

# Rules and Self-Citation\*

Ori Simchen

## Abstract

I discuss a neglected solution to the skeptical problem introduced by Lewis Carroll's "What the Tortoise Said to Achilles" (1895) in terms of a self-citational inferential license. I then consider some responses to this solution. The most significant response on behalf of the skeptic utilizes the familiar distinction between two ways of accepting a rule: as action-guiding and as a mere truth. I argue that this is ultimately unsatisfactory and conclude by opting for an alternative conception of rules as representations of behavior deployed for various purposes, some theoretical and others practical. This alternative conception does not allow the skeptical problem to get off the ground.

## 1 Introduction: Carroll's Regress

Lewis Carroll's "What the Tortoise Said to Achilles" has been a philosophical favorite ever since its publication in *Mind* in 1895. A parable of sorts, it is a dialog between Zeno's familiar characters of Achilles and the Tortoise concerning what is described as "a race-course, that most people fancy they can get to the end of in two or three steps, while it *really* consists of an infinite number of distances, each one longer than the previous one" (278). The upshot is a pernicious form of rule-skepticism personified in the character of the Tortoise. As is common, we can streamline the original presentation of the problem by looking at a schema of an argument in MP form:<sup>1</sup>

- (1)  $p$
- (2) if  $p$ , then  $q$

---

\*To appear in *Journal for the History of Analytical Philosophy*.

<sup>1</sup>I regard argument-schemas as arguments throughout as a matter of terminological convenience.

(3)  $q$ .

The Tortoise of the story concedes the truth of (1) and (2) and resists drawing the conclusion (3). Why the Tortoise so resists has been the topic of much speculation.<sup>2</sup> What is relatively uncontroversial, however, is that the Tortoise raises the possibility of not accepting the conclusion despite its obvious entailment by the premises. The claim that the conclusion does in fact follow from the premises is then added, upon the Tortoise's insistence, as an additional premise, yielding:

(1)  $p$

(2) if  $p$ , then  $q$

(2.1) from (1) and (2), (3) follows

(3)  $q$ .

The Tortoise next concedes the truth of the premises but resists drawing the conclusion in an analogous way, yielding the further enriched variant:

(1)  $p$

(2) if  $p$ , then  $q$

(2.1) from (1) and (2), (3) follows

(2.11) from (1) and (2) and (2.1), (3) follows

(3)  $q$ .

And so it goes, on to the next variant:

(1)  $p$

(2) if  $p$ , then  $q$

(2.1) from (1) and (2), (3) follows

(2.11) from (1) and (2) and (2.1), (3) follows

---

<sup>2</sup>For my own earlier take on the issue, see Simchen (2001). In that paper I was concerned to show that the Tortoise's deliberative stance, which inspires a distinct form of normative skepticism, isn't necessary. In what follows I argue that, appearances to the contrary notwithstanding, the stance in question isn't really possible after all. For a useful bibliography of the secondary literature, see Imholtz and Moktefi (2016).

(2.111) from (1) and (2) and (2.1) and (2.11), (3) follows

(3)  $q$ .

And so on. It appears that the Tortoise's repeated demand for further and further inferential licenses to legitimate drawing conclusion (3) cannot be met. The rule-skeptical generalization is that drawing conclusions is never fully legitimate. There is always room for failing or refusing to draw a conclusion entailed by accepted premises while accepting the claim that the conclusion indeed follows from them.

## 2 Self-Citation as Regress Blocker

There is a neglected response to the skeptic that would spell trouble for the Tortoise's deliberative stance straight away. Intuitively, if a premise addressing an argument's premise-set is added to an argument, one would expect it to address the extant premise-set in its entirety, itself included. An added inferential license is just such an added premise. The Tortoise ignores this intuitive requirement of self-inclusion for the added premise, however, thus launching a regress. With this in mind, let us begin, as before, with the Tortoise failing or refusing to draw conclusion (3) from accepted premises (1) and (2). This time, however, we enrich the original premise-set with a premise that includes its own citation:

(1)  $p$

(2) if  $p$ , then  $q$

(2.1') from (1) and (2) and (2.1'), (3) follows

(3)  $q$ .

The self-citation of (2.1') is achieved via self-denotation.<sup>3</sup> Now, what can be said about the Tortoise's reticence when it comes to this enriched variant (1),(2),(2.1')/(3)? By analogy to the original case, we would have to ascribe to the Tortoise the acceptance of (1) and (2) and (2.1') without accepting that from (1) and (2) and (2.1'),

---

<sup>3</sup>'Self-denotation' rather than 'self-reference' for reasons that would take us too far afield. Suffice it to say that in the real time production of (2.1'), the token so produced wouldn't be available (yet) to act as a referent for the produced constituent token of '(2.1')'. See Simchen (2013) for further discussion of the contrast between self-denotation and self-reference and its ramifications.

(3) follows.<sup>4</sup> But if the Tortoise accepts (1), (2), and (2.1'), without accepting that (3) follows from (1), (2), and (2.1'), this can only mean that the Tortoise accepts (1), (2), and (2.1') without accepting (2.1') itself: (3) following from (1), (2), and (2.1') is just what (2.1') "says". And this entails that the Tortoise accepts (2.1') and does not accept (2.1'), which is impossible.<sup>5</sup>

Before moving on, let us bring these observations to bear on the argument of the original story:

- (A) things that are equal to the same are equal to each other
- (B) the two sides of this triangle are things that are equal to the same
- (Z) the two sides of this triangle are equal to each other.

With the addition of the premise

- (C) if  $A$  and  $B$  are true,  $Z$  must be true,

we get an argument whose conclusion the Tortoise maddeningly resists:<sup>6</sup>

"If  $A$  and  $B$  and  $C$  are true,  $Z$  *must* be true," the Tortoise thoughtfully repeated. "That's *another* Hypothetical, isn't it? And, if I failed to see its truth, I might accept  $A$  and  $B$  and  $C$ , and *still* not accept  $Z$ , mightn't I?" (Carroll 1895: 279)

But if we substitute

---

<sup>4</sup>If a rationale for accepting (2.1') is needed (a big 'if' – see below) it's that the Tortoise accepts the claim that (3) follows from (1) and (2) alone, which is just (2.1), in which case the Tortoise already accepts the claim that (3) follows from (1) and (2) and any additional premise, including (2.1'), which is just (2.1'). This assumes that the Tortoise understands that for the kind of entailment at issue, anything following from a set of premises follows from any premise-superset of that set (monotonicity). Given the Tortoise's original stance, there is no reason to think this rationale for accepting (2.1') isn't available. Having said that, in the terms introduced by the original story it isn't clear that a rationale for accepting (2.1') is needed in the first place. After all, no rationale is provided in the original story for the Tortoise's acceptance of the equivalent of (2.1) other than being asked by Achilles to do so. See Carroll (1895: 279).

<sup>5</sup>Note that (2.1') is offered as justificatory support for reasoning to (3). Unlike (2.1) as per (1),(2)/(3), (2.1') wouldn't provide an adequate answer to the *post hoc* explanation-targeting why-question "Why did you conclude that  $q$  from the claim that  $p$ , the claim that if  $p$ , then  $q$ , and the claim that  $q$  follows from the previous two claims together with this very claim?". The answer "Because  $q$  follows from the claim that  $p$ , the claim that if  $p$ , then  $q$ , and the previous two claims together with this very claim" violates the irreflexivity of explanation.

<sup>6</sup>The 'must' expresses deducibility. We set aside the occurrences of 'this' in  $B$  and  $Z$ .

( $C'$ ) if  $A$  and  $B$  and  $C'$  are true,  $Z$  must be true

for the original  $C$ , Achilles can answer the Tortoise's second question with a resounding *no* and walk away. For if the Tortoise accepts  $A$  and  $B$  and  $C'$  and still fails to accept  $Z$ , this can only mean, in the terms set by the original story, that the Tortoise doesn't accept that if  $A$  and  $B$  and  $C'$  are true,  $Z$  must be true. But this last claim is just  $C'$  itself, which the Tortoise accepts. In other words, the Tortoise both accepts  $C'$  and does not accept  $C'$ , which is impossible.

We note that (2.1') is distinct from the conditional

(2.1\*) if (1) and (2) and (2.1\*) are true, then (3) is true,

which is problematic in a way dramatized by Curry's paradox. A (2.1\*)-type conditional seems to allow us to infer *anything* from (1) and (2). Consider, for example,

(2.1<sup>†</sup>) if (1) and (2) and (2.1<sup>†</sup>) are true, then the moon is made of green cheese.

Assuming the truth of (1), (2), and (2.1<sup>†</sup>) as premises, we may conclude that the moon is made of green cheese by truth-functional implication due to the equivalence of the truth of (2.1<sup>†</sup>) and (2.1<sup>†</sup>) itself. Discharging the premises from the conclusion, we get (2.1<sup>†</sup>) from no premises whatsoever. So if we assume (1) and (2), and then add (2.1<sup>†</sup>), we may conclude that the moon is made of green cheese from (1) and (2) alone due to the equivalence of each of (1), (2), and (2.1<sup>†</sup>) and its truth. The relevance of all this for the case at hand of adding (2.1') to (1) and (2) is limited, however, due to the fact that (2.1'), unlike (2.1\*) or (2.1<sup>†</sup>), isn't Curry-paradoxical.<sup>7</sup>

But going back to the self-citational version of the original story, we do have the Curry-paradoxical  $C'$ . We can prove that the moon is made of green cheese from  $A$  and  $B$  alone with the aid of the following  $C'$ -type conditional:

( $C^\dagger$ ) if  $A$  and  $B$  and  $C^\dagger$  are true, then the moon is made of green cheese.

We need only assume the truth of  $A$ ,  $B$  and  $C^\dagger$  as premises, rely on the equivalence of the truth of  $C^\dagger$  and  $C^\dagger$  itself, detach the latter's consequent, and then discharge the premises from the conclusion to yield  $C^\dagger$  from no premises whatsoever. Assuming next  $A$  and  $B$  as premises, and adding  $C^\dagger$ , allows us to draw  $C^\dagger$ 's consequent as conclusion via the equivalence of each of  $A$ ,  $B$ , and  $C^\dagger$  and its truth, from  $A$  and

---

<sup>7</sup>Whether or not a formalized version of (2.1') would give rise to a validity variant of Curry's paradox is controversial. For discussion, see Cook (2014). But the result of this controversy is largely beside the point of the present discussion due to the availability of the variant (1),(2),(2.1''),(2.1''')/(3) discussed below, which achieves the effect of self-cituality without the inclusion of a self-citational premise.

$B$  alone. It would appear that at least in the original argumentative setting the Tortoise can resist our proposed regress-blocker in terms of the self-citational  $C'$  by citing its Curry-paradoxicality.

We can replicate, however, the regress-stopping effect of self-citationality without actually engaging in it. Begin again with the Tortoise failing or refusing to draw conclusion (3) from accepted premises (1) and (2). This time enrich the original premise-set with two premises, each including the other's citation:

(1)  $p$

(2) if  $p$ , then  $q$

(2.1'') from (1) and (2) and (2.1'''), (3) follows

(2.1''') from (1) and (2) and (2.1''), (3) follows

(3)  $q$ .

Suppose the Tortoise in this strengthened version accepts (1), (2), (2.1''), and (2.1'''), but doesn't accept that (3) follows from (1), (2), (2.1''), and (2.1''').<sup>8</sup> Then, given the kind of entailment at issue, the Tortoise will not accept that (3) follows from any subset of this premise-set. And so, the Tortoise will not accept that (3) follows from (1), (2), and (2.1'''), and will not accept that (3) follows from (1), (2), and (2.1''). But that means that the Tortoise doesn't accept (2.1'') and (2.1''') after all, which contradicts their acceptance.

Going back to the original argumentative setting, we observe the same pattern:

(A) things that are equal to the same are equal to each other

(B) the two sides of this triangle are things that are equal to the same

(C''') if  $A$  and  $B$  and  $C''''$  are true,  $Z$  must be true

(C''') if  $A$  and  $B$  and  $C''$  are true,  $Z$  must be true

(Z) the two sides of this triangle are equal to each other.

Given that the 'must' expresses deducibility, if the Tortoise accepts  $A$ ,  $B$ ,  $C''$ , and  $C''''$  without accepting that if  $A$ ,  $B$ ,  $C''$ , and  $C''''$  are true,  $Z$  must be true, then the Tortoise will not accept that if  $A$  and  $B$  and  $C''''$  are true,  $Z$  must be true – i.e.  $C''$

---

<sup>8</sup>See footnote 4. The Tortoise accepting (2.1) will accept (2.1'') and (2.1''') as well. If (3) follows from (1) and (2) alone, then, in particular, it follows from (1), (2), and any additional premise.

– and will not accept that if  $A$  and  $B$  and  $C''$  are true,  $Z$  must be true – i.e.  $C'''$ . And this entails that the Tortoise both accepts  $C''$  and  $C'''$  and doesn't accept them, which is impossible.

Curry's paradox doesn't afflict  $C''$  and  $C'''$ . Consider the attempt to deduce the claim that the moon is made of green cheese from  $A$  and  $B$  alone with the aid of the following conditionals:

( $C^*$ ) if  $A$  and  $B$  and  $C^{**}$  are true, then the moon is made of green cheese

( $C^{**}$ ) if  $A$  and  $B$  and  $C^*$  are true, then the moon is made of green cheese.

Assuming the truth of  $A$ ,  $B$ ,  $C^*$ , and  $C^{**}$ , the shared consequent of  $C^*$  and  $C^{**}$  follows by truth-functional implication due to the equivalence of the truth of  $C^*$  and  $C^*$  itself and of the truth of  $C^{**}$  and  $C^{**}$  itself. Discharging the premises from this conclusion gets us

( $C^{***}$ ) if  $A$  and  $B$  and  $C^*$  and  $C^{**}$  are true, then the moon is made of green cheese

from no premises whatsoever. But if we now attempt to detach  $C^{***}$ 's consequent by adducing premises  $A$ ,  $B$ ,  $C^*$ , and  $C^{**}$ , we get the conclusion that the moon is made of green cheese but from the four premises. In particular, we don't get this conclusion from  $A$  and  $B$  alone.

These considerations can assuage our concerns about the Tortoise's stance. The Tortoise is revealed as accepting (2.1') while not accepting it (or accepting (2.1'') and (2.1''') while not accepting them, but we set this more complex argumentative setting aside in everything that follows). But if the description of the Tortoise's stance regarding (1),(2),(2.1')/(3) implies a contradiction, then that stance loses much of its interest. All that remains is a failure or refusal to draw the relevant conclusion. The relevance of these observations to our overall assessment of the Tortoise's stance in the original story is as follows. By tweaking the inferential license added to the premise-set of the original argument, our description of the Tortoise's stance is shown to entail a contradiction. The stance in the tweaked version is thus shown to be impossible. But then it seems that the Tortoise's stance in the original story is saved from impossibility by a fluke, as it were. We have a strong pre-theoretical sense that it isn't possible to acknowledge an inferential license as applying to one's inferential behavior without regarding the behavior as thereby licensed. What is it to acknowledge such a license as obtaining, after all, if not to regard it as licensing behavior? The Tortoise's stance in the original version is made possible by the fact that the inferential license added to the premise-set doesn't include its own citation. But the issue of self-citacionality seems otherwise irrelevant to the stance in question.

As such, the Tortoise's stance in the original case loses much of its interest as well.

Consider an analogy. We have a strong pre-theoretical sense that it isn't possible to sip a thirst-quenching, transparent, odorless, tasteless (henceforth TTOT) liquid without sipping water because a TTOT liquid is just water as far as our sipping behavior goes.<sup>9</sup> Now suppose an argument is given the upshot of which is someone sipping a glass of TTOT liquid while not sipping a glass of water. It happens that there is no possibility of sipping a glass of TTOT liquid *that naturally occurs in three states* without sipping a glass of water because a TTOT liquid that naturally occurs in three states is water as a matter of metaphysical necessity. It also happens that there is exactly one possible liquid other than water that is TTOT, which happens to occur only on the other side of the known universe and naturally occurs in two states. Then the original case of sipping a glass of TTOT liquid without sipping a glass of water is saved from impossibility by a metaphysical fluke which happens to occur on the other side of the known universe and which seems otherwise completely irrelevant to our sipping behavior. In other words, but for the existence of the TTOT liquid on the other side of the universe, sipping a glass of TTOT liquid while not sipping a glass of water would have turned out to be impossible exactly in accord with our pre-theoretical verdict on the matter.

Similarly in the present case, we have a strong pre-theoretical sense that it isn't possible to acknowledge an inferential license as obtaining without treating inferential behavior as licensed. The tweaked version of the story invoking the self-citational (2.1') bears this out, showing that the target stance in the tweaked case is indeed impossible. As mentioned earlier, an added premise addressing an extant premise-set should intuitively address the entire set, itself included, and this intuitive requirement of self-inclusion is expressly flouted by the Tortoise's (2.1) (or *C*) and subsequent iterations. So now it seems that the original story narrowly escapes impossibility by the absence of self-citatoriality, getting off on a technicality which is otherwise irrelevant when it comes to the Tortoise's stance towards inferential licenses. I submit that the Tortoise's stance in the original story thus loses much of its interest as well. It doesn't culminate in a threatening skepticism about rules after all.

---

<sup>9</sup>Pack into being TTOT all the superficial characteristics of water and abstract from the case the philosophical lore surrounding theoretical identifications in natural science. The sense of impossibility here is meant to be pre-theoretical. To make the example more "attitudinal", the reader is invited to substitute *de re* wanting for sipping and keep everything else the same.



### 3 A Skeptical Response

The regress blocker we've been considering is open to an obvious response. The so-called impossibility in the Tortoise's stance, it is now claimed, arises from a superficial take on the issue of acceptance. The Tortoise can accept (2.1') (or accept  $C'$ ) in one sense – as a premise, as true – while failing to accept it in another sense – as a rule, as action-guiding. With proper disambiguation, the apparent contradiction of accepting the self-citational premise while not accepting it is revealed as merely apparent. Let us turn to explore this response in more detail.

A familiar take on the original story maintains that the Tortoise's repeated demand for more and more rules licensing the transition from premises to conclusion exhibits a failure or refusal to attend to an important distinction between premises and rules within one's reasoning. The distinction is allegedly between what inferential behavior turns on, the stuff upon which the thinker is acting in reasoning, the premises; and what makes the behavior a case of genuine reasoning, the rule as incorporated into the relevant bit of inferential behavior. The Tortoise, it is claimed, accepts the rule as an extra premise. But this is shown to be irrelevant to the inferential task at hand, which requires the thinker to incorporate or accept the rule in a different sense into inferential conduct and thereby proceed to the conclusion. Thus Ryle (1946) concludes: "Acknowledging the maxims of a practice presupposes knowing how to perform it. Rules, like birds, must live before they can be stuffed" (11). And Sellars (1949) adds that "a rule, properly speaking, isn't a rule unless it *lives* in behaviour, rule-regulated behavior, even rule-violating behavior... A rule is *lived*, not *described*" (315). The skeptical response we are now considering utilizes this distinction between rules as lived and rules as described. Call acceptance of a rule as described *acceptance*<sup>\*</sup> and acceptance of a rule as lived *acceptance*<sup>\*\*</sup>. Rather than conclude that the Tortoise accepts (2.1') and doesn't accept (2.1'), which is impossible, we now say that the Tortoise accepts<sup>\*</sup> (2.1') and doesn't accept<sup>\*\*</sup> (2.1'). Impossibility averted. The skeptical upshot is that one can always fail to comply with the rule even while acknowledging it as being in force. Acknowledgment that a rule holds is powerless in the face of skeptical reticence. But the question remains how on this way of thinking we are supposed to think of the incorporation of rules into inferential behavior other than as being acknowledged to hold.

Consider again what the Tortoise in the original story is supposed to be doing wrong. The Tortoise fails or refuses to draw a conclusion without the relevant license, which is then cited as an additional premise. The Tortoise then proceeds to raise an analogous concern in the new inferential setting, thus launching the regress. At each step the Tortoise supposedly accepts the rule. And yet at each step the Tortoise fails

or refuses to draw the conclusion licensed by the rule. It is incumbent on a proponent of the present response to the challenge of the self-citational licence to explain what it is to accept\* the rule while failing to accept\*\* it. How can anyone acknowledge an inferential license as a truth about one's inferential behavior without treating it as licensing one's behavior? There is little use in saying "by behaving in a Tortoise-like fashion", which is what we are trying to explain.<sup>10</sup> To accept a license as a truth about one's behavior is to treat the behavior as thereby licensed. Similarly, to accept a ship bearing a name as a truth about the ship is to treat the ship as so named; to accept my having promised you to  $\phi$  as a truth about us is to regard me as having so promised and you as the promisee; to accept a description of France as hexagonal as a truth about France is to treat France as so described. We seem to lack an independently motivated distinction between acknowledging a license as being in effect and regarding it as licensing the behavior. Of course to *claim* that one acknowledges a license as being in effect isn't to acknowledge the license as being in effect. We can mouth the words that the license is in effect without acknowledging that the license is in effect, as evidenced by our behavior. And we can acknowledge a license as being in effect without expressing our acknowledgment. What we cannot do, however, is acknowledge the license as being in effect (as a truth about one's behavior) without acknowledging it as being in effect (as licensing the behavior). When asked why we behave in certain ways we often cite rules at various levels of specificity. When asked why (3) is entered as conclusion upon the acceptance of (1) and (2) as premises, for example, we might answer by citing the fact that (3) follows from (1) and (2). A fuller answer might include the fact of our acknowledgment of the fact that (3) follows from (1) and (2). Such answers to the why-question offer reasons for the inferential behavior at issue. The fact that (3) follows from (1) and (2), or the fact of acknowledgment of the fact that (3) follows from (1) and (2), answer the question why the behavior culminates in entering (3) as conclusion. So in what sense can it still be maintained that the license has to be incorporated into the behavior in some way other than acknowledging that it obtains? There is an unmistakable air

---

<sup>10</sup>Ryle (1946) comments on the contrast between acceptance\* and acceptance\*\* by saying:

What has gone wrong? Just this, that knowing how to reason was assumed to be analysable into the knowledge or supposal of some propositions, namely, (1) the special premisses, (2) the conclusion, plus (3) some extra propositions about the implication of the conclusion by the premisses, etc., etc., *ad infinitum*. (6-7)

This in effect answers the question how one can acknowledge an inferential licence as a truth about one's inferential behavior without incorporating it as a license into one's inferential behavior by citing the Tortoise's behavior as a case in point.

of mystery here.<sup>11</sup>

## 4 Rules as Representations

The mystery dissipates as soon as we reorient our thinking about rules and discard the assumption that inferential rules must somehow be incorporated into inferential behavior as ingredients, so to speak. A rule of conduct is a representation of behavior as patterned in a certain way. To say that the rule is a representation isn't to say that it should be thought of as merely stating or describing that the behavior occurs. Far from it. We use such representations of behavior for a variety of purposes, from explaining and predicting to instructing and enjoining would-be practitioners to engage in the relevant practice. Whatever our purposes for using rules of conduct may be, a rule for adding integers is a representation of arithmetical behavior, a traffic rule is a representation of the behavioral negotiation of traffic, a rule for inferring is a representation of inferential behavior, and so on.

Wittgenstein (2009) paints a vivid picture of rules as representations in such passages as the following:

We use a machine, or a picture of a machine, as a symbol of a particular mode of operation. For instance, we give someone such a picture and assume that he will derive the successive movements of the parts from it. (Just as we can give someone a number by telling him that it is the twenty-fifth in the series 1, 4, 9, 16, ...) (§193)<sup>12</sup>

The picture of the machine, by analogy to the rule, represents the real-world movement of the machine, the behavioral output, as patterned. To infuse the machine's performance with the picture of the machine, or by analogy to infuse the behavior of the person continuing the ascending series of perfect squares with the rule for continuing the series, is to mistake the representation for what it represents – a use-mention conflation of sorts.

---

<sup>11</sup>In his Presidential Address before the Pacific Division of the American Philosophical Association, Albritton (1985) persuasively argues that the will's freedom isn't restricted by the inability to act accordingly. It is easy to mistake constraints on what we do for constraints on our deciding what to do. If I'm prevented from moving around freely by being held down, this *by itself* doesn't prevent me from choosing to move around freely. And a deficiency or inability to do something shouldn't be read backwards, as it were, into an alleged earlier deficiency or inability to choose to do it. Along similar lines, a deficiency or inability to behave according to an inferential license shouldn't be read backwards into some deficiency or inability in the earlier acceptance of the license in question.

<sup>12</sup>See also the discussion in Wittgenstein (2009: §§193-194) and (1978: §§I-122-130).

If rules of conduct are representations of behavior as patterned, the license captured by (2.1) in the first iteration of Carroll’s regress represents a certain pattern of proceeding from (1) and (2). The argument (1),(2),(2.1)/(3) is the same as (1),(2)/(3) but for the inclusion in the former of (2.1). The mode of inference is the same in both, the only difference being that (1),(2),(2.1)/(3) includes (2.1) as an additional (and idle) premise. Once we see the rules as representations of inferential behavior, we no longer need to assume that they require incorporation into behavior in some mysterious way, or that described rules aren’t really rules, as Sellars puts it. The truth of the matter, on the contrary, is that insofar as rules are representations of behavior, they can also be described in turn. To speak of rules as “lived” is to skirt dangerously close to mistaking the representation for what it represents. The behavior is lived of course, but it, in turn, is represented by the rule. As a response to the Tortoise of the original story, the emphasis on rules as lived is unsuccessful. At each stage of the regress the Tortoise supposedly accepts a rule that enjoins that the desired conclusion follows from the extant premise-set. The explanatory burden then shifts to the skeptical proponent of the Ryle-Sellars line to say why such acceptance is ineffectual. In his Presidential Address before the Aristotelian Society Ryle (1946) describes the perils of failing to heed the distinction between knowing that and knowing how, and more specifically between “accept[ing] rules in theory” and “apply[ing] them in practice”, by adding, parenthetically: “This is Lewis Carroll’s puzzle in ‘What the Tortoise said to Achilles’. I have met no successful attempt to solve it” (6). Treating Carroll’s regress as a serious difficulty is predicated on succumbing to a mythology of rules as somehow operative in behavior.<sup>13</sup> Regarding rules as representations of behavior instead defuses such worries. The Tortoise fails to conclude (3). Why the Tortoise so fails may be an interesting question about tortoise psychology but need not culminate in a worrisome skeptical threat about rules. The Tortoise’s failure is the failure to draw the relevant conclusion – end of story. And it isn’t only the end of the story in the tweaked version via (2.1’). It is the end of the story in its original formulation as well. The detour through self-citation was needed to dramatize the incoherence in the Tortoise’s claimed attitude towards rules. But that attitude was incoherent all along.

---

<sup>13</sup>A sophisticated reading of Ryle’s distinction between intelligent behavior and mechanical habit due to Bäckström and Gustafsson (2017) characterizes the Rylean distinction as a formal distinction in category, where a category delineates the range of claims and questions that make sense regarding the item at issue. Such a reading renders the Tortoise as someone who, appearances to the contrary notwithstanding, is simply not behaving intelligently. Why Ryle (1946: 26) says, regarding the problem we’ve been discussing, “I have met no successful attempt to solve it”, remains unclear.

## 5 Conclusion

An inferential license *qua* rule of inferential conduct is a representation of inferential behavior. But rules of inference also play a crucial role in setting up formal systems in logic. A formal language is specified by providing an inventory of basic signs and “formation rules” that specify which strings of signs qualify as formulas. A formal system is specified by providing “transformation rules” of two kinds: axioms and rules of inference. A formal system as a whole can represent a type of real-world reasoning. The axioms can represent where reasoning of the represented type may begin without further justification. The rules of inference can represent how the reasoning may in fact proceed. Once we attend to such matters of representation we are less likely to suppose that rules of inference are somehow invariably operative in the inferential behavior being modeled by the formal system as a whole. The thought that rules of inference are operative in this way blurs the distinction in level between the representation and what it represents.

Lewis Carroll and his contemporaries, including, notably, the Russell of *The Principles of Mathematics* (1903: Ch.3), did not see things in this way. The attempt to meet the Tortoise’s reticence by enriching the set of premises of an obviously valid argument to include the relevant inferential license is but one detail of their shared commitment to a single-level approach to logic. The fully mature idea of a formal system had to wait until the transformative work of Carnap, Gödel and others in the ensuing decades.<sup>14</sup> But once the idea of a formal system became prominent in logic, another idea became prominent in philosophy – that insofar as formal systems represent swaths of reasoning, each significant aspect of those representations is itself representational.<sup>15</sup> It is but a short step from here to the thought that a rule of inference of a formal system stands for something operative in the reasoning represented by the system as a whole. The view then impels the further insistence that whatever is thus operative must not be included as a mere premise on pain of regress. But this, as we saw, including as it does the problematic idea of rules as somehow incorporated into rule-governed behavior, is not independently motivated.

Finally, there may be yet another thought lurking behind the insistence that rules must somehow be operative in rule-governed behavior. It is the conviction that whatever else rule-governed behavior may be, surely the rule discloses an essential aspect of the behavior, part of its very nature, what makes rule-governed behavior

---

<sup>14</sup>Further details on the contrast between the later focus on formal systems in logic and the earlier conception of the subject may be found in van Heijenoort (1967) and Goldfarb (2005).

<sup>15</sup>But see Kaplan’s (1975: 722) notion of “artifact of a model” for a compelling critique of this move from a model being representational to every aspect of the model being representational.

what it is. Consider, for example, the following passage from Sellars (1949):

We distinguished above between action which merely conforms to a rule and action which occurs because of a rule and pointed out that in so far as actions merely conform to it, a rule is not a rule but a mere generalization. On the other hand, we must not say that a rule is something completely other than a generalization. The mode of existence of a rule is as a generalization written in flesh and blood, or nerve and sinew, rather than in pen and ink. (299)

On such a view, the rule is supposed to disclose the nature of the behavior it governs not unlike the way the physiology of a bodily process reveals what it really is at bottom. But why think the rule should so reveal what the behavior it governs really is remains unclear. We may study various forms of behavior and identify rules that govern them. The rules we identify are representations used for a variety of purposes. The idea that those representations are inevitably entrusted with more than their theoretical or practical roles as representations of behavior, namely, with nature-disclosure when it comes to being governed by rules, has little to recommend it.

There is in fact little reason to suppose that representing behavior by specifying a rule tells us what the behavior is in the most demanding sense. A behavioral pattern is a regularity in human behavior; it can be studied from a variety of explanatory perspectives at various scales. Nothing suggests that the rule we describe ourselves as operating under, to ourselves or to others, in performing simple addition reveals what the practice of addition really is at bottom. The same goes for inferential behavior. What those behaviors are can be approached from a variety of angles, some personal (as when we explore matters of justification, for example), some sub-personal (as when we explore implicated cognitive processes at the individual level), and yet others super-personal (as when we explore advantages of certain practices over alternatives at a social level). The standard rule for addition often belongs to the first personal take on the practice, perhaps a representation for pedagogical purposes. It isn't the job of such a representation to tell us what the practice of adding really is in the most demanding sense. Indeed, even for standard scenes of instruction, where rules are given to novices as means of initiating them into and guiding them through the practice, the rules don't disclose what exactly is to be done in the most demanding sense. A rule for addition, for example, doesn't tell us whether in adding 57 to 68 we should first add the tens, then the units, and then add the results; or add the units first and carry the one over to the tens; or add from left to right (or top to bottom in a column) or from right to left (bottom to top); or what. Rules for inferring likewise don't tell us the order in which we should take

multiple premises into account. Rules are silent, as can only be expected, on the many ancillary details that are relevant for actual behavior that accords with them. Be that as it may, thinking of rules as representations of behavior doesn't allow the skeptical problem we've been discussing to get off the ground.<sup>16</sup>

---

<sup>16</sup>For discussion of earlier versions of this material, I am indebted to Paul Egré, Robin Bianchi, Pascal Engel, Jeremy Fantl, Kathrin Koslicki, Dave Liebesman, John Woods, Marc Ereshefsky, Yudi Huang, Simon-Pierre Chevarie-Cossette, Nicole Wyatt, Antoine Taillard, Richard Zach, and two anonymous referees. Thanks are also due to audiences at the Institute of Philosophy at Neuchâtel, at the Logic4Peace conference, and at the University of Calgary.

## References

- Albritton, Rogers. (1985) 'Freedom of the Will and Freedom of Action', *Proceedings and Addresses of the American Philosophical Association* 59: 239-251.
- Bäckström, Stina and Gustafsson, Martin. (2017) 'Skill, Drill, and Intelligent Performance: Ryle and Intellectualism', *Journal for the History of Analytical Philosophy* 5: 40-55.
- Carroll, Lewis. (1895) 'What the Tortoise Said to Achilles', *Mind* 4: 278-280.
- Cook, Roy. (2014) 'There is No Paradox of Logical Validity', *Logica Universalis* 8: 447-467.
- Goldfarb, Warren. (2005) 'On Gödel's Way In: The Influence of Rudolf Carnap', *Bulletin of Symbolic Logic* 11: 185-193.
- van Heijenoort, Jean. (1967) 'Logic as Calculus and Logic as Language', *Synthese* 17: 324-330.
- Imholtz, Clare and Moktefi, Amirouche. (2016) 'What the Tortoise Said to Achilles: A Selective Bibliography', *The Carrollian: The Lewis Carroll Journal* 28: 125-127.
- Kaplan, David. (1975) 'How to Russell a Frege-Church', *Journal of Philosophy* 72: 716-729.
- Russell, Bertrand. (1903) *The Principles of Mathematics* (London: George Allen & Unwin).
- Ryle, Gilbert. (1946) 'Knowing How and Knowing That', *Proceedings of the Aristotelian Society* 46: 1-16.
- Sellars, Wilfrid. (1949) 'Language, Rules and Behavior', in Hook, S. (ed.), *John Dewey: Philosopher of Science and Freedom* (New York: Dial Press): 289-315.
- Simchen, Ori. (2001) 'Rules and Mention', *Philosophical Quarterly* 51: 455-473.
- Simchen, Ori. (2013) 'Token-Reflexivity', *Journal of Philosophy* 110: 173-193.
- Wittgenstein, Ludwig. (1978) *Remarks on the Foundations of Mathematics*, Revised Edition (Cambridge, MA: The MIT Press).
- Wittgenstein, Ludwig. (2009) *Philosophical Investigations*, 4<sup>th</sup> Edition (Oxford: Wiley-Blackwell).