

3 Reasoning, Rules, and Representation

Paul D. Robinson and Richard Samuels

Introduction

Regress arguments have had a long and influential history within the philosophy of mind and the cognitive sciences. They are especially commonplace as a bulwark against representational or *intentional* theories of psychological capacities. For instance, arguments of this sort played a prominent role in debates concerning the theory of transformational grammar (Chomsky, 1969a; 1969b; Harman, 1967, 1969), the language of thought hypothesis (Fodor, 1975, 1987; Laurence & Margolis, 1997), the massive modularity hypothesis (Fodor, 2000; Collins, 2005), and intentional accounts of intelligent activity quite broadly (Ryle, 1949; Fodor, 1968; Dennett, 1978). Typically, the regress is presented as one horn of a dilemma:

To explain the manifestation of some kind of capacity, *C*, the theorist postulates an (intentional) psychological process of kind *P*. But, the critic suggests, the successful operation of any *P* process itself depends upon some prior manifestation of *C*. Thus: Either it is necessary to postulate a second psychological process of kind *P*, and so on, *ad infinitum*, or alternatively, one must grant that *C* can be explained without positing *P*.

Thus the proposed intentional theory is either broken-backed or redundant. Or so proponents of regress arguments would have us believe.

Recently, a similar argument—which we simply call the *Regress*—has surfaced in philosophical debate regarding the nature of reasoning. Participants in this debate are not concerned with everything that gets called ‘reasoning’. Rather, they focus on a relatively circumscribed range of reasoning-like phenomena—which they call *active reasoning* or *inference*¹a kind of

1 Three comments regarding terminology. First, as is common in the present context (but see Broome, 2013, 292), we use ‘active reasoning’ and ‘inference’ interchangeably. Second, although it is slightly infelicitous to use ‘inference’ in this restricted sense, it should be read as such unless explicitly modified—e.g., as in *sub-personal* inference. Third, as is typical, we take it to be true, more-or-less by definition, that a process or activity is active only if it is person-level. As such, we count no sub-personal processes as active.

person-level, conscious, voluntary activity, which at least in paradigmatic instances results in the fixation of belief. It is widely assumed that active reasoning in this sense is fairly pervasive among human beings; that it can involve attitudes with markedly differing contents; that simple, consciously made, deductive inferences are a prototypical case; and that errors in active reasoning are both possible and, indeed, fairly commonplace. For philosophers interested in active reasoning, then, the core explanatory challenge is to provide an illuminating account of the nature of this psychological capacity.

Within this context, the presumed significance of the Regress is that it (allegedly) undermines a family of highly influential accounts of inference—what might be called *intentional rule-following* (or IRF) theories. To a first approximation, such theories make a pair of commitments. First, they suppose that inference essentially involves following rules concerning the premises from which one reasons:

(Rule-Following View): All active reasoning involves rule-following operations.

In addition, they impose the following necessary condition on rule-following:

(Intentional View): All rule-following involves intentional states, which *represent* the rules being followed.

In brief, the Regress purports to show that if such accounts were correct, *any* instance of active reasoning—no matter how apparently simple—would be a *supertask* involving an infinite number of rule-following operations. In which case, contrary to fact, it would be impossible for finite creatures such as us to actively reason.

If the Regress were sound, it would have serious implications for philosophical debate regarding the nature of inference. What may be less obvious is that it would also have significant consequences for scientific theories of reasoning, and cognition more broadly. Within the psychology of reasoning, quite generally, and the psychology of *deductive* reasoning in particular, it is commonplace to suppose that reasoning relies on mentally represented rules.² This commitment is perhaps most apparent in *mental logic* accounts, where it is explicitly hypothesized that there are “deduction rules that construct mental proofs in the system’s working memory” (Rips, 1994, 104). But the commitment is also apparent among dual-process theorists who routinely suppose

2 It is worth noting, in this regard, that philosophers writing on inference have tended to focus on cases in which we reason in accordance with logical rules, such as modus ponens. Further, Boghossian (2008, 499) claims that whereas denying the Rule-Following View of reasoning in general seems false, denying the Rule-Following View of deductive reasoning in particular seems ‘unintelligible’.

that System 2 processes involve intentional states that represent rules (e.g., Sloman, 1996). Moreover, we suspect—though won't argue here—that even those who explicitly reject the mental logic approach *also* presuppose the existence of intentional states that represent rules. For example, *mental models* accounts seem to presuppose the existence of such states, albeit where the presumed rules are for the manipulation and inspection of iconic models denoting possibilities, as opposed to the construction of mental proofs through chaining linguistic entities such as sentences (e.g., Johnson-Laird, 2008).

Of course, such accounts of reasoning are contentious and may turn out to be false. But on the face of it, this should be an empirical issue, addressed by empirical means. If the Regress is sound, however, such theories should be rejected a priori. Further, as we will show, since the Regress does not turn essentially on assumptions about the nature of *active* reasoning per se, the argument, if sound, would apply to a far broader class of phenomena. Specifically, as we will see, it would apply, with minimal modification, to processes that are unconscious and sub-personal, and, hence, not active. If sound, then, the Regress would have ramifications for a wide array of theories in many regions of cognitive science, including theories of perception.

Fortunately, the Regress is not sound. Formulations of the argument are invariably underspecified, and once presented in suitably perspicuous fashion, it becomes clear that the Regress relies on assumptions no sensible version of IRF should endorse. The primary burden of this chapter is to show why this is so.

Here's how we proceed. In Section 1, we explain the IRF account of reasoning in more detail, and set out some of its *prima facie* virtues. In Section 2, we aim to explain the general structure of the Regress and provide the most charitable formulation of the argument that we can. In Section 3, we discuss a standard—and we think correct—response to this original regress: to posit sub-personal inferential processes. We show that this response provides a plausible way to block the original regress. But following suggestions from Boghossian and others, we also a) show how to develop a Revenge Regress, which targets IRFs about sub-personal processes, and b) explain how to use this result to develop a Strengthened Regress, which fills the gap in the original argument. Finally, in Section 4 we explain why the Strengthened Regress is still subject to a serious objection, and in Section 5 we address two responses to this objection.

1. The Virtues of Intentional Rule-Following Accounts of Inference

The IRF is not so much a single account of inference as a family of proposals that share a common commitment to the Rule-Following View of inference and to the Intentional View of rule-following. In our view, such proposals merit serious consideration because they possess a host of explanatory virtues. We are especially sympathetic to variants of IRF that incorporate some form of computationalism about mental processes—a class that includes the

sort of ‘classicism’ advocated by Fodor and Pylyshyn (1988), versions of connectionism (e.g., Smolensky, 1988), and some recent Bayesian approaches to cognitive modeling (e.g., Perfors, Tenenbaum, & Regier, 2011). These frameworks are among the most plausible extant approaches to the study of higher cognition in general, and reasoning in particular.

Although this is not the place to discuss the virtues of IRFs in detail, a brief reminder should make clear that much is at stake if the Regress is sound. First, consider some of the *prima facie* explanatory virtues that accrue merely as a result of adopting the Rule-Following View (cf. Boghossian, 2014, 4, 12):

- *A theory of reasoning should discriminate reasoning from mere causation by belief (and other intentional states).* Not all instances of beliefs causing other beliefs are inferences. Notoriously, there are “deviant” causal chains involving beliefs that are obviously non-inferential.³ The rule-following account helps to explain the difference. Very roughly, in the case of inference, the influence of belief is wholly mediated by rule-following operations and in the other cases, not.
- *Since not all reasoning is good reasoning, we should prefer, on grounds of generality, an account that covers both the good and the bad.* Rule-following accounts can capture this desideratum. On such views, one can reason badly, either by following a bad rule or by making mistakes in one’s attempt to follow good rules. In contrast, good reasoning only occurs when one correctly follows a good rule.
- *A theory of reasoning ought to explain the sorts of generality that are exhibited by inference.* For example, it is widely recognized by philosophers and psychologists that we are capable of reasoning about an exceedingly broad array of topics—roughly, any topic for which we possess concepts. Moreover, our inferences often exhibit similar patterns or ‘logical forms’ across these various topical domains. Rule-following accounts provide promising explanations of such phenomena. Specifically, if some inferential rules are akin to logical rules in being largely ‘content independent’ or ‘formal’, then we have a partial explanation of why we are able to reason about so many different subject matters. Further, if we suppose that humans follow these rules in lots of different contexts, we will have an explanation of why inferences in different domains exhibit similar forms.
- *A theory of active reasoning should both subsume and explain the difference between deductive and inductive reasoning.* Once again, the rule-following picture offers a natural account. When reasoning deductively, the relevant rule-following operations involve deductive rules, and when one reasons inductively, the relevant rules are inductive ones.

3 Example: Suppose John believes that he’s late for class and that this realization makes him sweat. If on the basis of this experience he came to believe that he was sweating, we could have a case of causation by belief, but not inference.

No doubt there are other issues that the Rule-Following View might help address, but let's turn to the Intentional View. As we see it, there are two deep and closely related explanatory motivations for this view. The first is what we call the *Guidance Problem*. The aforementioned explanatory virtues of the Rule-Following View all turn on the assumption that rules can in some sense *guide* our cognitive activities. But how is this possible? After all, a rule *qua* rule is 'just an abstract object' and so presumably incapable of exerting any causal influence (Boghossian, 2014, 13).

Here's where the Intentional View enters the picture. Although rules as such cannot guide cognition, intentional states that encode or represent such rules can. For in addition to their representational properties, intentional states have other properties that are causally relevant—various physical and structural properties, for example. On the Intentional View, then, rules guide behavior in an attenuated sense: they are the contents of intentional states—rule-representations—that are causally implicated in reasoning.⁴

A second and related virtue of the Intentional View is that it helps resolve a very old problem for rule-following accounts of cognition. In brief, such accounts presuppose a distinction between following a rule and mere *accordance* with a rule (see e.g., Hahn & Chater, 1998, 203f.). Without such a distinction, rule-following *per se* will be of little use in explaining what is distinctive of reasoning. For it will turn out that all processes describable by a rule—that is, all processes that display regularity in their behavior—are rule-following processes. In which case, it will be no more true of reasoning that it involves rule-following than it will be of, say, the planets that they 'follow' a rule when conforming to Kepler's laws of planetary motion (Fodor, 1975).

Again, we think that the Intentional View provides a credible approach to this problem. According to this approach,

what distinguishes what organisms do from what the planets do is that a *representation of the rules they follow constitutes one of the causal determinants of their behavior.*

(Fodor, 1975, 74)

4 Although it does not require endorsing the Intentional View in its full generality, it is also worth noting that the idea that rules are encoded by intentional states helps explain what is otherwise a puzzling fact about human beings—namely, we are capable of learning rules that influence our behavior on the basis of "one-shot" instruction or linguistic communication. For example, if, at passport control, the guard tells me, "Stand behind the yellow line, until you are called", I stand behind the yellow line and wait to be called! On the basis of one exposure to instruction, my behavior is modified so that I follow the rule. This is readily explained on the assumption that, on the basis of linguistic processing, I come to possess one or more intentional states that represent the content of the guard's utterance—i.e., the rule.

In contrast, where the planets are concerned,

at no point in a causal account of their turnings does one advert to a structure which encodes Kepler's laws and causes them to turn. The planets *might* have worked that way, but the astronomers assure us that they do not.

(ibid.)

In summary, if the Intentional View is correct, we have a prima facie plausible way both to resolve the Guidance Problem and to draw the rule-following/rule-accordance distinction. Moreover, since a solution to these problems is a prerequisite for the Rule-Following View to have any explanatory value, it is exceedingly attractive to combine the Intentional and Rule-Following views in the manner proposed by IRFs.

Of course, the aforementioned is defeasible, and matters would be quite different if there were powerful independent reasons to reject IRFs. With this in mind, we turn to the Regress.

2. The Regress

Although a number of theorists have invoked variants of the Regress, Boghossian's discussion strikes us as the most perspicuous, to date, and for this reason, we focus primarily on it here. In Section 2.1, we lay out Boghossian's general strategy. In Section 2.2, we explain how he aims to establish a crucial premise of the argument, what we call the Rule Application Condition. Then in Section 2.3, we sketch the Regress itself, and in Section 2.4, we provide a more regimented formulation of the argument.

2.1 The General Strategy

It is important to distinguish the Regress, advocated by Boghossian and others, from a range of superficially similar worries. The relevant regress is *not* an epistemic one. It is not, for example, a regress with respect to justification, or reasons for belief. Nor is it a regress concerning the determination of meanings or contents of the sort associated with Kripke's Wittgenstein (Kripke, 1982). Finally, it is not a definitional regress wherein the definiendum—'inference'—is to be defined in terms of 'rule-following', which in turn is to be defined in terms of 'inference', and so on. Rather, the problem allegedly posed by the Regress is a regress of mental *operations*. The worry, in brief, is that IRFs place active reasoning beyond the grasp of finite creatures by turning every instance of inference into a *supertask*: an infinite sequence of rule-following operations to be performed in a finite period of time.

Here is the general strategy: to generate the desired regress, Boghossian seeks to show that the IRF entails the following interlocking pair of conditions:

- Condition 1:* Each inference I_i requires a rule-following operation R_i .
Condition 2: Each rule-following operation R_j requires some further inference I_j .

Given these conditions, we can generate a regress of mental operations by cycling between them:

- Suppose I draw inference I_1 ;
 by Condition 1: I perform a rule-following operation R_1 ;
 by Condition 2: I draw an inference I_2 .
 by Condition 1: I perform a rule-following operation R_2 ;
 by Condition 2: I draw an inference I_3 .
 And so on . . .

A regress of mental operations ensues. In which case, if the conjunction of the Rule-Following View and Intentional View entail these conditions, then IRFs turn all inferences into supertasks.

2.2. Establishing Condition 2

How does Boghossian seek to establish that IRFs are committed to Conditions 1 and 2? Since the Rule-Following View *asserts* that all inference involves rule-following operations, Condition 1 is easily secured. Condition 1 is just a rendering of the Rule-Following View. In contrast, neither the Rule-Following View nor the Intentional View asserts Condition 2. Boghossian's main argumentative burden, then, is to show that they entail it.

How is this to be done? Rather than focusing on rules of inference, Boghossian initially discusses a simple *decision* rule with the aim of drawing out some general morals regarding what, on the Intentional View, would be required for active, person-level rule-following:

Suppose I receive an email and that I answer it immediately. When would we say that this behavior was a case of following:

(Email Rule) Answer any email that calls for an answer immediately upon receipt!

as opposed to just being something that I happened to do that was in conformity with that rule?

Clearly, the answer is that it would be correct to say that I was following the Email Rule in replying to the email, rather than just conforming to it, when it is because of the Email Rule that I reply immediately.

(2014, 13)

Of course, this immediately raises an instance of the Guidance Problem: what is it to follow this rule, as opposed to merely *conforming* to it? Since the E-mail Rule ‘qua rule, is just an abstract object’ it cannot *directly* guide behavior. Instead, if the Intentional View is correct,

my behavior is to be explained via some state of mine that represents or encodes that rule.

(Ibid.)

Merely positing such an intentional state does not, however, fully explain this particular instance of rule-following activity. There also needs to be a process in which this rule-representation might figure so as to guide my behavior. And, according to Boghossian, it is plausible that this process conforms to the following pattern:

I have grasped the rule, and so am aware of its requirements. It calls on me to answer any email that I receive immediately. I am aware of having received an email and so recognize that the antecedent of the rule has been satisfied. I know that the rule requires me to answer any email immediately and so conclude that I shall answer this one immediately.

(Ibid.)

Of course, this is only one specific instance of rule-following activity. Nevertheless, Boghossian takes it to illustrate what, on the Intentional View, active rule-following *in general* would require:

On this Intentional construal of rule-following, then, my actively applying a rule can only be understood as a matter of my grasping what the rule requires, forming a view to the effect that its trigger conditions are satisfied, and drawing the conclusion that I must now perform the act required by its consequent.

(Ibid.)

Notice—and this is the crucial point—that this appears tantamount to claiming that “on the Intentional view of rule-following, rule-following requires inference” (ibid.). More precisely, Boghossian appears to be insisting that, on the Intentional View,

Rule Application Condition: For a person-level rule-following process to utilize a rule-representation, it must contain an inferential sub-process—an inference from the rule to what the rule calls for under the circumstances.

And, of course, if this is true, then so too is Condition 2. That is, the Rule Application Condition entails that each rule-following operation R_j requires some further inference I_j .

2.3. *The Regress Within Reach*

If the aforementioned is correct, then an *intentional* rule-following account of inference is committed to both Conditions 1 and 2. As Boghossian puts it,

On the one hand, we have the Intentional View of rule-following, according to which applying a rule always involves inference. On the other hand, we have the Rule-Following picture of inference according to which inference is always a form of rule-following.

(2014, 14)

Further, if this is so, then it would seem that any instance of person-level rule-following must involve an infinite series of further rule-following operations. If, for example, I actively follow the E-mail Rule, then I must draw an inference in order to follow it, and since, by assumption, this involves rule-following, I must draw another inference, which requires another instance of rule-following, and so on *ad infinitum*. Thus Boghossian concludes,

These two views . . . can't be true together. Combining the two views would lead us to conclude that following any rule requires embarking upon a vicious infinite regress in which we succeed in following no rule.

(2014, 14)

Boghossian is not alone in drawing this pessimistic conclusion. For instance, for the same reason, Wright claims that if rule-following requires a state that carries a content that licenses the inferential transition, then it is “uncertain that any coherent—regress-free—model can be given of what inferring actually is” and hence,

we must drop the idea that inference is, everywhere and essentially, a kind of rule-following. That, in outline, is the solution to the problem of the Regress.

(2014, 32f.)

Similarly, Broome claims that if rule-following requires an explicit representation of a rule, then

you would have to determine whether each particular case of potential reasoning falls under the rule. Doing so would require reasoning, which would again require following a rule. There would be a circle.

(2014, 632)

In short, some very influential philosophers maintain that the Regress undermines IRF.

2.4. *The Regress Regimented*

With the earlier exegetical work complete, we are now in a position to set out the Regress in full dress. Our aim is to capture the details and spirit of Boghossian's discussion as charitably as possible, though without logical lacunae. What follows is our best effort.

The Regress proceeds from the assumption that IRF is true, to the untenable conclusion that active reasoning—or inference—is impossible for finite creatures. And since the Regress targets IRFs about *active* reasoning, it is natural to formulate the Rule-Following View and Intentional View in person-level terms. That is,

- (1) Any process of inference is a kind of person-level rule-following.
- (2) Any process of person-level rule-following utilizes a person-level rule-representation.

Here (1) and (2) clearly characterize a version of IRF about inference, or active reasoning. But as we saw earlier, without additional premises, they do not suffice to generate a regress. Rather, one must further maintain a version of what we earlier called the Rule Application Condition:

- (3) For a person-level, rule-following process to utilize a personal-level rule-representation, it must contain an inferential sub-process—a person-level inference from the rule to what the rule calls for under the circumstances.

Here 'sub-process' is to be understood as referring to a *proper* part of the person-level, rule-following process.⁵ It follows from (1), (2), and (3) that

- (4) Any inferential process involves an inferential sub-process.

Since any such inferential sub-process is itself an inference, it will also involve an inferential sub-process, and so on, *ad infinitum*. Hence by iteration on (4) we have

- (5) Any inference requires infinitely many inferential sub-processes.

But given that the performance of infinitely many inferences cannot be carried out in finite time, it follows from (5) that

- (6) Inference is impossible for finite beings like us.

⁵ This is required to block an interpretation of (3) on which the rule-following process is *identified with* the inference from the rule to what the rule calls for. For if such an identification is made, (3) will not generate a regress of operations.

Yet active reasoning *is* possible for creatures such as us. At any rate, this is what Boghossian, Broome, Wright, and almost *everyone*—including the present authors—suppose.⁶ In which case, on the assumption that the IRF is true, we appear to have reason to reject the IRF.

3. Sub-personal Processes, Revenge, and the Strengthened Regress

3.1 *Getting Sub-personal*

How should proponents of IRFs respond to the Regress? If premises (1)–(3) are true, then regress ensues. But it is plausible to reject premise (2) in favor of a weaker requirement. Specifically, proponents of IRFs may allow that person-level rule-following *sometimes* involves person-level rule-representations, while insisting that *sub-personal* rule-representations may also play the requisite role. The resulting variant of the Rule-Following View can be formulated as follows:

- (2***) Any process of person-level rule-following utilizes a rule-representation that is either personal or sub-personal.

This modification evades the Regress. Moreover, it does so in an independently motivated and independently plausible fashion. Although the personal/sub-personal distinction is a notoriously vexed one (see Drayson, 2012), for present purposes, the crucial requirement is—as Boghossian recognizes—that person-level processes are “processes of which we are, in some appropriate sense, aware” (2008, 483). In contrast, sub-personal states are “not consciously accessible to the thinker” (2014, 15). Yet if this is how we are to draw the personal/sub-personal distinction, it should be clear that any plausible IRF will need to insist that person-level rule-following quite typically involves *sub-personal* rule-representations. This is because, as Boghossian, Broome, and many others recognize, active reasoners very typically *lack* conscious awareness of following a rule. In which case, proponents of IRFs have exceedingly good reason to insist that rule-representing states are often sub-personal. Moreover, this has nothing to do with regress worries *per se*. Rather, it is mandated by the antecedent assumption that active reasoning is a commonplace cognitive activity, along with the overwhelmingly plausible empirical claim that active reasoners very frequently lack conscious awareness of any relevant rule or rule-representing state. More generally, the point is that once a theory posits representational states to explain a cognitive phenomenon, the hypothesized states must be sub-personal if the agent lacks conscious awareness of them. In this regard, proponents of IRFs are

⁶ Presumably, some eliminativists, behaviorists, and the like would deny this.

in the same predicament as psycholinguistics who posit sub-personal representations of syntactic rules, or vision scientists who posit sub-personal representations of edges. And, in our view, this is not bad company to keep.

3.2. *The Revenge Regress*

We have argued that weakening (2) in the proposed manner both evades the earlier regress and is independently plausible. Nonetheless, the proponent of IRF is not out of the woods yet. For, as critics note, a closely analogous sub-personal regress can be generated. Thus Boghossian maintains,

In the present context, going sub-personal presumably means identifying rule- acceptance . . . not with some person-level state, such as an intention, but with some sub-personal state . . . Let us say that [such a state] is some sub-personal intentional [i.e., representational] state in which the rule's requirements are explicitly encoded. Then, once again, it would appear that some inference (now sub-personal) will be required to figure out what the rule calls for under the circumstances. And at this point the regress will recur.

(2008, 498)

The core insight of the earlier passage is that merely extending the Intentional View to cover sub-personal rule-following does little to alter the overall structure of the IRF. In which case, one might think that if utilizing a person-level rule-representation requires an inferential sub-process, then utilizing a sub-personal rule-representation will also require an inferential sub-process—albeit a *sub-personal* one. And if this is so, then we can generate a *Revenge Regress* that mirrors the original:

- (1*) Any sub-personal inference is a kind of sub-personal rule-following.
- (2*) Any process of sub-personal rule-following utilizes a sub-personal rule-representation.
- (3*) For a sub-personal rule-following process to utilize a rule-representation, it must contain an inferential sub-process—a *sub-personal* inference from the rule to what the rule calls for under the circumstances.

From (1*– 3*) it follows that

- (4*) Any sub-personal inference involves a sub-personal inferential sub-process.

And since any such inferential sub-process is itself a sub-personal inference, by iteration on (4*) we may infer the following:

- (5*) Any sub-personal inference requires infinitely many sub-personal inferential sub-processes.

Finally, given that the performance of infinitely many sub-personal inferences cannot be carried out in finite time, it follows from (5*) that

- (6*) Sub-personal inference is impossible for finite sub-personal systems like ours.

3.3 *Regress Strengthened*

No doubt, this conclusion will be welcome to those already suspicious of intentional explanations of the sort found in cognitive science. For the proponent of IRF, however, the Revenge Regress is exceedingly unfortunate. Supposedly, by allowing for sub-personal rule-representation, IRFs have an independently plausible way to escape the original Regress. But if the Revenge Regress is sound, the escape route is blocked, and the IRF is left without a way to account for active reasoning.

In our experience, the significance of the Revenge Regress is not always clearly appreciated. One problem is that it targets a different phenomenon from the earlier Regress—i.e., *sub-personal* inference. Why, then, should it be relevant to theories of *person-level* inference? Another problem is that proponents of the Regress never spell out in detail how the Revenge Regress interacts with the original one in order to strengthen the case against IRFs further. In view of this, it would be helpful to fill the gap by showing how to combine the Revenge Regress with the original argument in order to develop a *Strengthened Regress*. Again, here is our best effort.

First, assume the Rule-Following View:

- (1) Any process of inference is a kind of person-level rule-following.

Next, in view of the response to the original Regress, reject (2) and replace it with

- (2**) Any process of person-level rule-following utilizes a rule-representation that is either personal or sub-personal.

Now we require two variants of the Rule Application Condition. The first we retain from the original argument:

- (3) For a person-level, rule-following process to utilize a personal-level rule-representation, it must contain an inferential sub-process—a person-level inference from the rule to what the rule calls for under the circumstances.

However, the replacement of (2) by (2**) requires that we supplement it with another variant of the Rule Application Condition:

- (3**) For a personal-level, rule-following process to utilize a *sub-personal* rule-representation, it must contain an inferential sub-process—a

sub-personal inference from the rule to what the rule calls for under the circumstances.

The crucial difference between (3) and (3**) is, of course, that the former specifies what is involved in using person-level rule-representations, whereas the latter specifies what is involved when actively reasoning with sub-personal rules. These premises commit the intentionalist not to (4) but to

(4**) Any person-level inference involves either a person-level or sub-personal inferential sub-process.

Moreover, the Revenge Regress still commits the intentionalist to

(4*) Any sub-personal inference involves a sub-personal inferential sub-process.

Suppose we try to carry out a process of active reasoning. By (4**), it involves a sub-process of either active reasoning or sub-personal inference. If it involves the former, then (4**) will also apply to that sub-process. Thus if successive iterations were always to lead to a further sub-process of active reasoning, then they would generate the original regress. But if at any stage active reasoning involves a sub-personal inference, then by iteration on (4*), the Revenge Regress is generated. So we have shown not (5) but rather

(5**) Any inference requires infinitely many person-level or sub-personal inferential sub-processes.

Hence for the by now familiar reason,

(6) Inference is impossible for finite beings like us.

QED.

4. Rejecting the Strengthened Regress

Although the Regress is widely supposed to show that IRFs are untenable, we maintain that such a view is unwarranted. Even in its strengthened form, the Regress is unsound.

Our first pass response to the Strengthened Regress is to reject (4*)—the claim that any sub-personal inference involves additional inferential sub-processes. We take it to be obvious that any remotely plausible theory of sub-personal inference—intentionalist, or otherwise—must reject this commitment, since it is viciously regressive all by itself. But, of course, (4*) is a consequence of premises (1*)–(3*) of the Revenge Regress. So if we are to reject (4*), it must be because one of those premises is false. Further, since our response to the original Regress was to advocate an IRF for sub-personal inference, we are committed to rejecting (3*) since (1*) and (2*)

simply describe the Rule-Following and Intentional Views as they apply to sub-personal inference. Our challenge, then, is to argue that it is legitimate to reject (3*).

How is this to be done? Premise (3*) is a sub-personal version of the Rule Application Condition. It maintains that sub-personal rule-following of the kind envisaged by the Intentional View requires an inferential sub-process from the rule to what the rule calls for under the circumstances. The obvious way to justify the rejection of (3*), then, is to explain how a sub-personal process might utilize a rule-representation without thereby containing an inferential sub-process. We think that this challenge can be met and indeed that the right response is an exceedingly familiar one.

4.1. *The 'Primitivist' Strategy*

Causal-explanatory regress is among the most commonplace theoretical challenges to intentional theories in cognitive science. It should come as no surprise, then, that cognitive scientists have a routine strategy for quashing such worries. In our view, this strategy works extremely generally, including for IRFs about sub-personal inference.

Obviously, the intentionalist about rule-following processes cannot maintain—on pain of regress—that rule-guided psychological processes *always* involve further rule-guided psychological processes. Yet the intentionalist need not make this commitment. Instead, they can—and often do—posit a level of *primitive* processing mechanisms. Such processors may take rule-representations as inputs. In which case, the primitive processes they subserve will, in a sense, be rule-guided—though only in the thin sense that a rule-representation is causally implicated in the process because it is an *input* to the processor. In contrast to non-primitive processes, however, primitive ones are *not* rule-guided in a richer sense. That is, they are not rule-guided in the sense that they involve *further* rule-guided or inferential sub-processes. Thus, if non-primitive processes—such as those involved in active reasoning—ultimately decompose into primitive ones, then we have a general view of psychological processes on which rule application regresses cannot occur.

4.2. *Primitive Processes and Reflexes*

The earlier primitivist proposal is, of course, exceedingly well-known (Block, 1995; Dennett, 1978; Fodor, 1968, 1987; Fodor & Pylyshyn, 1988; Pylyshyn, 1980). But to see how it helps address the Regress, it is useful to clarify the notion of primitive processes. We think that this is usefully done by comparing them with prototypical (monosynaptic) reflexes.⁷ Primitive

⁷ Or, at any rate, a caricature of reflexes.

processes are closely analogous to prototypical reflexes in two crucial respects and, importantly, disanalogous in another. A first point of similarity is their *automaticity*. Given the relevant input conditions, a reflex generates a fixed behavioral output. Knock a knee, and it flexes. Analogously, provide input to a primitive processor, and it too flexes automatically—though not to lift a knee, but to output a representation.

A second similarity is that, in contrast to non-primitive intentional processes, the input-output relations of both reflexes and primitive processes are not inferentially mediated. Given a blow to the knee, it flexes, and as far as we know, no intervening stage of the process involves an inference or rule-following operation. The same holds for primitive psychological processes. They too have no intermediate stages that involve further inference or rule-following.

Yet there is, of course, an important difference between prototypical monosynaptic reflexes, and primitive psychological processes. In the case of reflexes, such as the patellar or corneal reflexes, the input is not a representation. Crudely put, it is a mere physical magnitude—a stimuli. In contrast, for primitive processes to play their assigned role within an intentional psychology, it is necessary that their inputs are representational. Indeed, in the cases of interest here, it is necessary that they represent *rules*. Primitive processors of the relevant sort, then, must be automatic, non-inferentially mediated, rule-applicators. By virtue of being rule-applicators, they can underwrite an intentional account of sub-personal inference, and by virtue of being automatic and non-inferentially mediated, they evade the concern that the application of any rule requires a further inferential step. They thus provide an alternative model of sub-personal rule application, which permits the proponent of IRF to reject (3*).

4.3. *Primitive Processes and Stored Program Computers*

The aforementioned might well sound rather mysterious were we to lack any model of how sub-personal inference could “bottom out” in processes that are reflex-like and yet rule-guided in the thin sense outlined earlier. But we *do* possess a model of such processes. For what we are describing is closely akin to a core aspect of standard, stored program computers. Such computers take programs (rules) as input, and many of their sub-processes involve rule-governed sub-processes (inferences). But computers are organized in such a manner that sooner or later all this rule-governed activity decomposes in a set of reflex-like operations, which do not rely on any further inferential activity. Indeed, their possession of this characteristic is among the central reasons that the concept of a stored program computer became so important to cognitive science. For it provides a model of how a system can be rule-guided without thereby succumbing to regress problems (Dennett, 1982; Fodor, 1975, 2000). The notion of a primitive process is simply a generalization of this aspect of stored program computers,

formulated in a manner that remains neutral regarding the precise nature of the processors and operations involved in human cognition.

4.4. *Primitive Processes and Cognitive Architecture*

It is worth stressing that the aforementioned is little more than standard background theory in much of cognitive science. This is because a hypothesized set of primitive processes and operations is a core facet of what, by deliberate analogy with computer science, is ordinarily called *cognitive architecture*.

As Zenon Pylyshyn noted a very long time ago, a cognitive architecture consists, at least in part, of “those functions or basic operations of mental processing that are themselves not given a process explanation” (1980, 126). That is, they are psychological functions and operations that are not to be explained in terms of other *psychological* processes—specifically, processes that deploy rules and representations (*ibid.*). In this respect, Pylyshyn continues, they are quite unlike “cognitive functions in general . . . [which] are . . . explainable . . . in terms of rules and representations” (*ibid.*). Instead, primitive processes and operations are “appealed to in characterizing cognition” and “are themselves explainable biologically rather than in terms of rules and representations” (*ibid.*)

By broad consent, it is an empirical matter *which* specific cognitive processes and operations are primitive. That there *are* such processes and operations is, however, widely—and we think correctly—assumed to be a presupposition of any sensible intentional psychological science and for the very same reason that primitive operations are a prerequisite for any sensible version of IRF. Without such operations, regress ensues. Again, as Pylyshyn observed long ago, the positing of primitive processes and operations avoids “a regress of levels of interpreters, with each one interpreting the rules of the higher level and each in turn following its own rules of interpretation” (*ibid.*).

5. Counterarguments

Positing primitive sub-personal processes allows proponents of IRFs to reject the Rule Application Condition—(3*)—and thereby neutralize the Regress, even in its strengthened form. Yet, as already noted, this regress-blocking strategy is an exceedingly familiar one from cognitive science. So it is somewhat surprising that it receives so little attention in the literature on active reasoning.

Why might this be? One obvious possibility is that the cost of primitivism is in some way too high—that it staves off the Regress, only to raise other no less serious problems for IRFs. In this section, we conclude by briefly considering two possible problems of this sort, which are hinted at in the literature on active reasoning.

5.1. *Positing Primitive Processes Addresses the Strengthened Regress Only at the Expense of Succumbing to Well-Known Kripkensteinian Rule-Following Problems*

Primitive rule-applicators take rule-representations as inputs. But one might find it deeply puzzling how such inputs could have rules as their *contents*. How, for example, might an input determinately represent modus ponens as opposed to some other rule? The obvious suggestion is that it represents the rule in virtue of the effects it has on the processor itself—that it induces modus ponens-like behavior in the processor. Yet this suggestion appears to raise Kripke’s familiar Wittgensteinian worries about rule-following. Out of the frying pan and into the fire.

We are tempted to give the earlier worry short shrift. The Regress, as understood by its advocates (and by us), is entirely independent of Kripke’s problem. Our aim here has been to address the Regress. If Kripke’s problem remains unaddressed, so be it. That’s a problem for another day.

Of course, if primitivism generated special Kripkensteinian worries for IRFs, then this quick-fire response would ring hollow. But we deny that it has such consequences. First, Kripke’s problem is orthogonal to the issue of whether one adopts the primitivist proposal. If Kripke’s problem is a serious one for IRF, then it applies equally to primitive processes *and* non-primitive ones. Kripke’s problem concerns the possibility of internalizing a determinate rule, given that it is supposed to cover a potential infinity of cases. Assuming the Intentional View of rule-following, this reduces to the issue of what it is for an intentional state to determinately *represent* a specific, infinitary rule. Further, if IRF is correct, then both primitive and non-primitive processes rely on rule-representations of very much the same sort. In which case, it is hard to see why Kripkean concerns would not arise equally for both sorts of processes. In short, positing primitive processes should make no difference to whether or not Kripke’s problem undermines IRF.

Second, we deny that Kripke’s problem is especially troublesome for IRFs as such, whether or not they endorse primitivism. To be clear, IRFs are theories about a class of psychological processes—i.e., inferential ones. In contrast, as Boghossian notes,

Kripke’s problem arises against the backdrop of a naturalistic outlook relative to which it is difficult to see how there could be determinate facts about which infinitary rule I have internalized.

(2014, 13)

For proponents of IRF, to “internalize” a rule is to represent it. In which case, Kripke’s problem clearly arises for IRFs only when one further demands a *naturalistic* account of rule-representation. In contrast, the problem has no traction if one “waives naturalistic constraints”—e.g., by allowing for primitive facts regarding the content of rule-representing states.

Speaking personally, we are not much inclined toward this sort of non-naturalism. But that's beside the point. Our point is that Kripke's problem is not a problem for primitivism as such, or even IRFs as such. Indeed, it is not a problem about psychological processes at all. Rather, it is a problem for naturalistic theories of *content*. Moreover, it is one that arises for them entirely independently of issues to do with *rule-following*. If one accepts that we can so much as *think* about determinate, infinitary rules—e.g., modus ponens—then the problem arises for naturalistic theories of content.

5.2. *Although Positing Primitive Processes May Save an IRF for Sub-personal Inference, It Does So Only at the Expense of Rendering IRF Untenable for Active Reasoning*

At one stage, Boghossian considers a proposal about sub-personal inference, which may appear to resemble primitivism to a considerable degree. It goes like this:

I consider [the premises] (1) and (2). I do so with the aim of figuring out what follows from these propositions, what proposition they support. A sub-personal mechanism within me “recognizes” the premises to have a certain logical form. This activates some sub-personal state that encodes the MP rule which then puts into place various automatic, sub-personal processes that issue in my believing [the conclusion] (3).
(2014, 15)

Setting aside the challenge posed by the Revenge Regress, which he maintains is “importantly correct”, Boghossian is prepared to imagine that some reasoning works like this. However, he continues,

That is not the sort of reasoning that this paper is about—rather, it is about person-level reasoning, reasoning as a mental action that a person performs, in which he is either aware, or can become aware, of why he is moving from some beliefs to others.

No such process of reasoning can be captured by a picture in which (a) reasoning is a matter of following rules with respect to the contents of our attitudes and (b) our following rules with respect to the contents of our attitudes is a matter of automatic, subconscious, sub-personal processes moving us from certain premises to certain conclusions.

If this is so, then it may seem that, by introducing such sub-personal processes, we fail to account for the sort of active reasoning we sought to understand in the first place. And since this sort of sub-personal process looks much like primitive rule application, it may further appear that endorsing primitivism thereby undermines the prospect of an IRF about active reasoning.

Appearances, at least in this instance, are misleading. As we noted in Section 3, it is widely if not universally supposed that active reasoners often, though not invariably, lack conscious awareness of the rules they are following (cf. Boghossian, 2014, 12).⁸ In which case, it ought to be common ground that there are two different kinds of active reasoning:

AR₁: Active reasoning for which there is conscious awareness of the premises and the conclusion, but not the rule.

AR₂: Active reasoning for which there is conscious awareness of the premises, the conclusion, *and* the rule.

If the representation of the rule is sub-personal, we have AR₁. If the representation of the rule is person-level, we have AR₂—the sort of reasoning that Boghossian says he seeks to explain.

In our view, a theory of reasoning ought to capture *both* these kinds of active reasoning. And, as far as we can tell, our response to the Regress in no way prevents us from doing so. If we insisted that *all* rule application was sub-personal, then we could not. But we make no such commitment. Indeed, we allow for at least four different sorts of rule-governed process.

- We allow for AR₂ because we accept that, in some cases, people are conscious of the rules they follow as well as their premise and conclusion attitudes.
- We allow for AR₁ because we posit rule-representations that are sub-personal and, hence, not the subject of conscious awareness.
- We allow for non-primitive, sub-personal inferences where one has no conscious awareness of premises, conclusion, or rule.
- Finally, we allow for primitive sub-personal rule application processes, which involve no inferential sub-processes but are rule-governed, in a thin sense, by virtue of taking rule-representations as input.

The main point of positing this hierarchy of processes was to allow for dependencies that block the Regress in a plausible fashion. Thus, for example, in some cases, AR₂ may rely on AR₁ so that a consciously accessible rule is applied via an inferential sub-process whose rule is not, itself, consciously accessible. Further, we allow that such AR₁ processes may rely on non-primitive, sub-personal inferences. And, of course, we *insist* that all such cascades must at some point rely on primitive rule application processes. Far from failing to accommodate AR₂, we maintain that we accommodate it, and many other sorts of inference beside.

⁸ Furthermore, that children can engage in active reasoning is taken to be an important reason to avoid conceptual oversophistication in their accounts (see, e.g., Broome, 2013, 229, 236; Boghossian, 2014, 6).

6. Conclusion

We started by touting the prima facie explanatory virtues of IRFs. We then argued that, by positing a cascade of different sorts of rule application processes, IRFs can accommodate active reasoning in a non-regressive fashion. Specifically, we argued that these resources allow proponents of IRFs to address the Regress, even in its strengthened form. We concluded by suggesting that our solution does not generate any obviously untenable consequences for IRFs. In view of this, and contrary to the opinion of many, we conclude that the Regress fails to undermine intentional rule-following accounts of reasoning.

References

- Block, N. (1995). The mind as the software of the brain. In E. Smith & D. Osherson (Eds.), *An invitation to cognitive science, volume 3: Thinking* (pp. 377–425). Cambridge, MA: MIT Press.
- Boghossian, P. (2008). Epistemic rules. *Journal of Philosophy*, 105, 472–500.
- Boghossian, P. (2014). What is inference? *Philosophical Studies*, 169, 1–18.
- Broome, J. (2013). *Rationality through reasoning*. Oxford: Wiley-Blackwell.
- Broome, J. (2014). Normativity in reasoning. *Pacific Philosophical Quarterly*, 95, 622–633.
- Chomsky, N. (1969a). Linguistics and philosophy. In S. Hook (Ed.), *Language and philosophy: A symposium* (pp. 51–94). New York: New York University Press.
- Chomsky, N. (1969b). Comments on Harman's reply. In S. Hook (Ed.), *Language and philosophy: A symposium* (pp. 152–159). New York: New York University Press.
- Collins, J. (2005). On the input problem for massive modularity. *Minds and Machines*, 15, 1–22.
- Dennett, D. (1978). *Brainstorms: Philosophical essays on mind and psychology*. Montgometry, VT: Bradford Books.
- Dennett, D. (1982). Styles of mental representation. *Proceedings of the Aristotelian Society*, 83, 213–226.
- Drayson, Z. (2012). The uses and abuses of the personal/subpersonal distinction. *Philosophical Perspectives*, 26, 1–18.
- Fodor, J. (1968). The appeal to tacit knowledge in psychological explanation. *The Journal of Philosophy*, 65, 627–640.
- Fodor, J. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.
- Fodor, J. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.
- Fodor, J. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: MIT Press.
- Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3–71.
- Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? Exhaustive? Empirically distinguishable? *Cognition*, 65, 197–230.
- Harman, G. (1967). Psychological aspects of the theory of syntax. *The Journal of Philosophy*, 64, 75–87.
- Harman, G. (1969). Linguistic competence and empiricism. In S. Hook (Ed.), *Language and philosophy: A symposium* (pp. 143–151). New York: New York University Press.

- Johnson-Laird, P. (2008). *How we reason*. Oxford: Oxford University Press.
- Kripke, S. (1982). *Wittgenstein on rules and private language: An elementary exposition*. Cambridge, MA: Harvard University Press.
- Laurence, S., & Margolis, E. (1997). Regress arguments against the language of thought. *Analysis*, 57, 60–66.
- Perfors, A., Tenenbaum, J., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, 118, 306–338.
- Pylyshyn, Z. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, 3, 111–132.
- Rips, L. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, MA: MIT Press.
- Ryle, G. (1949). *The concept of mind*. London: Hutchinson.
- Slovan, S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3–22.
- Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, 1–23.
- Wright, C. (2014). Comment on Paul Boghossian, ‘What is inference?’. *Philosophical Studies*, 169, 27–37.