to as the *Adoption Problem* is a challenge to the idea that the principles for logic can be rationally revised. The argument is based on a reconstruction of unpublished work by Saul Kripke. As the reconstruction has it, Kripke essentially extends the scope of William van Orman Quine's regress argument against conventionalism to the possibility of adopting new logical principles. In this paper we want to discuss the scope of this challenge. Are all revisions of logic subject to the regress problem? If not, are there interesting cases of logical revisions that are subject to the regress problem? We will argue that both questions should be answered negatively.

1 Introduction

What is commonly referred to as the *Adoption Problem* is a challenge to the idea that the principles for logic can be rationally revised. The argument is based on a reconstruction of unpublished work by Saul Kripke.[1] As the reconstruction has it, Kripke essentially extends the scope of William van Orman Quine's regress argument (Quine, 1976) against conventionalism to the possibility of adopting new logical principles. In this paper we want to discuss the scope of this challenge. Are all revisions of logic subject to the regress problem? If not, are there interesting cases of logical revisions that are subject to the regress problem? We will argue that both questions should be answered negatively. Kripke's regress does not arise for all rules of inference and not even for the adoption of those rules that are of relevance for the discussion of the rational revisability of logic.

We will begin the paper in section 2 with a rather brief summary of the use that Quine made of the regress argument against a conventionalist conception of logic and sketch Quine's own view on the revisability of logic. Kripke seems to claim that the point that Quine makes against conventionalism should equally apply to Quine's own view on the rational revisability of logic. In section 3 we will look at which logical principles are at all subject to a potential regress problem and we will discuss whether the principles that are potentially subject to a regress problem are principles that are of relevance for the discussion of the rational revisability of logic. Our arguments in section 3 will thereby follow the specific setup that Kripke introduced for the discussion of the regress problem. In section 4 we will look at actual cases of proposed logical revisions in order to show how the more abstract considerations of the previous sections may apply to "real life" cases.

Since we arrive at a largely negative evaluation of Kripke's argument, we will close the paper in section 5 by considering alternative targets for Kripke's argument. Perhaps Kripke doesn't primarily target Quine's view on the revisability of logic (as the Kripke

[1]See Padro (2015) and Devitt (2018).

scholars Padro and Devitt have it) but Quine's view on logic in general. However, as we will argue in that section, also for these alternative targets Kripke's regress argument doesn't pose a real challenge.

2 The Adoption Problem

According to Padro (2015), Kripke uses the following example to illustrate the problem of adoption:

Ravens

> Let's try to think of someone – and let's forget any questions about whether he can really understand the concept of "all" and so on – who somehow just doesn't see that from a universal statement each instance follows. But he is quite willing to accept my authority on these issues – at least, to try out or adopt or use provisionally any hypotheses that I give him. So I say to him, 'Consider the hypothesis that from each universal statement, each instance follows.' Now, previously to being told this, he believed it when I said that all ravens are black because I told him that too. But he was unable to infer that this raven, which is locked in a dark room, and he can't see it, is therefore black. And in fact, he doesn't see that that follows, or he doesn't see that that is actually true. So I say to him, 'Oh, you don't see that? Well, let me tell you, from every universal statement each instance follows.' He will say, 'Okay, yes. I believe you.' Now I say to him, ' "All ravens are black" is a universal statement, and "This raven is black" is an instance. Yes?' 'Yes,' he agrees. So I say, 'Since all universal statements imply their instances, this particular universal statement, that all ravens are black, implies this particular instance.' He responds: 'Well, Hmm, I'm not entirely sure. I don't really think that I've got to accept that.' (Padro, 2015, fn. 49)

2.1 Quine against conventionalism

Lewis Carroll's similar dialogue between a tortoise and Achilles has famously been used by Quine (1976) in order to show that the logical positivists'[2] conventionalism about logic

[2] Who the target of Quine's paper 'Truth by convention' eventually is, is not clear. Quine doesn't explicitly say that it is Carnap and there are reasons to think he targeted his own view (Ebbs (2011)) and that of C.I. Lewis (Morris (ming)).

is in trouble. Conventionalism about logic (of the kind that Quine considers) explains why logic should have a special status: Logical principles are knowable *a priori* and necessarily true. According to conventionalism, we decide to maintain the statements of logic "independently of our observations of the world" and thus assign them a truth-value by convention. This accounts for their epistemic and modal status.

Although Quine expresses considerable sympathy for the view (granting that it is "perhaps neither empty nor uninteresting nor false"), he nevertheless sees it facing a difficulty that he summarizes as follows:

> Each of these conventions [Quine refers here to the schematic axioms of propositional logic] is general, announcing the truth of every one of an infinity of statements conforming to a certain description; derivation of the truth of any specific statement from the general convention thus requires a logical inference, and this involves us in an infinite regress. (Quine, 1976, 103)

In Carroll's dialogue, the tortoise challenges Achilles to get it to infer in accordance with Modus Ponens. Achilles fails to achieve this even though the tortoise is ready to accept Modus Ponens as a true principle. For Quine, the upshot of that dialogue is that logic can't be based on convention alone, since it seems that we need to have the ability to apply the supposed conventions and derive consequences from them in order to follow them. But then logic must be prior to such conventions (rather than the other way around).

> In a word, the difficulty is that if logic is to proceed *mediately* from conventions, logic is needed for inferring logic from the conventions. (Quine, 1976, 104)

Quine does see a way for the conventionalist to address this difficulty. What if we can adopt a convention "through behaviour" (Quine, 1976, 105) instead of adopting it via explicitly announcing it first? Perhaps the explicit formulation of these conventions can come later, once we have language and logic and all that at our disposal. For Quine this is a live option, but not one that he is still willing to describe as logic being based on "convention". From Quine's behaviorist point of view, behavior that adopts a rule is indistinguishable[3] from behavior that displays firmly held beliefs. Since the label 'convention' is then without explanatory power, we can drop it from our account of logic.[4]

2.2 Kripke against Quine

As Padro (2015) explains, Kripke now turns the regress argument against Quine himself. Quine had famously suggested in 'Two dogmas of empiricism' (Quine, 1953) that not even logic is immune to revision. Empirico-pragmatic considerations may lead us to the adoption of a new logic. A view that is, of course, quite compatible with the idea that logic is nothing but firmly held belief in the first place. Perhaps – so Quine's own example – we may decide to a adopt a logic that drops the principle of excluded middle because it may help to simplify quantum mechanics (Quine, 1953). However, Kripke seems to believe that Quine's picture, viz. that we can treat principles of logic just like any other empirical hypothesis, is prone to the exact same objection that Quine mounted against conventionalism. Padro cites Kripke as follows:

> . . . the Carnapian tradition about logic maintained that one can adopt any kind of laws for the logical connectives that one pleases. This is a principle of tolerance, only some kind of scientific utility should make you prefer one to the other, but one is completely free to choose. Of course, a choice of a different logic is a choice of a different language form.
>
> Now, here we already have the notion of adopting a logic, which is what I directed my remarks against last time. As I said, I don't think you can adopt a logic. Quine also criticizes this point of view and for the very same reason

[3] In fact, Quine only makes the much weaker observation that it would be "difficult to distinguish" a behavioral adoption of conventions from behavior that displays firmly held beliefs.

[4] See Azzouni (2014) and Cohnitz and Estrada-González (2019) for a discussion of conventionalism and Quinean arguments against it. Thanks to the work of David Lewis and others we now have a much clearer idea of how behavior that is based on firmly held belief can be distinguished from behavior that is guided by an implicitly adopted convention.

> I did. He said, as against Carnap and this kind of view, that one can't adopt a logic because if one tries and sets up the conventions for how one is going to operate, one needs already to use logic to deduce any consequences from the conventions, even to understand what these alleged conventions mean.
>
> This is all very familiar as a criticism of Carnap. Somehow people haven't realized how deep this kind of issue cuts. It seems to me, as I said last time, obviously to go just as strongly against Quine's own statements that logical laws are just hypotheses within the system which we accept just like any other laws, because then, too, how is one going to deduce anything from them? I cannot for the life of me, see how he criticizes this earlier view and then presents an alternative which seems to me to be subject to exactly the same difficulty. (Padro, 2015, 113)

Padro and Devitt interpret Kripke as targeting in particular Quine's idea that logic is revisable and that we can adopt a *new* logic. We will follow their reconstruction (but will discuss in the last section of this paper whether that is the best interpretation of Kripke's attack on Quine). According to this reconstruction of the argument, logic is not only not based on convention, but logic can't be rationally revised either, because whatever empirico-pragmatic reasons we may have for preferring some alternative logic, we can't adopt a new logic. Presumably the argument is then that the adoption of a new rule (as in Kripke's example) would already presuppose the logical competence that allows us to apply the rule. However, as in Kripke's example, if that competence is in fact the very rule we are supposed to adopt, then this can't work.

A *prima facie* reasonable reaction to the argument – due to Michael Devitt (2018), for instance – is to distinguish the way in which we come to know the propositional form of a logical rule, its representation, such as 'from a universal statement, each instance follows', and the way in which a person/agent can come to be *governed* by a rule, a state that may not necessarily require a representational form of the rule. The first kind of knowledge may be dubbed *declarative*, the second *procedural*. According to this first reaction, therefore, the sort of revision involved in Carroll's example concerns the fact that declarative knowledge of a rule alone may not be sufficient to rationally revise one's logical beliefs. But this does not rule out the possibility of training someone in acquiring procedural knowledge of a new rule.

A similar position is assumed by Graham Priest (2014), although framed in his distinction between the *logica docens*, *utens*, and *ens*. The logic we teach (*docens*) can be revised by means of a broadly abductive methodology. What is commonly called a 'logic', for Priest, should in fact better be seen as a 'logical theory', namely a substantial body of knowledge concerning some notion of logical consequence. Now a logical theory can be rationally revised in the same way as other scientific theories can be revised, namely by comparing it with alternatives according to theory-choice criteria such as explanatory power, strength, adequacy to data, unifying power, and whatever else these may be. The logical theory we teach, therefore, can be rationally revised, and so can the logical theory we *use*. How? Simply by training students in a chosen *logica docens*. To connect Priest's approach to rational revisability of logic with the Carroll-Kripke example, what seems to be clear is that for Priest the process of acquisition of a rule is not a local procedure, but rather a global process of acceptance of a logical theory that goes well beyond the rules of a formal system. This point will be further expanded in §4.

In the next three section we leave aside these attempts to undermine the Adoption Problem by denying a significant role to the declarative knowledge of a rule. We assume that the declarative knowledge of a rule does indeed play a role in one's actual adoption and consider in more detail how such process could actually work. As it will turn out in section 3 and 4, there is no problem of adoption that would arise for the *revision* of logic (as Kripke seems to claim). It is true that one needs *some* logical rules in order to be able to adopt and apply new ones, but in pretty much all cases in which one has already a logic, these rules will be available. The adoption problem – as a problem for a Quinean revisionist about logic – seems to be just a pseudo-problem.

2.3 Logica Utens

Although we will set aside Priest's solution to the problem of adoption, it will still be useful for our discussion to help ourselves to a distinction between *logica docens* and *logica utens*. The former is an explicit theory that may or may not be formalized in precise mathematical terms. For all we know, Aristotle started the business of developing a *logica docens*.

A *logica utens*, on the other hand, is – in our terminology – the logic that an individual reasons with under idealized circumstances. More concretely, the *logica utens* is constituted by an individual's dispositions to accept or reject inferences as "logically valid" (under favourable circumstances), including their dispositions to correct their judgments about the matter after reflection or after having received additional information. The *logica utens* is thus not simply all of a person's inferential behavior, including everything that they themselves would (if they gave the matter more thought) recognize as fallacies. Nor is it a person's pre-scientific, unschooled explicit theory of logic (as Peirce sometimes seems to imply in his usage of the Latin terminology). The *logica utens* may in fact be intransparent (to some degree) to the reasoner.

We use *logica utens* in order to refer to a set of dispositions of an individual, but yet assume that these dispositions have the right kind of normative force that would allow us to say that a reasoner can fail to reason in accordance with her own standards. We thereby set the so-called rule-following problem aside. Kripke(nstein) would perhaps not be happy with this, but we take it that the adoption problem can be discussed independently of Kripke's scepticism about meaning and the possibility of a private language/logic (but we will come back to this point in section 6). While Aristotle is widely credited with having started the business of developing a *logica docens*, *homo sapiens* much earlier started to develop a *logica utens*.

3 Patterns of adoption

3.1 What can we adopt?

As noticed already in Cohnitz and Estrada-González (2019), when one looks carefully at the Carroll-Kripke example, it becomes clear that not all rules are equally problematic. Consider the following version of our original dialogue in which universal instantiation is now replaced by the introduction of the existential quantifier. It involves subjects A and B and we assume, for the sake of the argument, that B is not able to perform inferences according to *Existential Introduction.* As before, we assume that B is willing to cooperate in accepting and reasoning according to the hypotheses that A provides.

A. Consider the hypothesis that, if some predicate φ holds of an individual t, then there is at least one individual that satisfies φ.

B. OK, I am considering it.

A. This piece of paper is white, isn't it?

B. Yes.

A. Now 'this piece of paper is white' is telling us that the predicate 'is white' applies to this piece of paper, therefore since if some predicate φ holds of an individual t, then there is at least one individual that satisfies φ, so *there is something that is white.*

B. Sure, thanks!

In the above dialogue, unlike what happens in the Kripke case, nothing prevents B from following and accepting A's instructions. The reason is that *no prior* understanding of Existential Introduction is needed for B to follow the instructions given by A.

However, there is something else that *needs* to be presupposed by B. First of all she needs the ability of inferring via Modus Ponens. This is the lesson we learnt from

Carroll's example. Moreover, in the light of Kripke's example, it would prima facie seem that also Universal Instantiation is required. However, both in Kripke's example and here we need much less than the Universal Instantiation in full generality. Consider A's last sentence: it presupposes the capability of recognizing the validity of the step that goes from an argument of the form $\langle \varphi(t/v), \exists v\varphi \rangle$, *for all* φ, to an argument of the form $\langle P(t/v), \exists vP \rangle$ for a particular P. Similarly, in Kripke's example, the step that prevents the receiver of the instructions from agreeing on the desired conclusion is her incapability of recognizing the validity of the inference from an argument of the form $\langle \forall v\varphi, \varphi(t/v) \rangle$ to one of the form $\langle \forall vP, P(t/v) \rangle$. In both cases, it is a form of universal instantiation that is at stake. But at a closer look, the inferences under considerations are in fact of the form:

(SCS) for any *formula* φ, if $\Phi(\varphi)$, then $\Phi(P/\varphi)$, for some fixed argument pattern Φ.

(SCS) is a very distinguished form of Universal Instantiation. In the first place the quantifiers range over a fixed *set*, more specifically a set of formulas of the language. Under the natural assumption that the languages that we speak are countable, the size of such set is then no greater than $\aleph_0$, whereas no such assumption is required for the general form of Universal Instantiation. Moreover, (SCS) has a form that is well-known to logicians: it is a schematic substitution rule, according to which, by accepting the schema, one accepts all its specific instances in the language under consideration.

This discussion can be generalized by formulating a more abstract recipe for adoption contained in the box below.

Of course the extent to which (SCS) is a logical rule can be debated at length: it can

[5]The reader might worry that arriving at the antecedent of the displayed conditional requires an additional rule, introduction of the conjunction in particular. However, in many logics, including classical logic, this can be reformulated as a series of nested conditionals:

$$\Phi_1(\vec{A};\vec{t}) \to (\Phi_2(\vec{A};\vec{t}) \to (\ldots \to (\Phi_k(\vec{A};\vec{t}) \to \Psi(\vec{A};\vec{t}))\ldots)$$

pattern for adoption:

1. One starts with a schematic logical principle of the form

   (1) if $\Phi_1(\vec{X};\vec{z})$ and ... and $\Phi_k(\vec{X};\vec{z})$, then $\Psi(\vec{X};\vec{z})$,

   with $\vec{X}$ and $\vec{z}$ possibly empty strings of variables of finite length. Here the X_i's are one sort of variables to be replaced with formulas, and the z_j's are meta-variables for terms possibly including a different sort of variables for objects. Some machinery for renaming variables, if needed, is also assumed.

2. One is then given a schematic instance of the antecedent of the conditional

   $$\Phi_1(\vec{A};\vec{t}) \text{ and} \ldots \text{and } \Phi_k(\vec{A};\vec{t})$$

   for $\vec{A}$ formulas of the language and $\vec{t}$ actual terms in the language.[5]

3. (scs) enables one to go from 1 to

   $$\text{if } \Phi_1(\vec{A};\vec{t}) \text{ and} \ldots \text{and } \Phi_k(\vec{A};\vec{t}), \text{ then } \Psi(\vec{A};\vec{t}),$$

4. by Modus Ponens applied to 2 and 3, one concludes $\Psi(\vec{A};\vec{t})$, thereby inferring according to (1).

even be argued that it is *the* logical rule, as it is possible to axiomatize, say, classical logic, by resorting to axioms involving specific predicate letters – and not axiom schemata or rule schemata – and some principle akin to (scs). For our concerns, however, what matters is that the form of universal instantiation that Kripke suggests is presupposed by our capability of acquiring Universal Instantiation is not as strong. Rather, it is a very specific form of universal instantiation that has much to do with our ability of recognizing and combining syntactic patterns.

The problems encountered with the adoption of a logical rule – as far as Kripke's example is concerned – boil down, therefore, to the necessity of certain presuppositions to the process, in particular the presuppositions of the validity of Modus Ponens and the validity of the very specific form of universal instantiation (SCS).

3.2 Where can we adopt?

In general, revisions can reasonably involve either (i) dropping some principle from the set of one's logical beliefs, or (ii) adding principles to it.[6] We call the former process DROP, and the latter ADD.

Most cases of proposed logical revision at the heart of modern and contemporary debates involve DROP. Starting with classical reasoning, intuitionists proposed to drop the law of excluded middle or, equivalently, to weaken one of the rules for negation. Paracomplete and paraconsistent logicians also propose to drop one of the rules for negation, although their weakening of classical negation is more severe than the one proposed by the intuitionists. Some subtler proposals are also possible. Supervaluationists, for instance, agree with all inferences of classical logic of the form $\langle \Gamma, \varphi \rangle$, but disagree on inferences with multiple conclusions.[7]

But if one focuses on DROP, it seems clear that there are no major problems for the adoption of a new rule. If one is in fact already able to infer by means of a rule, it is always possible to adopt *restrictions* of the rule without falling prey of the examples considered above. One might see paraconsistent logic, for instance, as resulting from classical logic via the restriction of Modus Ponens to formulas that are not truth value gluts. Faced with the Carroll's story, the 'adoption' of restricted Modus Ponens for the paraconsistent logician would not pose any problem.

What about ADD? Let us consider different scenarios here. Revision *upwards*, so

[6] Of course it is possible that the proposed adoption in question leads from a set of logical beliefs to another which is inconsistent with the previous one, but in the reasonable cases in which this happens one can always describe this process as the result of first dropping some rule and then adding to the remaining principles some other principles.

[7] For instance, they drop the classical inference $\langle \{\varphi \vee \neg\varphi\}, \{\varphi, \neg\varphi\} \rangle$.

to speak, may involve different starting points. Consider the representation in figure 1. Let's assume that we can order the logics under consideration from weak to strong.

$$\text{weak} \;\text{———}\; \bullet \longrightarrow \textit{classical logic} \longrightarrow \bullet \;\text{———}\; \text{strong}$$

Let the arrows represent the direction of revision. The first arrow on the left represents a revision that takes a subclassical logic as its starting point and revises "upwards" in the direction of (or to) classical logic. The second arrow represents the case of upwards revision that takes classical logic as a starting point.

Prima facie there are good reasons to doubt the significance of ADD, if one assumes that the process of adoption has classical logic as its starting point and restricts oneself to the propositional case. The Post completeness of classical propositional logic tells us that the only consequence relation that properly extends it is the trivial one.

On the other hand, when we do not restrict ourselves to the propositional case and consider first-order logic, which isn't Post-complete, we also know that Modus Ponens and Universal Instantiation are already in place. But then revision that follows our schema for adoption is also unproblematic – new rules can be learned and applied since they can be brought in conditional form. For instance, we might consider a higher-order version of the rule of existential introduction:

(2) from $\varphi(R)$, infer $\exists X \varphi(X)$

with R a set variable which is free for X in φ. As before, the adoption of such rule would require the capability of applying (SCS). In the specific case of (2), the schematic variable needs to be of a suitable type; it should be capable of taking variables like X as arguments. This process, however, is still carried out once a suitable language is fixed. The substitution involved in the adoption of (2) does not require any substantial decision on the semantic status of the different types of variables.

This leaves us with upwards revision where some subclassical logic is our starting point. Here the only problematic candidates seem to be those that either don't have Modus Ponens or do not have (scs). A logic without Modus Ponens is difficult to conceive of. True, there are logics, e.g. some paraconsistent logics, that do not have Modus Ponens, but this is usually seen as a major problem for these systems that puts their very logicality into doubt.

What about (scs)? It is a common assumption in much of contemporary semantics that natural languages must (in some way, (Cohnitz, 2005)) be *compositional*. How else could it be explained that we can use and understand new sentences with novel meanings? However, compositionality requires some form of systematic syntactic decomposition and of keeping track of how, for example, argument places of predicates are filled. It is hard to see why such capacity shouldn't already be sufficient for the kind of schematic substitution that Kripke's example requires. Compositionality by itself guarantees that competence with a sentence like 'Sam kisses Martin' entails competence with 'Martin kisses Sam', 'Reinold kisses Julie' – this fact is behind the systematicity argument for compositionality (Szabó, 2000). But then the basic skills involved in processing a compositional language (treating linguistic items as schematic and (re)combinable with other linguistic items of certain syntactic categories) already allow one to reason in accordance with (scs). This skill doesn't seem to be in need of "adoption".[8]

For our purposes it suffices to note that (scs) is weaker than the rule of Universal Instantiation. And (scs) will be a very basic (logical or linguistic) skill that everyone masters who masters some logic (and perhaps that everyone masters who masters some language). In other words, logic without schematic substitution is just as difficult to conceive of, if the logic is supposed to represent our actual *logica utens*. Not just any logical rule we learn, but learning any new compositional phrase requires mastery of

[8]To be precise, for the application of (scs) in reasoning, we need not only the ability to compose new expressions, but also to decompose them. This requires compositionality as well as *inverse compositionality* (Pagin, 2003).

schematic substitution.[9] Again, any logic that is supposed to model an actual *logica utens* will have to contain (SCS) then.

Of course, there can be "logics" that are weaker than classical logic and that do not contain Modus Ponens or (SCS). But the question isn't whether there are logic-like formal systems that may or may not allow reasoning that would enable to grasp the application conditions of a new rule. The question is whether there is any formal system that models a possible *logica utens* such that it enables the reasoner to adopt a new rule. If any application of logical rules requires some (suitably restricted form of) Modus Ponens and (SCS) and if from that a reasoner can obtain a (a suitably generalized) form of Modus Ponens and (SCS) that is sufficient for grasping the application conditions for a new rule, then every logic that is a possible *logica utens* will allow upwards adoption. If this is right, then Kripke's "adoption problem" does not actually pose a problem for the adoption of a new logic.

But Kripke's scenario is anyway highly artificial. No one adopts a logic simply because some oracle told them that the principle behind it is logically valid. We may come to reason in new ways, because we adopted a new theoretical perspective on matters of validity. What this process may look like and how it gets initiated will be the focus of the next section.

4 Adoption in a logical theory

We have argued that revision of logic by adoption of a new logical principle is best understood as a revision of one's *logica utens*. In the scenario envisaged by Kripke, the agent was asked to follow the instructions of some logical oracle. In this section we consider the patterns of adoption isolated earlier in the more realistic context of a *logical theory*, defined loosely as a collection of principles governing the core notions

[9] And, as we argued above, schematic substitution is implicit in our mastery of composing and decomposing complex expressions in general.

involved in one's specific account of logical consequence: these notions might involve global accounts of notions such as truth-preservation, predication, negation, implication, assertion, formality, consistency, provability and so on, and therefore giving a full account of one's preferred logical theory is often a highly non-trivial matter.

4.1 Deflationary and inflationary views of logical theories

The characterization of logical theories just sketched is not the only one considered in the literature. It more or less aligns to what Hjortland (2017) calls *non-deflationary* logical theories. Following this terminology, a typically deflationary account is the one articulated in Williamson (2017), which holds that the ultimate task of logical theories is to unravel general claims about the world. Meta-linguistic notions such as truth and validity are not the primary concern of logic, which is essentially a *non-metalinguistic* enterprise pointed at discovering absolutely general laws of reality. In this, logic does not differ from physics, or from metaphysics; it only proceeds at a much higher level of abstraction.

Williamson suggests that a logical theory is a collection of nonmetalinguistic generalizations corresponding to logical truths. This picture is motivated by the following process: Williamson starts from valid inferences in some logic $\mathcal{S}$ in a language $\mathcal{L}_{\mathcal{S}}$ – e.g. $\neg\neg\varphi \therefore \varphi$. It proceeds by extending $\mathcal{L}_{\mathcal{S}}$ with new, higher-order variables of the same type as formulas of $\mathcal{L}_{\mathcal{S}}$ and by replacing the entailment relation with a conditional – in our example, this turns $\neg\neg\varphi \therefore \varphi$ into $\neg\neg X \to X$. The process is then completed by universally quantifying over the free higher-order variables of the translation of the logical claim under considerations. A logic, in this view, is a collection of claims such as $\forall X\, (\neg\neg X \to X)$. Endorsing a logic is endorsing a collection of universally quantified claims: since there is no reason to consider higher-order quantification as more metalinguistic than first-order quantification (Williamson, 2017, p. 329), a logical theory is no more metalinguistic than any other theoretical enterprise seeking universal laws, such as physics itself.

Given our analysis, the problem of adoption in a deflationary logical theory of the kind just sketched does not arise. Already the process of turning a purported valid inference into a universal generalization of the appropriate type requires a prior understanding of quantification. It is hard to see how this understanding may not involve something as basic as (SCS): this is especially clear in the step that requires the expansion of one's language with variables of the appropriate type. The very adequacy of this process seems to rest on the capability of instantiating such variables with formulas of $\mathcal{L}_{\mathcal{S}}$, as required by (SCS). Moreover, the substitution of the entailment sign with a suitable conditional certainly presupposes a conditional that satisfies Modus Ponens. How can the reduction be put to use, if one cannot retrieve the original inference by assuming an instance of the antecedent of the law-like conditional and conclude its consequent via Modus Ponens? The structural assumptions required by Williamson's view of logical theories therefore presuppose both (SCS) and Modus Ponens; our analysis of the pattern for adoption entails that the circularity involved for the adoption of a new rule does not arise in the presence of such principles.

Finally, it seems also clear that to accept a deflationary view of logical consequence of the sort just given one should also give an account of the role of second-order quantification, its semantics, and of the semantics of predication of properties to objects, which is discussed at length in Williamson (2013) and which is surely not uncontroversial (see for instance Bacon et al. (2016) for a critique of Williamson's view). In other words, when one moves from the proponent's intention to the actual development of the proposal, the very distinction between a deflationary and an inflationary logical theory does not appear to be so clear-cut. We now move to more substantial logical theories, and ask whether the problem of adopting a logical inference may arise in that context.

4.2 Logical theories and metatheory

Logical theories, in the abstract – and more substantial – sense considered in this section, can be seen as the formal counterpart of *logicae utenses*. In the same way as a logica utens encodes the agent's dispositions towards a class of inferences (or meta-inferences), a logical theory enriches this acceptance of a class of validities with a collection of meta-theoretic claims concerning semantic and proof-theoretic notions associated with such inferences. For instance, the logical theory of intuitionistic logic includes an account of what is a canonical or direct method of verification, as opposed to an indirect one. Similarly, the logical theory of paraconsistent logic involves a characterization of negation and falsity that substiantially differs from the classical exclusive approach to negation. Taken at face value, claims of the sort just described belong to the metatheory of one's logic. And such metatheory typically amounts to a fragment of classical or intuitionistic mathematics. There have been interesting attempts, in the context of some approaches to semantic paradoxes, to align a weaker nonclassical approach – generally substantially weaker than intuitionistic logic, since semantic paradoxes affect classical and intuitionistic logic alike – in the object theory with a nonclassical metatheory (Leitgeb, 2007; Bacon, 2013). Such attempts, however, are at best at an initial stage and cannot yet be considered to be actual rivals of a classical or intuitionistic metatheory.

How can the problem of adoption be formulated in this richer framework? There are, it seems, at least three fundamental ways to look at the question, depending on one's stance towards the structure of logical theories. First, one can keep all metatheoretic principles fixed, by considering them in a purely instrumental role, and take into account only adoption and revision for the object-theoretic logical inferences. The second is to consider at face value the logical tools that one relies on in their metatheory and ask whether, in the light of Kripke's examples, such principles can be adopted or revised. The third is to generalize the problem and conceive of, if not all, some principles of the metatheory as broadly logical and investigate whether *they* can be adopted or revised:

revision of such quasi-logical principles may determine a change in one's object-linguistic validities. The next three sections will deal with such options.

4.3 Logics in logical theories

According the first reading of logical theory, the cluster of metalinguistic notions that account for a collection of logical validities plays a mere instrumental role. One should not consider the logical inferences one draws in reasoning *about* their object-language validities as truly logical, but mere instruments to define and compare such inferences. This view is analogous to the way in which some advocates of nonclassical solutions to semantic paradoxes conceive of the role of their classical metatheory: although the metatheory contains classical concepts of truth, satisfaction, or property predication, they should only be seen as tools to provide safe foundations to *real* truth, satisfaction, predication, which are the object-linguistic ones – cf. for instance (Field, 2008, §5.6).

For our purposes there is no need to debate at length the coherence of this position. It may for instance be argued that, even if one grants the possibility of distinguishing between real and instrumental concepts, say, of truth, it is much more difficult to rely on such distinction for logical principles. Since logic is in general not tied to a specific language, accepting a logical principle just means accepting it across the board, regardless of whether they are instantiated in the language endowed with what one considers to be the real truth predicate. At any rate, even if one grants this distinction in the context of logical principles, our analysis carries over without substantial modifications. Since (SCS) is presupposed by any competent language user, the only problematic case of revision may be the one in which one starts from a logic that does not feature Modus Ponens. That is, a case of what we called ADD above. Unlike what happens in the local process of adoption above, however, in the present case there is a more recognizable tension. The agent finds themselves in a puzzling situation: they can rely on the metalinguistic inference:

(3) If 'φ' and '$\varphi \to \psi$' obtain, then 'ψ' obtains,

but not on its object-linguistic translation '$\varphi, \varphi \to \psi \therefore \psi$', where we assume that the meta-linguistic 'if…then…' translates the entailment sign, and the 'obtains' is a Tarskian truth predicate for the object language. It is clear that, for the instrumentalist position to have any bearing at all, one cannot rely on the assumption that the object-language is included in the metalanguage. The disquotational nature of Tarskian truth for truth-free sentences – where of course, 'truth-free' here means free of the Tarskian, metalinguistic truth predicate, not free of a possibly object-linguistic one – would entail that $\varphi, \varphi \to \psi \therefore \psi$ is a theorem of the metatheory: this would create an immediate conflict. The only reasonable option is that '$\varphi, \varphi \to \psi \therefore \psi$' belongs to a language that *extends* the language of one's metatheory with, say, primitive notions that account for the meaning of the inference sign '$\therefore$'. The language of the metatheory, by contrast, would be a purely extensional language such as, for instance, the language of ZFC. It is in this particular case that the problematic weak-to-strong revision considered above resurfaces.

As we stated in the previous section, the absence of Modus Ponens in one's logic is sufficiently problematic *per se*. In the present context, however, together with an account of how one can put forward a logic in which Modus Ponens fails, the instrumentalist needs also to claim that, although one can always rely on a certain logical inference for the language in which a *meaningful translation* of an object-theoretic inference is given, this fails to be the case for the original, non-translated language. This is not inconceivable. We only feel that the burden of proof is on the proponent of such a logical theory to explain how one can do without Modus Ponens and at the same time uphold such fundamental uses of Modus Ponens in one's metatheory.

4.4 Logics of logical theories

A second way of revising a logical theory might be to revise the *logical principles* of one's overall logical theory, including the logic of metalinguistic concepts. In the abstract case, it is clear that this is no more nor less problematic than allowing for a revision of object-linguistic logical principles: the logical component of one's logical theory is simply a collection of inference patterns that one recognizes as valid *in the more general language of the metatheory*. There seem to be no substantial differences between the analysis of the local adoption problem above and the present case: again, the only problematic cases might be cases of ADD, in which from a weaker metatheory one moves to a stronger metatheory.

For instance, to consider a case that is compatible with what we deemed "actual" metatheoretic frameworks for validity, one might ask whether the intuitionistic logician is able to adopt a classical perspective on validity. In the current setting, this can simply be reduced to the problem of whether one can instruct an intuitionist to infer according to, say, double negation elimination $\neg\neg\varphi \therefore \varphi$. But in the presence of (SCS) and Modus Ponens, we have seen that this is unproblematic: one starts with exhibiting a specific doubly negated instance $\neg\neg A$ of $\neg\neg\varphi$; by (SCS), one provides the intuitionist with the concrete instance of – a suitable translation of – the original principle 'if $\neg\neg A$, then A'. From $\neg\neg A$ and 'if $\neg\neg A$, then A', the agent that possesses the general capability of inferring by Modus Ponens can immediately conclude A.

In practice, since we agreed that, to date, intuitionistic or classical foundational frameworks are the only reasonable candidates for the logic of the metalinguistic components of one's logical theory, we can safely conclude that no worries of circularity can arise in this second reading of logical theories.

4.5 Quasi-logical notions in logical theories

We are left with the third notion of revision for one's logical theory. This is, arguably, the option that is closest to actual cases of revision of one's logical assumptions. Paraconsistent and paracomplete logicians motivated by semantic or logical paradoxes, for instance, aim at a revision also of foundational tools, such as comprehension axioms, that are needed to define their notion of logical consequence. In this context, one considers not only a collection of logical inferences, but also the principles of quasi-logical notions such as truth, property predication, and consequence as passible of revision. For instance, one can require their notion of consequence to be squared in terms of preservation of disquotational truth: as we shall see in a moment, this might affect the genuinely logical validities that are admissible in the framework. Can the worries of circularity adumbrated in the local case of adoption in the previous sections have some bearing on such cases of revision?

Prima facie it seems that this scenario evades our initial question. After all, even if it is hard or problematic to revise or adopt quasi-logical principles, our case against circular worries in previous sections may well still hold. However, in the analysis of concrete cases, the revision of one's metalinguistic concepts may determine cases of adoption of crucial rules such as Modus Ponens that are precisely the cases that are problematic in the light of Kripke's argument. We will now consider one such case taken from the recent literature.

Disquotational truth and consequence. Let us consider a popular account of the semantics of a language $\mathcal{L}$ containing its own truth predicate Tr.[10] An atomic, non-semantic, sentence $P(t)$ is true iff t is indeed P, and false if t is not P. A conjunction is true iff both conjuncts are true, false iff at least one conjunct is false. A universally

[10] We can safely assume that the language As it is customary in the literature, this can be achieved either by working in a model of a direct axiomatization of concatenation or, equivalently, via an arithmetical setting.

quantified sentence is true iff all its instances are true, false if at least one instance is false. A truth ascription $\mathsf{Tr}\ulcorner\varphi\urcorner$ is true iff φ is true, false iff φ is false. In other words, we are assuming that our logical theory features a largely compositional and disquotational truth concept.[11]

Let us now assume for the sake of the argument that this picture of compositional, self-referential truth is fundamentally correct. So correct that we want to extend this account of truth – which is now grounded truth – to *consequence*. In other words, we want to know which arguments, now still informally conceived, are licensed by our picture of truth.[12] The first is to define 'consequence', or 'follows from', as:

> CONSEQUENCE1: ψ is a CONSEQUENCE1 of φ iff whenever φ is true, ψ is also true.

It follows that, in this picture of consequence via truth-preservation, we are allowed to reason with sentences that do *not* have a detrminate truth value. For instance, CONSEQUENCE1 will validate the inference from a Liar sentence λ to λ, even if λ does not have a determinate truth value. More generally, any inference of the form (φ, φ) will be licensed by CONSEQUENCE1, regardless of the semantic status of φ.

This brings us to our next point. One plausible desideratum that one may want to impose on their logical theory is the following: any semantic notion should ideally be internalizable into the object language. This, for instance, would avoid problematic asymmetries between the expressive and deductive power of the object and metatheory.

[11] This informal picture can of course be translated into a precise inductive definition of the set of true and false sentences in the style of Kripke (1975) and Martin (1984). In particular, if one identifies syntactic objects with natural numbers, the clauses just sketched can be turned into a monotone operator on sets of natural numbers taking one extension of the truth predicate to another extension until no more sentences can be added – i.e. until one reaches a *fixed point*. At the fixed point the truth becomes *transparent*: a sentence φ, possibly including the truth predicate, is in the fixed point if and only if $\mathsf{Tr}\ulcorner\varphi\urcorner$ is. This internalization of truth in the language has the additional advantage, that for many is considered to be the *main* advantage of the construction, that avoids resorting to a metalinguistic notion of truth in the semantics, and approximates the ideal of a *semantically closed* language. For a more recent reference see for instance the 'silence' strategy defended in Horsten (2012) and the discussion in Field (2008).

[12] Without loss of generality, we now deal with pairs of sentences only.

What happens if we try to internalize CONSEQUENCE1 via an object linguistic operator or predicate? For simplicity, let us try to introduce a connective $\twoheadrightarrow$ for CONSEQUENCE1. It seems intuitively correct to let $\twoheadrightarrow$ be governed by the clauses:

($\twoheadrightarrow$1) if ψ is CONSEQUENCE1 of φ, then $\varphi \twoheadrightarrow \psi$ is true;

($\twoheadrightarrow$2) if φ and $\varphi \twoheadrightarrow \psi$ are true, then ψ is also true.

This option is a non-starter. The resulting logical theory would in fact be trivial (see Appendix A). The paradox suggests, however, another option to give a semantically closed logical theory starting with the notion of grounded truth. Let us consider the a new notion of consequence, where again truth and falsity have to be understood as *determinate truth* and *determinate falsity* respectively:

CONSEQUENCE2: ψ is a CONSEQUENCE2 of φ iff either φ is false or ψ is true.

The obstacles we found in the internalization of the notion of CONSEQUENCE1 have now disappeared. In fact, by following the pattern above, we can introduce a new connective $\hookrightarrow$ in our language corresponding to CONSEQUENCE2 and governed by clauses($\hookrightarrow$1) and ($\hookrightarrow$2) that are analogous to ($\twoheadrightarrow$1) and ($\twoheadrightarrow$2). For instance, it is not the case that λ is a consequence of λ itself, for the simple reason that, according to the picture of semantic groundedness, λ is neither determinately true nor determinately false; therefore, already the first step of the paradoxical reasoning of Appendix A is blocked: no sentence entails itself – and therefore other sentences semantically equivalent to itself – unless it is determinate. The resulting theory is indeed paradox-free, as it can be shown by an easy modification the fixed point construction provided in Nicolai and Rossi (2017), and it displays notions of truth and consequence that go hand in hand.

Change the logical theory, change the logic. In moving from CONSEQUENCE 1 to CONSEQUENCE 2, it seems, one is only motivated by extra-logical concerns related to one's logical theory. However, the logics associated with such consequence notions are quite different, and the move from one to the other may *prima facie* display a pattern that resembles the only possibly problematic case of revision that we dubbed ADD in previous sections. Both concepts of consequence are in fact based on disquotational truth: however, CONSEQUENCE 1 cannot be internalized in the object language of one's logical theory, whereas CONSEQUENCE 2 can. Furthermore, the logical inferences licensed by CONSEQUENCE 1 are close to what is commonly known as First Degree Entailment (FDE). In such logic, neither the rule of conditionalization 'if from φ you can infer ψ, then $\varphi \to \psi$', nor Modus Ponens hold. By contrast, the logic associated with CONSEQUENCE 2 satisfies *all classical rules for connectives, including conditionalization and modus ponens*.[13] However, it *does not* satisfy the structural rule of identity or reflexivity: 'from φ, infer φ'.

The case at hand amounts to a case in which, by modifying broadly logical concepts of one's logical theory, one can find themselves in the situation of adopting a form of Modus Ponens from logical assumptions that do not include Modus Ponens in the first place. Prima facie this case is not covered by what has been said so far. It should be clear, however, that from the perspective of a classical or intuitionistic logical core of one's logical theory both the formulation of FDE and the formulation of the fully operational, non reflexive logic of CONSEQUENCE 2 are all restrictions of classical rules of inferences: in the former case, one restricts the rules of the conditional. In the latter case, the restriction operates at the level of structural inferences. To recall our discussion in §4.3, the present case may be simply understood at the level of extensional translations of object-language inferences in the methematical support theory built-in in one's logical theory.

To the extensional definitions of CONSEQUENCE 1 and CONSEQUENCE 2 there correspond collections of object language inferences. No change of logical assumptions, nor

[13]To be precise, the logic associated with CONSEQUENCE 2 satisfies all classical *metainferences*. In a natural deduction setting, this would amount to satisfying both introduction and elimination rules for the conditional. In a natural deduction formulation of FDE, by contrast, neither the introduction nor the elimination rule for the conditional hold. For details see again Nicolai and Rossi (2017).

tabula rasa adoption of logical principles seems required. Since the present case deals with the particularly challenging case of "adoption" of Modus Ponens, our reaction should easily generalize to several similar cases of adoption of quasi-logical notions.

5 Alternative Quinean Targets for Kripke's Argument

For all we have argued so far it seems that there is no adoption problem that would pose an obstacle or challenge to the idea that we can rationally revise our *logica utens*. Neither in the abstract scenario that Kripke discusses nor in actual cases is it plausible to assume that we lack the resources to apply new logical rules in reasoning.

As we explained in section 2, we took it on the authority of Kripke scholars that are more familiar with Kripke's actual writing[14] on the matter that his real target is Quine's view on the revisability of logical principles. In light of the fact that Kripke's argument seems to utterly miss the target here, we would now like to briefly discuss whether Kripke in fact had a different aspect of Quine's view about logic in mind when he claimed that Quine's argument against Carnap applies in the same way against Quine's own conception.

We could identify four possible alternative targets that are part of Quine's conception of logic and may, at least *prima facie*, be affected by the proposed regress. The candidates are in turn the adoption of a first logic, the transition from the acceptance of a principle to the adoption of certain behavior, the problem of the missing normative force of purely descriptive logical principles, and the *knowledge that*/*knowledge how*-distinction. We will discuss the candidates in this order.

[14] To the extent that there is such. We are only aware of the few quotes that Padro provides and that we have already presented in full.

5.1 The Adoption of the First Logic

So far we have considered the Kripkean challenge as being directed at Quine's idea that we can adopt a *new* logic. So it was legitimate in our argument to suppose that some logic and some language is already in place and that an individual has on the basis of some *reasoning* arrived at the conviction that she should adopt a different way of reasoning, that she should adopt a new logic.

But perhaps Kripke's challenge is indeed closer to Quine's original point against conventionalism and concerns the question how – on Quine's view – logic could have ever gotten off the ground. After all, also on the conception that logic is just general, firmly held belief, there seems to be the issue that firmly believing Modus Ponens does not yet allow you to reason with it, if you don't yet have that capacity. Thus, as a general theory of what logic is, Quine's theory isn't better than conventionalism, since it still is open to the challenge that it can't explain how the first logical principles could have been adopted in absense of an already existing logic.

Although this well may be so, it is not clear that this is a challenge that Quine needs to address. Or, in other words, it seems to us that Quine, quite clearly, does not have to address it. Quine presents a picture according to which the first principles of logic are not adopted as a result of engaging with some explicit formulation of the principles (as conventionalism has it), but where they get adopted in behavior and only later are reconstructed in terms of explicit reasoning principles or rules. This adoption in behavior does not require that Quine's theory of belief revision applies to it, so he does not at all need to explain how *homo sapiens* managed to develop structured reasoning that is describable in terms of schematic inference principles. This should be part of a general naturalistic account of how higher cognition and reasoning in general developed. To require that Quine's conception of logic provides some detailed explanation of this process is completely inadequate.

5.2 From Belief to Behavior

A second potential target for the regress argument is Quine's emphasis on belief. Quine considers logic to be nothing but firmly held *belief.* But adopting a logic is not just adopting some belief. It is adopting a way of reasoning. There are two ways to make that challenge. The first would be to see this as a critique of Quine's behaviorism. For Quine, having a certain belief (for example, the belief that Modus Ponens is valid) just means to show certain forms of behavior (for example to reason in ways that are licensed by Modus Ponens). But perhaps that's too short-sighted. As the regress argument shows (on this interpretation), one may accept a belief (viz. that Modus Ponens is valid) and yet fail to show the appropriate behavior (e.g. to assent to implications that are licensed by Modus Ponens). The "regress argument" then doesn't show that there indeed is a regress problem, but that there may be a problem of a certain kind of "stubborness": someone may count as having grasped and adopted a certain belief, but just doesn't act in a way that may be canonical for the ascription of that belief.

This may be a reasonable challenge to the idea that 'S beliefs that p' can be analyzed as 'S is disposed to assent to this and that under conditions such and such'. But this doesn't seem to be a specific problem for Quine's theory of logic than rather a problem for Quine's theory of belief. However, while the regress argument *displays* the problem, it doesn't actually establish anything that could seriously be regarded as an argument for the claim that such an analysis must fail. It seems still perfectly reasonable to just respond to such a regress argument that it merely shows that the person in the dialog who doesn't reason in accordance with, for example, Modus Ponens has not yet actually adopted the relevant belief.

5.3 The Normative Force of Logical Principles

A closely related challenge (one that actually makes use of the regress) is to interpret the regress argument as pointing out that logic is *normative.* Logic tells us how we ought to reason. However, the general principles that are featured in the regress arguments are not norms or imperatives. They don't say anything about how anyone should reason. Therefore there is a gap between adopting the belief that a certain logical principle is true and adopting the norm that one ought to reason in a certain way. Quine, who takes logical principles to be just like any other general scientific hypotheses overlooks this.

As Besson (2016) explains this could work only if we'd lack a bit of non-propositional knowledge, like an imperative or a rule, when we merely have accepted the propositional knowledge that Modus Ponens is a valid principle. Is there a plausible candidate for the normative knowledge that we lack? The recent discussion of the normative force of logic strongly suggests that there isn't (for an overview, see Cohnitz and Estrada-González (2019)). In order for the regress to get off the ground, we'd need an imperative or a rule that would "move" a subject to reason in accordance with the logical principle at issue. However, as we have learned from Harman (1986) and others, logical principles can't give rise to such rules. It simply isn't always rational to use Modus Ponens and endorse q whenever you believe p and $p \supset q$ for some p and q. However, a weaker principle that would, say, allow that it is rationally permissible to believe q whenever you believe p and $p \supset q$ for some p and q is plausible, but also easily available in the form of propositional knowledge (see Besson (2016) for details).

5.4 Knowledge that and knowledge how

This leaves us with a last candidate which again tries to explain the problem of the regress by a certain insufficiency of the merely propositional knowledge that we acquire, when we accept the claim that Modus Ponens is valid. We mentioned in the beginning in section 2 that Devitt and Priest both see the problem of adoption as primarily an issue of acquiring certain *knowledge how* after one has convinced oneself of the relevant *knowledge that*.

Take a familiar analogy: from reading a book about how one rides a bike, one doesn't know yet how to ride a bike in the sense that one won't be able (yet) to ride a bike. The latter will require certain practical competence, a skill, that is not identifiable with any kind of propositional knowledge. The acquisition of that skill might require training. In the regress argument, the subject accepts Modus Ponens but doesn't have the skill to apply it, she thus gets a new bit of propositional knowledge which she doesn't know how to apply either, and so forth.

Devitt and Priest seem to think that also the adoption of logic requires that we train ourselves in the application of a rule in order to be able to apply it. However, as our discussion above shows, the competence that rule application of logical principles requires is merely the competence with basic rules like MP or SCS. The relevant knowledge how is the mere capacity to reason in the first place. Adoption of a new logic thus does not require training in new rules.

Another question may be what it takes to "see" new implications that one didn't see as implications with the "old" logic, or how one can get to stop seeing implications that aren't implications according to a new logic. That is an interesting question for empirical psychology, but not a principled obstacle to the adoption of a new logic. After all, we don't "see" all implications of what we think is our current *logica utens*, and we do make inferences that fail to be licensed by that logic.

6 Conclusions

We showed that the so called Adoption Problem does not pose a serious obstacle to the idea that logic is rationally revisable, nor to any other aspect of a generally Quinean conception of logic. How our cognitive capacities – and reasoning in particular – developed is a fascinating question, but as soon as there was a way for *homo sapiens* to reason about reasoning, there also was a rational way to develop that capacity further.

References

Azzouni, J. (2014). A defense of logical conventionalism. In Rush, P., editor, *The Metaphysics of Logic*, chapter 2, pages 32–48. Cambridge University Press.

Bacon, A. (2013). Non-classical metatheory for non-classical logics. *Journal of Philosophical Logic*, 42(2):335–355.

Bacon, A., Hawthorne, J., and Uzquiano, G. (2016). Higher-order free logic and the prior-kaplan paradox. *Canadian Journal of Philosophy*, 46(4-5):493–541.

Besson, C. (2016). Norms, reasons and reasoning: A guide through lewis carroll's regress argument. In Star, D., editor, *The Oxford Handbook of Reasons and Normativity*. Oxford University Press.

Cohnitz, D. (2005). Is compositionality an apriori principle? In Werning, M., Machery, E., and Schurz, G., editors, *The Compositionality of Concepts and Meanings: Foundational Issues*, pages 23–58. De Gruyter.

Cohnitz, D. and Estrada-González, L. (2019). *An Introduction to the Philosophy of Logic*. Cambridge University Press.

Devitt, M. (2018). *The Adoption Problem in Logic: A Quinean Picture*. Unpublished Manuscript.

Ebbs, G. (2011). Carnap and quine on truth by convention. *Mind*, 120(478):193–237.

Field, H. (2008). *Saving truth from paradox*. Oxford University Press, Oxford.

Harman, G. (1986). *Change in View*. MIT Press.

Hjortland, O. T. (2017). Anti-exceptionalism about logic. *Philosophical Studies*, 174(3):631–658.

Horsten, L. (2012). *The Tarskian Turn.* MIT University Press, Oxford.

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, 72:690–712.

Leitgeb, H. (2007). On the metatheory of field's 'solving the paradoxes, escaping revenge'. In Beall, J. C., editor, *Revenge of the Liar: New Essays on the Paradox.* Oxford University Press.

Martin, R. L. (1984). On representing true-in-l'in l robert l. martin and peter w. woodruff. In Martin, R. L., editor, *Recent Essays on Truth and the Liar Paradox*, page 47. Oxford University Press.

Morris, S. (forthcoming). Quine against lewis on truth by convention. *Pacific Philosophical Quarterly.*

Nicolai, C. and Rossi, L. (2017). Principles for object-linguistic consequence: from logical to irreflexive. *Journal of Philosophical Logic.* Online-First, DOI: https://link.springer.com/article/10.1007/s11225-017-9727-y.

Padro, R. (2015). *What the Tortoise Said to Kripke: the Adoption Problem and the Epistemology of Logic.* CUNY Academic Works.

Pagin, P. (2003). Communication and strong compositionality. *Journal of Philosophical Logic*, 32(3):287–322.

Priest, G. (2014). Revising logic. In Rush, P., editor, *The Metaphysics of Logic*, pages 211–223. Cambridge University Press, Cambridge. DOI: 10.1017/CBO9781139626279.016.

Quine, W. V. O. (1953). Two dogmas of empiricism. In *From a Logical Point of View*, pages 20–46. Harvard University Press.

Quine, W. V. O. (1976). Truth by convention. In *The Ways of Paradox and other Essays*, chapter 11, pages 77–106. Harvard University Press, revised and enlarged edition edition.

Szabó, Z. G. (2000). *Problems of Compositionality*. Routledge.

Williamson, T. (2013). Logic, metalogic and neutrality. *Erkenntnis*, 79(S2):1–21.

Williamson, T. (2017). Semantic paradoxes and abductive methodology. In Armour-Garb, B., editor, *Reflections on the liar*. Oxford University Press, Oxford.

Appendix A

We show that, once the connective $\twoheadrightarrow$ is in the language of our logical theory based on CONSEQUENCE 1, our background assumptions on the expressive capabilities of our language – capable of diagonalization – entail the existence of a Curry sentence κ which is equivalent to $\mathsf{Tr}\ulcorner\kappa\urcorner \twoheadrightarrow \bot$, where $\bot$ is a fixed absurdity (e.g. $0 \neq 0$ if we are working in an background arithmetical language). Then we can reason as follows:

- if κ is true, then so is $\mathsf{Tr}\ulcorner\kappa\urcorner \twoheadrightarrow \bot$ by the assumptions on syntax;
- by transparency we obtain $\mathsf{Tr}\ulcorner\kappa\urcorner$, and therefore $\bot$ is true by ($\twoheadrightarrow$2) under the assumption of the truth of κ;
- but then, if κ is true, so is $\bot$, that is $\bot$ is CONSEQUENCE1 of κ;
- therefore, by ($\twoheadrightarrow$1) and transparency, also $\mathsf{Tr}\ulcorner\kappa\urcorner \twoheadrightarrow \bot$ is true;
- but then $\mathsf{Tr}\ulcorner\kappa\urcorner$ and $\bot$ is true by ($\twoheadrightarrow$2) under no assumption.

The last line is clearly contradictory: $\bot$ is cannot certainly be true. Something has gone wrong, and it seems to compromise our project of a semantically closed logical theory.