

Intrinsic Responsibility for Rule-Following*

Ulf Hlobil

Abstract

The paper explores responses to an inconsistent quartet of theses regarding rule-following. In addressing this inconsistent quartet, two lines of thought pull in opposite directions. On the one hand, it can seem that rule-following cannot require acts that shape or guide themselves or acts that require infinitely many similar acts. On the other hand, rule-following seems to require that we are responsible for our acts of rule-following in a special way. It is difficult to see how these thoughts can be coherently combined.

1 Introduction

The goal of this paper is to investigate and clarify a potential tension between thinking of rule-following as a natural phenomenon¹ and thinking

*Thanks for very helpful comments and discussion goes to Olivia Sultanesu, Robert Brandom, Shuhei Shimamura, Ryan Simonelli, and Rea Golan.

¹In light of a reviewer's comment, I want to stress that I will use "natural phenomenon" in a stipulative way in this paper, namely as determined in its meaning by my stipulative definition of "naturalism" below. My philosophical points are independent of whether anyone in the literature means what I mean by "naturalism" or "natural phenomenon."

of rule-following as a cultural phenomenon. My arguments below cannot decide whether this tension can be resolved; but they suffice, I think, to throw new light on the issue and to bring the problem into sharper relief.

As I will explain in more detail below, what I mean by “thinking of rule-following as a natural phenomenon” is, to a first approximation, the claim that rule-following must be something that finite beings can do and that does not involve acts that are identical to what guides or shapes these acts. What I mean by “thinking of rule-following as a cultural phenomenon” is, to a first approximation, the claim that in virtue of following a rule, the agent cannot reject responsibility for behaving in accordance with the rule and for correctly applying the general rule to her particular case at hand. It will become clear below that we can view rule-following as the paradigm of acts that give rise to this kind of responsibility, which I call “intrinsic responsibility.”

My considerations below have similarities to—and have been influenced by—ideas due to Boghossian (2012) and Wright (2012; 2007). In contrast to Boghossian and Wright, however, my considerations are independent of theories of meaning and content.

The paper is structured as follows: In Section 2, I present an exhaustive classification of coherent accounts of rule-following. In Section 3, I introduce the thought that the distinction between rule-following and merely behaving in accordance with a rule implies that rule-following requires some “uptake” of the rule and an application of the rule to one’s case at hand. Sections 4–6 introduce the notion of intrinsic responsibility and use it to bring out a tension with the ideas in Sections 2–3. Section 7 considers objections, and Section 8 concludes.

2 An Issue Regarding Rule-Following

There is a family of issues regarding rule-following that involve regresses (see Reiland, 2024; Sultanescu, 2023; Boghossian, 2012; Wright, 2012, 2007).

Kant formulates one such issue as follows:

If [... logic] wanted to show generally how one subsumes under these [logical] rules, i.e., how one should decide whether something falls under them or not, then this could only happen by another rule. Precisely because this would be another rule, however, it would again require some guidance by the faculty of judgment. And it hence becomes clear that while the understanding can be instructed by, and furnished with, rules, the faculty of judgment is a special gift that cannot benefit from instruction but only from exercise. (Kant, KrV, B172, my translation)

Here Kant notices that if the application of a rule is itself a case of rule-following, then this launches us on a vicious regress that makes rule-following an impossible task. Kant's solution is to introduce the faculty of judgment as an ability to apply rules to particular cases without its acts being themselves cases of rule-following.

We can formulate Kant's regress as an inconsistent quartet of theses, thereby revealing some implicit assumptions.

Rule-Following: If subject *S* follows a rule *R*, then *S* applies *R* to her case at hand.

Ruly-App: If *S* applies *R* to her case at hand, then this application of a rule is itself a case of rule-following.

No-Auto-App: The following of some rule and the aspect (or part or cause) of it that is the application of the rule to the case

at hand (and any of its aspects (or parts or causes)),² are two distinct cases of rule-following.

Finitude: There are cases of rule-following that do not require that the subject engages in infinitely many distinct cases of rule-following.

To see that these four theses are inconsistent, consider the following reasoning.

1. By Finitude, we can assume that S follows a rule, R_1 , and S does not engage in infinitely many cases of rule-following.
2. By Rule-Following, S applies R_1 to the case at hand.
3. By Ruly-App, S 's application of R_1 to the case at hand, is itself a case of following some rule R_2 .³
4. By Rule-Following, S applies R_2 to the case at hand.
5. This goes on to infinity, unless, some of the acts of rule-following or rule application that occur in distinct places in this chain of rule applications are identical acts; but that is ruled out by No-Auto-App.
6. Therefore, S must engage in infinitely many cases of rule-following, which contradicts our assumption.

Since the four theses are inconsistent, any account of rule-following must reject at least one of them. This gives us the following exhaustive (but

²The parenthetical addition "and any of its aspects" is there to ensure that No-Auto-App entails not only that the following of a rule and its application are distinct but also entails that the following of a rule and any act of rule application that is an aspect of the application of the rule are also distinct. Thus, No-Auto-App concerns the transitive closure of the relation of *being an aspect of an act of rule-following*, and it thereby covers the whole chain of rule applications that is mentioned in line 5 of the argument below.

³Nothing in this argument requires R_1 and R_2 to be distinct rules.

not mutually exclusive) categorization of all coherent accounts of rule-following, according to which of the four theses above they reject.

No-App: Rule-following does not require that one applies the rule that one follows to one's case at hand.

Unruly-App: Applying a rule to a case at hand is not always itself a case of rule-following.

Auto-App: The following of some rule and the aspect (or part or cause, or an element in the transitive closure of this aspect-of relation) of it that is the application of the rule to the case at hand are sometimes identical.

Infinity: Whenever anyone follows any rule, they are engaging in infinitely many distinct cases of rule-following.

As is clear from the quote above, Kant adopts an account of rule-following that belongs to the Unruly-App category. He holds that applying ("subsuming under") a rule is an act of the faculty of judgment, whose acts are not cases of rule-following. However, I am not interested in Kant exegesis. I am interested in how we should think about rule-following in light of the inconsistent quartet above.

One way to structure the options can be formulated in terms of thinking about rule-following as a natural and as a cultural phenomenon.⁴ What could we mean by "understanding rule-following as a natural phenomenon" or as a "naturalistically intelligible phenomenon"? It is a fraught issue what naturalism is and whether it is a fruitful philosophical concept (see Stroud, 1996; Spiegel, 2022; Raleigh, 2024). It seems to me that

⁴Nothing here hangs on whether anyone agrees with my use of terms like "natural phenomenon" and "naturalism." I will clarify in the next paragraph how I use these terms.

the—I think, unclear—use of “naturalism” in philosophy today suggests that accounts of the Auto-App and Infinity type are not compatible with “naturalism” in this—unclear—sense. However, I do not want to consider whether this claim is true, as it is irrelevant to my project in this paper. I will rather define what I mean by “naturalism” for the purposes of this paper as the joint denial of Auto-App and Infinity.⁵ If there are people who call themselves “naturalists” and who endorse Auto-App or Infinity, I am interested in their accounts of rule-following; but they won’t count as naturalists for my purposes in this paper. Similarly, what I mean by “natural phenomenon” is something that can be satisfactorily explained by an account that is compatible with naturalism, in the just defined sense. Readers who are reluctant to use “naturalism” and similar terms in accordance with my definitions are invited to replace such terms with a contextually appropriate phrase about the joint denial of Auto-App and Infinity.

Since I want to investigate whether and how we can think of rule-following simultaneously as a natural (in my stipulative sense) and as a cultural phenomenon, I will henceforth focus exclusively on accounts of rule-following of the No-App and Unruly-App variety. In the next section, I will argue against No-App type accounts.

3 Applying Rules

Any adequate account of rule-following must distinguish rule-following from merely behaving, and being disposed to behave, in accordance with a rule. In this section, I introduce a suggestion for how to draw the necessary distinction.

⁵Below I will extend this definition from rule-following to what I call “responsibility-source acts.” This must wait, however, until I have assembled the necessary pieces.

Behaving in accordance with a rule—and being disposed to do so—is neither necessary nor sufficient for rule-following. It is not necessary because one can follow a rule and nevertheless fail to behave in accordance with the rule if one follows the rule incorrectly, as when one applies the rule incorrectly to a case at hand, or when one makes a mistake in one's performance. Moreover, acting in accordance with a rule is not sufficient for rule-following: one can behave in accordance with a rule—and be disposed to do so—for reasons that have nothing to do with the rule. In order to follow a rule, rather than just to behave in accordance with it, what one does must be shaped or guided by the rule in some to be determined, appropriate way.

The distinction between rule-following and behaving in accordance with a rule is not the distinction between acts that fall under the rule, or can be appropriately assessed by the rule, and acts that are in accordance with the rule but do not fall under it or cannot be appropriately assessed by it.⁶ One's behavior in social situations, for instance, falls under the rule that one ought to be polite and it is appropriate to assess such behavior by this rule of politeness, but this does not imply that people who behave in accordance with the rule of politeness necessarily follow the rule of politeness in such situations. It might be a lucky coincidence that they are polite. Conversely, one can follow the rule not to step on boarders between tiles in the floor while one does not make a mistake and cannot

⁶Here I disagree with Brandom (1994, chap 1), who holds that the philosophical problems regarding rule-following can be solved by understanding how a norm can be implicit in a social practice. Brandom often starts with examples like bidding in an auction or taking the Queen's Shilling, which are acts by which one incurs obligations independently of whether one knows or accepts the norms that one makes applicable to oneself. Notice that these are not cases of rule-following. When I bid in an auction without knowing that I do so, or when I take the Queen's Shilling without knowing that I do so, then I do not follow rules of bidding or rules of enlisting in the Royal Navy. I doubt that Brandom can account for the phenomenon of rule-following on the basis of these other ways of making norms applicable.

be negatively assessed for stepping on a boarder in the tiling, except in the sense that doing so violates the rule one follows. Hence, acts can fall under rules and it can be appropriate to assess acts by rules without them being acts of following the rule, and one can follow rules without falling under them in any independent sense. Acts of rule-following must be shaped or guided by the rule, and this is not true of all acts that fall under rules or can be appropriately assessed by the rule.

What does it mean to be shaped or guided by a rule? A natural suggestion is that some appreciation of the rule must shape or guide the act. Brandom (1994, 30) attributes a version of this idea to Kant in the following formulation: “What is distinctive about us as normative creatures is the *way* in which we are subject to norms (for Kant, in the form of rules). As natural beings, we act according to rules. As rational beings, we act according to our *conceptions* of rules.”⁷ Tomasello (2016, 82) suggests that our ancestors developed normative practices by developing “a kind of second-personal responsibility to their collaborative partners—the original ‘ought’—that was not just a blind emotion or preference, but rather a sense of cooperative rational pressure that innervated their decision making.” What is common to these two suggestions is that following norms requires that “the agent’s take on the norm” appropriately shapes or informs (or “innervates”) what the agent does. It is a controversial issue, however, what “the agent’s take on the norm” could mean here.

What I will call “reductionism” about rule-following holds that what it is for an act to be shaped or guided by a rule is for the agent to have or to manifest the right sort of dispositions and perhaps other states, all of which can be explained in non-intensional terms. Anti-reductionists

⁷Kant himself says that an imperative, which he takes to be the central practical norm, is “a rule whose representation makes a subjectively-contingent action necessary, thus representing the subject as to be forced (necessitated) to be in accordance with the rule” (Kant, AA VI, 222, my translation).

hold that in order to follow a rule one must have an appreciation of the rule and what one does must be the result of applying this appreciation to one's current case; and this appreciation and application cannot be explained in non-intensional terms. The disagreement between reductionism and anti-reductionism may seem to be the crux of the problem of rule-following.⁸ However, I think that the importance of the issue is sometimes overestimated, and I hope to stay neutral regarding the disagreement.

In order to be neutral on the issue of reductionism, I will use the term "uptake" as a neutral technical term for an event and the resulting state in which an agent is in broadly cognitive rapport with some content or fact, such as when a rule or fact is encoded or stored or represented or learned; where it is left open whether this "uptake" can be described in non-intensional terms. Developing dispositions (that can be individuated in non-intensional terms) to behave in accordance with a rule may count as uptake of the rule in this sense; nothing below turns on this issue.

Irrespective of whether we are reductionists or anti-reductionists about rule-following, we need some way to explain the difference between rule-following and being in accordance with a rule. The idea I want to investigate is that this requires, at a minimum, that there is some uptake of the rule and some uptake of the case at hand and these uptakes are combined in an appropriate way, which shapes or guides a doing that is a case of rule-following. By way of fixing terminology, I will use "applying a rule to the case at hand" for the event or act or state that shapes or guides the doing that is the rule-following. We can summarize the idea as follows:

⁸The disagreement between reductionism and anti-reductionism is crucial for the interaction between considerations of rule-following and the theory of content and meaning. However, the theory of content and meaning is not my concern in this paper.

Rule-App: Someone is following a rule in doing ϕ only if she is applying the rule to the case at hand, i.e., there is some uptake of the rule and some uptake of the case at hand and these uptakes combine in shaping or guiding the agent's doing of ϕ .

Notice that Rule-App is compatible with reductionism about rule-following because it leaves open the possibility that the uptakes of the rule and the case at hand can be individuated in non-intensional terms. Notice also that Rule-App makes no assumption about the uptake's format, availability to consciousness, explicitness, about whether it is a representation, or the like. All that Rule-App says is that if a doing is not merely a case of someone or something behaving in accordance with a rule but rather a genuine case of rule-following, then the general rule and the particular case at hand must be brought together in (or by) the someone or something that is following the rule and this "bringing together" must shape or guide the doing that is the rule-following. Indeed, Rule-App says very little and is a minimal condition on rule-following.

Although Rule-App is a minimal condition on rule-following, it rules out No-App above. It rules out No-App because it says that rule-following requires that one applies the rule one follows to the case at hand, and that is precisely what No-App denies.

If what I said in this section is correct, there is some reason to accept Rule-App. For, it promises to enable us to draw a distinction that we need to draw in order to make sense of rule-following. Moreover, naturalism about rule-following and Rule-App jointly imply that Unruly-App is the only viable type of account of rule-following. So, with the above argument and Kant's authority impressing themselves upon us, we must examine Unruly-App in more detail. In order to put some pressure on Unruly-App in later sections, I will next introduce the notion of intrinsic responsibility.

4 Intrinsic Responsibility

In this section, I introduce the notion of intrinsic responsibility and the related notion of responsibility-source acts. I will later use these notions to raise problems for Unruly-App. I will first explain what I mean by “intrinsic responsibility” and “responsibility-source acts.” Then I will explain why rule-following, as the culturally important phenomenon that we want to understand, must be a responsibility-source act. This yields a constraint for thinking of rule-following as a cultural phenomenon.

4.1 What is Intrinsic Responsibility?

In this subsection, I will first give a stipulative definition of intrinsic responsibility. Next I will illustrate the notion by giving some examples. Then I will introduce the related notion of a responsibility-source act.

I will say that subject S is intrinsically responsible for her doing ϕ being in accordance with a norm N just in case S does ϕ and it is impossible, in virtue of what it is to do ϕ , that S does ϕ and S also sincerely rejects to acknowledge being responsible for her ϕ -ing being in accordance with N . So, for any act-type ϕ such that the agent is intrinsically responsible for acts of that type, if S sincerely rejects to acknowledge responsibility for her apparent ϕ -ing being in accordance with the relevant norm N , this implies that S is not doing ϕ . What I mean by “rejecting to acknowledge responsibility for one’s ϕ -ing being in accordance with a norm N ” is what one does when one says that one doesn’t accept the norm N as governing one’s ϕ -ing and, hence, makes no mistake (by one’s own light) when one acts against it.

Asserting and believing are examples of acts that give rise to intrinsic responsibility.⁹ Suppose I (seem to) assert that p or (seem to) believe that p , and you point out that my assertion or belief violates the truth-norm on assertion and belief because p is false.¹⁰ If I now sincerely reply that I acknowledge no responsibility for my (apparent) assertion or (apparent) belief being true, then this implies that what I did was not really asserting or believing.¹¹ In such cases, I might not take myself to be (and need not be) blameworthy or an appropriate subject of punishment or the like. It is, however, impossible that what I did was asserting or believing and that I acknowledge that p is false but also reject the criticism that I made a mistake. So, not only do my assertions and beliefs fall under the norm that they should be true and can be appropriately assessed by this norm, but I must also admit such assessments as proper or applicable, to what I do, in order for my doings to be assertions and beliefs.

Notice that my account does not imply that we cannot intentionally lie. I can make assertions that I know to be false, and I may not care about my assertion being false. But if someone calls me out on my (apparent) lie and I respond to the criticism by sincerely saying that I don't take myself

⁹I think that there are many more examples for responsibility-source acts. For instance, I think that promising is a responsibility-source act. However, this presupposes that one rejects Scanlon's (1990) account of promising. The claim becomes plausible if one accept, as I do, Nieswandt's (2019) criticism of Scanlon.

¹⁰I here presuppose that assertion and belief are governed by a truth-norm. Nothing below turns on this. What is important for me is that asserting and believing are such that it is essential to these acts that they fall under some norm. If one takes this norm to appeal to knowledge or evidence and not to truth, then one should reformulate my point in the main text accordingly.

¹¹An anonymous reviewer worries that this claim is incompatible with the fact that we can assert and believe things like " p , but there is evidence against p " or " p , but I cannot vouch for its truth." I think that these possibilities turn on the fact that the truth of p is compatible with there being evidence against p and with my not being able to vouch for the truth of p . This comes out in the fact that asserting or believing things like " p , but p is not true" seems unintelligible, unless some very special explanation or context is provided.

to have done anything that falls under the norm that it ought to be true and, hence, I did not act against any norm that I acknowledge, then this implies that I did not make any assertion and, hence, did not in fact lie.¹²

I will use “responsibility-source acts for norm *N*” as a label for acts that give rise to intrinsic responsibility to be in accordance with some norm *N*. These are acts by which we bring ourselves under norms, in contrast to merely falling under these norms. In virtue of what we ourselves do, we make the norm applicable to us, by our own lights.

Let me explain. We can apply the aesthetic norm of beauty to sunsets, and sunsets may fall under this norm. But sunsets do not bring themselves under the norm of beauty. Similarly, moral norms apply to amoralists and skeptics regarding morality.¹³ Perhaps there is a broad sense in which these agents bring themselves under moral norms merely by acting intentionally or voluntarily; but they do not bring themselves under these norms in my narrower sense. They are, for instance, not intrinsically responsible for acting morally in taking my food because it is possible that they take my food while also rejecting to acknowledge any responsibility to act morally in taking my food, i.e., they may claim that they did not violate any genuine norm by taking my food.

By contrast, rule-following is a responsibility-source act. In following a rule, one brings oneself under the rule; one makes the rule appropriate for assessing what one does, not merely from the perspective of someone who accepts and wants to enforce the rule but from one’s own perspective,

¹²The situation is different for belief, where I hold that one cannot believe what one takes to be false, barring repression or confusion or dialethist solutions to semantic paradoxes or the like. This claim, however, is neither a premise nor an implication of my arguments in this paper.

¹³I am here presupposing that a subject’s beliefs and desires cannot undermine the applicability of moral norms. I think that a plausible, moderately externalist position regarding moral reasons suffices to ensure this; but the issue is outside the scope of this paper.

in virtue of one's doing being a following of the rule. To see this, suppose that someone points out that what you are doing is not in accordance with the rule that you are following. Can you sincerely reply by saying that you don't accept this standard of assessment? If you sincerely reply in this way, then this shows, I think, that you are not actually following the rule.¹⁴ If this is correct, we cannot reject to acknowledge the responsibility that we have for our rule-following being in accordance with the rule that we follow. Thus, we are intrinsically responsible for our rule-following.

Three clarifications are in order. First, what I said does not imply that responsibility-source acts require that the agent represents herself as responsible for being in accordance with the rule. It merely means that a rejection of such an acknowledgment is incompatible with performing the responsibility-source act.¹⁵ Second, what I said does not imply that the responsibility at issue has any importance or authority. We can follow silly or trivial rules. I might, for instance, follow the rule not to step on the borders in the tiling, and do so for no particular reason. When I realize that I stepped on the border between two tiles, I must acknowledge that I made a mistake—a trivial and utterly unimportant mistake, of course.

¹⁴In the sense at issue here, we are not following the grammatical rules of our first language (nor those of universal grammar). For these grammatical rules are usually such that we can say that we reject them (e.g., because they don't seem to us to be the correct rules) while we still count as uttering the sentences that we construct, unbeknownst to ourselves, according to the grammatical rules that we reject. We are merely non-accidentally behaving in accordance with these grammatical rules. The rules of grammar are, in this respect, like the rules that govern how we should adjust the lenses in our eyes, so as to see properly.

¹⁵Notice that this is not a merely negative condition. The performance of responsibility-source acts must have a positive property, namely the property of being incompatible with the mentioned rejection. One may claim that the performance can have this property only if it includes a representation of the norm or some similar representation. If such an argument is successful, then my account turns out not to be independent of the question whether rule-following requires the representation of the rule. However, as far as I can see, I can afford to leave this issue open for the purposes of this paper. (Thanks to an anonymous reviewer for pressing me on this point.)

Third, intrinsic responsibility applies only to the first-person perspective in the present tense. I can follow a rule and later decide that I don't accept the rule any more. At the time of rule-following, however, I cannot reject my responsibility for being in accordance with the rule. Similarly, I can mistake my mirror image for someone else, take the person to be following a rule, and reject any responsibility for the correctness of what that person does.

4.2 Why does Intrinsic Responsibility Matter?

Why is it important that rule-following is a responsibility-source act, that is, that rule-following gives rise to intrinsic responsibility? It is important because (a) rule-following is often taken to be a normative phenomenon that lies at the bottom of all other normative phenomena, and (b) the existence of responsibility-source acts is a necessary condition for the rich normative phenomena that are characteristic of human culture. These two ideas jointly imply that we must be able to explain rule-following as a responsibility-source act without appealing to other responsibility-source acts. In this sense, our account of rule-following must explain rule-following as a basic responsibility-source act.

Let me start with the idea (a) that rule-following lies at the bottom of all normative phenomena. We can find a version of this idea in Wright's work:

It is natural to think that in any area of human activity where there is a difference between *correct* and *incorrect* practice, which we achieve is (partly) determined by rules which fix what correct practice consists in, and which in some manner guide our aim. (Wright, 2007, 481)

Peregrin endorses a similar view in the opening of a recent book entitled *Normative Species*:

This book is about rules, and especially about the human capability to create, maintain and follow rules as a root of what makes us humans different from other animals [namely, what makes us a “normative species”]. Indeed, it is meant as an elaboration of Wilfrid Sellars’ visionary observation that “to say that man is a rational animal, is to say that man is a creature not of *habits*, but of *rules*”. I am convinced that scrutinizing this capability will let us understand who we humans are and what kinds of lives we lead. (Peregrin, 2023, 1)

Here Peregrin suggests that being rational, the kind of cultural life that humans lead, and belonging to a “normative species” are all intimately linked to rule-following. Moreover, Peregrin takes his topic to be continuous with the topic of recent empirical research on social norms, which (or a part of which) is sometimes called “norms psychology” (Heyes, 2024; Tomasello, 2016; Schmidt et al., 2011; Sripada et al., 2005). Heyes (2024, 13-14) explicitly characterizes norm psychology as considering norms in a wide sense (as opposed to the narrow focus on moral norms that she attributes to moral psychology), and she outlines a “cultural evolutionary account of norm psychology.”

According to these passages and authors, what is at issue in considerations of rule-following is a necessary condition for the normativity that underlies our rich cultural lives as humans. If this is our topic, then it will not do to explain rule-following in terms of other normative phenomena that have the same structural properties as rule-following. In particular, given that rule-following is a responsibility-source act, if responsibility-source acts are necessary for explaining the normativity that underlies our

rich cultural lives as humans, then we should not explain rule-following in terms of other (unexplained) responsibility-source acts. For that would merely move our questions to these other responsibility-source acts, which we would need to understand to explain the normativity that underlies our rich cultural lives.

Let's now turn to idea (b) above. Why are responsibility source acts a necessary part of the normativity characteristic of human culture that is our ultimate explanatory target in thinking about rule-following as a cultural phenomenon (at least according to the passages just quoted)? Intrinsic responsibility is a necessary condition for normative phenomena of the kind that is our ultimate explanatory target because it is the way in which norms first get a grip on us from our own perspective.

By performing responsibility-source acts we bring ourselves under norms. When we do this, others can criticize us in light of norms that necessarily apply to us by our own lights, in virtue of our act being a responsibility-source act. Others can hence criticize us in such a way that we must acknowledge that we are responsible for failures to live up to the norms that are used to criticize us.

If a normative practice does not include responsibility-source acts, then it is always open to the practitioners to reply to criticism by saying that they do not accept the norm that is used to criticize them, while still counting as performing the criticized act. From their own perspective, any negative reactions from other practitioners can appear as mere incentives or disincentives to change their behavior, in broadly the same way in which (nonnormative) nature can provide such incentives or disincentives. As a result, a normative conflict would appear to the practitioners as not

giving rise to any distinctively normative pressure but merely to welcome and unwelcome actions by others.¹⁶

Practices that include responsibility-source acts are different. For, intrinsic responsibility ensures a kind of normative friction. The normative criticism of responsibility-source acts creates a distinctively normative pressure from the perspective of the criticized agent. The agent is under an obligation, by her own lights, to react to such criticism in a way that does not reject the norm. This enables us to engage with each other against the background of accepted norms that cannot be rejected (at least not without thereby re-characterizing the acts performed by the opposing parties). Without intrinsic responsibility, we would have no way to ensure that our normative assessments of each other don't simply slide off one another without any normative friction. Without responsibility-source

¹⁶Here I disagree with Brandom (1994), who holds that we can explain the normativity of rule-following on the basis of the normativity that is implicit in a practice of sanctioning people in response to their behavior, such as the practice of treating people who raise their arm at an auction as bidding, independently of whether they know that their action will have this effect. As Brandom puts it, his idea is to construe "the normative attitude of taking or treating something as correct or incorrect in practice in terms of the application of positive or negative sanctions" (Brandom, 1994, 45). And he holds that we can understand all normativity, including the normativity at issue in rule-following, in terms of the normative attitude of taking or treating something as correct or incorrect in practice. Brandom acknowledges that the application of sanctions can always itself be assessed as correct or incorrect. However, I doubt that anything can count as the application of sanctions unless the person who applies the sanctions follows a rule in applying the sanctions and, hence, takes herself to be applying the sanctions correctly. To see this, suppose we have a society where people have dispositions to beat each other with sticks in certain situations but, for all these people know, their behavior of beating each other is random and is not the application of any rule. This does not seem to me to be a case in which these people apply negative sanctions to each others behavior. They merely accidentally train each other to behave in certain ways. If this is correct, Brandom cannot use the concept of applying sanctions to explain the normativity at work in rule-following. For related discussions of Brandom see (Pippin, 2005; Satne, 2017).

acts, normative assessments would degenerate into little more than name calling and threats of punishment.¹⁷

If what I just said is correct, then responsibility-source acts and intrinsic responsibility are a necessary part of any normative practice that can count as culture or as the kind of rich normative phenomenon characteristic of human life. It follows that we need an account of rule-following that does not presuppose an understanding of other responsibility-source acts.

The upshot of this section is that it is a constraint on thinking of rule-following as an important cultural phenomenon (in the sense intended by at least some researchers in this area) that we can account for rule-following as a responsibility-source act without appealing to other responsibility-source acts. In the remainder of this paper, I want to show that, given plausible further premises, this constraint on accounts of rule-following is in tension with the idea behind Rule-App from the previous section.

5 Acts Generating Intrinsic Responsibility

If what I said above is correct, our real topic when considering rule-following are responsibility-source acts for norms in general, not something that is specific to rules or the following of rules in contrast to norms and bringing oneself under norms more generally. In this section, I will first reformulate some of the above ideas about rule-following explicitly in terms of responsibility-source acts. In particular, I will formulate an analogue of Rule-App in terms of responsibility-source acts. Then I will formulate a thesis, which I call Source-Resp, that is inconsistent with the

¹⁷Some philosophers may hope to account for genuine normative friction in terms of punishment and assessments that need not be shared by the agent. I am pessimistic regarding the prospects of such a project, and I here assume that such a project cannot be carried out successfully. My conclusions in this paper are (henceforth implicitly) conditional on that assumption.

analogue of Rule-App above; and I will argue for Source-Resp in the next section.

Let's start by translating the four theses that span the field of coherent accounts of rule-following into the terminology of responsibility-source acts: The analogue of No-App for responsibility-source acts is the thesis that responsibility-source acts do not require that the norm is applied to the case at hand. The analogue of Unruly-App is the thesis that applying a norm to a case at hand is not always itself a responsibility-source act. The analogue of Auto-App is the thesis that a responsibility-source act and the application of the norm to the case at hand (or some rule application in the transitive closure of the aspect-of relation for the act) are sometimes identical. And the analogue of Infinity is the thesis that every responsibility-source act requires engaging in infinitely many distinct responsibility-source acts. These are the four options to avoid an inconsistent quartet of theses about responsibility-source acts that corresponds to the inconsistent quartet that I distilled out of the Kant quote at the start of this paper.

We can adjust the definition of naturalism accordingly and say that naturalism is the combined rejection of the idea that there can be responsibility-source acts that are identical to their norm applications and the rejection of the idea that there can be responsibility-source acts that involve infinitely many distinct responsibility-source acts or loops of such acts. In parallel to what I said above, this leaves open the analogues of No-App and Unruly-App. And, in parallel to what we did above, we can now rule out the analogue of No-App by accepting an analogue of Rule-App, namely the following thesis:

Source-App: Someone is performing a responsibility-source act for norm *N* only if she applies norm *N* to her case at hand (i.e., there is some uptake of *N* and some uptake of the particular case

at hand and these uptakes combine in shaping or guiding the responsibility-source act for norm N).

The motivation for Rule-App carries over to Source-App. We must distinguish between behaving (and being disposed to behave) in accordance with a norm and performing a responsibility-source act for the norm. Moreover, we must distinguish falling under a norm and bringing oneself under the norm by performing a responsibility-source act. As we saw in the previous section, this distinction is crucial because acts that give rise to intrinsic responsibility are a necessary condition for the normative cultural phenomena that we ultimately want to understand. Merely falling under a norm does not ensure the kind of normative friction that we need from responsibility-source acts. Source-App is a suggestion for making room for these distinctions by saying that bringing oneself under a norm requires that one applies the norm to one's case at hand. By contrast, neither a disposition to behave in accordance with a norm nor merely falling under the norm require such an application of the norm by the agent.

Let's consider where we are: That we reject the analogues of Auto-App and Infinity constitutes the acceptance of naturalism (in my stipulative sense). And that we moved from rule-following to responsibility-source acts and accept Source-App reflects our desire to think of rule-following as the basis of the broader cultural phenomenon of normativity. For it indicates that we are interested in norms in general, and that we are interested in a class of responsibility-source acts that can be understood without already presupposing an understanding of some responsibility-source acts.

Unfortunately, the conjunction of Source-App and naturalism is inconsistent with the following thesis:

Source-Resp: The application of a norm to a case at hand is itself a responsibility-source act.

To see this, suppose that S engages in a responsibility-source act for norm N_1 . Thus, by Source-App, S applies N_1 to her case at hand. By Source-Resp, this application of N_1 is itself a responsibility-source act for some norm N_2 .¹⁸ It follows by Source-App, that S applies N_2 to her case at hand. This either (i) goes on to infinity, or (ii) some responsibility-source act for a norm N_k and the application of N_k to the case at hand are identical, or (iii) there are loops of responsibility-source acts and applications. But all three options (i)–(iii) are ruled out by naturalism.

We thus reach the result that it is incoherent that naturalism is true and that there are responsibility-source acts that obey Source-App and Source-Resp. We can summarize this as follows:

Choice-Point: Given naturalism, responsibility-source acts either do not require that the norm under which they bring the agent is applied to the case at hand (thus rendering Source-App false (call these “brute responsibility-source acts”)) or the application of this norm is not itself a responsibility-source act (thus rendering Source-Resp false (call these “irresponsible norm applications”)).

In other words, if naturalism is true and we must make room for acts that give rise to intrinsic responsibility, then there must be either brute responsibility-source acts or irresponsible norm applications. That is, either we can become intrinsically responsible for being in accordance with a norm without the norm being applied to our case at hand,¹⁹ or we can become intrinsically responsible for being in accordance with a

¹⁸This argument does not require that N_1 and N_2 are distinct.

¹⁹This option is structurally similar to the position that rule-following must, in the basic cases, be blind (Boghossian, 2012; Wright, 2007), where rule-following is blind if it is not informed by some uptake of the requirements of the rule that one follows.

norm by an act of norm application for which we are not intrinsically responsible.

Translating this back into the language of rule-following: if naturalism is true and rule-following gives rise to intrinsic responsibility, then either there are cases of rule-following in which the rule is not applied to the case at hand (“brute rule-following”) or there are cases of rule-following in which we are not intrinsically responsible for the application of the rule to the case at hand (“irresponsible rule applications”).

If we accept Source-App on the basis of the considerations above, then we must reject that there are brute responsibility-source acts. This then forces us to say that there are irresponsible norm applications. We must thus say that there are acts of bringing oneself under a norm by applying the norm to one’s case at hand where one is not intrinsically responsible for this application of the norm. In the next section, I want to consider whether this is a plausible position.

6 Irresponsible Norm Applications

The result we have reached at this point is the following: if naturalism is true and responsibility-source acts require that the relevant norm is applied to the case at hand, then there must be responsibility-source acts in which the application of the norm is not itself a responsibility-source act. That is, Source-Resp must be false; there must be irresponsible norm applications. In this section, I will raise problems for rejecting Source-Resp.

The idea behind Source-App—and analogously for Rule-App—above was that one must apply the norm to one’s case at hand in order to bring oneself under a norm in a responsibility-source act. If we reject Source-Resp, while accepting Source-App and naturalism, we must hold that in following a norm an agent is bringing herself under a norm by applying

the norm to her case at hand, but she is not thereby bringing herself under the norm to apply the first norm correctly. The agent cannot reject responsibility for acting in accordance with a norm, but she can reject responsibility for applying this norm correctly to her case at hand.

This view strikes me as unstable. Rejecting all norms for applying a particular norm correctly to a particular case at hand, in effect, amounts to rejecting that the norm governs the case at hand. It is, for instance, incoherent to accept the truth-norm for one's current assertion of "It is raining" but to reject all norms for applying the truth-norm to this assertion. If one holds that there is no correct way to apply the truth-norm to a particular case, one thereby in effect rejects that the truth-norm governs this case. But if it is impossible to reject all norms to apply another norm correctly, with respect to some particular cases, while accepting the latter norm as governing those cases, then we cannot reject Source-Resp while endorsing Source-App with respect to any collection of cases. So, if Source-App is true for all responsibility-source acts, the agent cannot reject the respective norms that she follows and applies in each of these acts. It follows that, for any particular responsibility-source act, the agent cannot reject all norms for how to apply the norm that she follows correctly.

Formulating this idea in normative terms: it is unclear why applying a rule should make the agent intrinsically responsible for her act of norm-following if she isn't intrinsically responsible for applying the rule correctly. It is at best of dubious coherence to suggest that someone becomes intrinsically responsible for an act by applying a norm in a way that, for all she knows, is not subject to any norms she accepts.

Someone might suggest that all that a rejection of Source-Resp requires is that we have some default entitlement or *a priori* non-evidential entitlement to take our application of the relevant norm to be correct (see Wright, 2004b,a). Note, however, that what is at issue is not our justification or

entitlement to take our norm applications to be correct. Rather, what is at issue is whether we can apply a norm and also reject all norms for how to apply the former norm correctly. Even if we are by default entitled to assume that we apply norms correctly, it still seems that we should be open to the possibility and criticism that we misapplied a norm, which suggests that our application of the norm is governed by a norm that determines when the first norm is applied correctly and when it is misapplied. Hence, I don't think that an appeal to default entitlements can help at this point.²⁰

To sum up, naturalism implies that there are either brute responsibility-source acts or irresponsible norm applications. If there are brute responsibility-source acts, then an agent can become intrinsically responsible for being in accordance with a norm without applying the norm to her own case. It is unclear, however, how such cases differ from cases in which the acts can merely be appropriately assessed by the norm. In order to think of rule-following as a cultural phenomenon, there must be cases of following norms that go beyond norms being appropriately applicable. If, however, there are irresponsible norm applications, then agents become intrinsically responsible to be in accordance with a norm in virtue of applying the norm, while this application of the norm can seem to the agent not to be subject to any norm beyond those imposed on her by others. On neither of these two options, it seems, have we succeeded in thinking about the following of norms as the kind of cultural phenomenon that was our target. If this is correct, there is a deep tension between thinking of rule-following as a natural phenomenon (in my stipulative sense) and thinking of rule-following as a cultural phenomenon.

²⁰Thanks to an anonymous reviewer for pressing me on this point.

7 Objections

The Choice-Point from Section 5 above means that we must reject at least one of naturalism (in my stipulative sense), Source-App, and Source-Resp. Supposing that my argument for Choice-Point is valid, only three coherent positions are available: (a) One might doubt that thinking about rule-following as a cultural phenomenon requires that one accepts Source-App. (b) One might doubt that such a cultural perspective requires Source-Resp. (c) One might suggest that we accept and Auto-App or Infinity type account of rule-following. Since I simply assumed the rejection of Auto-App and Infinity (under the label “naturalism”), for the purposes of exploring this portion of logical space, an opponent will have to take position (a) or (b).

Here I want to consider a combination of (a) and (b). In the previous section, I have presented my considerations for thinking that a cultural perspective on rule-following requires Source-Resp. An opponent might point out, however, that if we can reject Source-App, then Source-Resp will fall with it. After all, if no norm application is required for norm-following, then no norm application that is itself a responsibility-source act is required either. Now, the opponent might hold that Source-App becomes implausible once we think of normativity as an essentially social phenomenon. Many readers of Wittgenstein tend to understand his following remarks in this direction:

It is not possible that there should have been only one occasion on which only one person followed a rule. It is not possible that there should have been only one occasion on which a report was made, an order given or understood, and so on. — To follow a rule, to make a report, to give an order, to play a game

of chess, are customs (usages, institutions). (Wittgenstein, 1953, §199)

Perhaps Wittgenstein suggests here that rule-following does not require that the rule is applied to a particular case at hand but that it is rather sufficient for rule-following that the act takes place in the appropriate social context. If so, Wittgenstein recommends a No-App type account, while rejecting Source-App and Source-Resp.

It is unclear to me, however, how one could square such a view with the realization that the kind of normativity that is central to a human life must allow for intrinsic responsibility. Perhaps the right social context can make it the case that nothing counts as a certain responsibility-source act if the agent rejects responsibility for the act. Even if this is so, however, we are facing the question what must be true of the agent when she engages in such acts. And Source-App is an answer to that question. If we don't want to bring back Source-App and remain in the No-App camp, we must provide an alternative account of how the agent brings herself under norms in responsibility-source acts. Unless and until a promising account of this sort is on the table, there is at least some pressure to accept Source-App. Once we accept Source-App, however, the issue of whether to accept Source-Resp arises again as in Section 6 above.²¹ Hence, I cannot see how this objection could block my arguments above (short of suggesting a novel account of how we bring ourselves under norms).

Up to this point, I disregarded position (c) by assuming what I labeled "naturalism." I want to end by briefly considering this position. Infinity

²¹As far as I can see, the same considerations apply to Sellars's (1969) proposal that acts of rule-following are shaped or guided by the rule, in the primitive case, in virtue of the rule being causally relevant in the training that rule-following agents have received. It is unclear how the causes of the training someone has received should make it impossible for them to reject responsibility for their acts. Perhaps there are some resources made available by the internalism present in *Empiricism and the Philosophy of Mind* (Sellars, 1997). But a discussion of this point is beyond the scope of this paper.

type accounts of rule-following strike me as implausible, and I will not discuss them here. However, I am sympathetic to Auto-App accounts, and there is a way of interpreting Wittgenstein that suggests that he advocated an account of broadly this kind. In a famous passage, Wittgenstein says that there must be a way of grasping a rule that manifests itself in particular cases of application.

It can be seen that there is a misunderstanding here from the mere fact that in the course of our argument we give one interpretation after another; as if each one contented us at least for a moment, until we thought of yet another standing behind it. What this shews is that there is a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call “obeying the rule” and “going against it” in actual cases. (Wittgenstein, 1953, §201)

Perhaps this passage suggests that there is a way of following a rule that is identical with an act of applying the rule to the case at hand. If rule-following is like that, doing what one does in following the rule and taking the rule to call for this action in the case at hand are one and the same act. For the case of belief, the idea might be that believing something and taking the belief to be true (and perhaps also taking it to be warranted) are the same act.²² And for the case of inference, the idea might be that making an inference and taking one’s premises to support one’s conclusion are the same act. In all these cases, the idea would be that a responsibility-source act is identical to a positive evaluation of itself that shapes or guides it.

²²For the case of beliefs and how we bring ourselves under the norms essential to believing, I take Boyle (2011) and Hieronymi (2009) to offer accounts that are at least in the vicinity of Auto-App type accounts.

Whether or not Wittgenstein recommends an account of the Auto-App variety, it is clear that such accounts would allow us to combine Source-App and Resp-App. If we want to accept naturalism in the vague sense in which the term “naturalism” is used in philosophy today, then accepting Auto-App requires that we explain how a process or acts that shapes or guides itself is naturalistically intelligible (in the vague sense of current philosophical discourse). For this, it will not suffice to think of such self-shaping or self-guiding processes or acts as including feedback loops (like thermostats) or the like, as a current current state shapes a future state and not itself in such feedback loops. Considering these issues is, however, beyond the scope of this paper.

8 Conclusion

If what I said in this paper is correct, there is a tension between naturalism (in my stipulative sense) and thinking of rule-following—and the following of norms more generally—as a cultural phenomenon. Let me rehearse my overall line of thought.

Naturalism rules out the idea that we can bring ourselves under norms in acts that guide or shape themselves and also rules out the idea that bringing oneself under norms requires infinitely many distinct acts of this kind.

Acknowledging the cultural dimension of following norms, however, pushes us towards Source-App and Source-Resp. For, in order to think of the following of norms as the basis of the kind of normativity that is characteristic of a human life, there must be a difference between bringing oneself under norms and merely falling under norms or being disposed to be in accordance with norms. This pushes us towards Source-App, which is the idea that following a norm requires applying the norm to one’s

case at hand. It is then difficult to see, however, how agents can become intrinsically responsible for an act by applying a norm to their own case if they are not also intrinsically responsible for applying the norm correctly to their own case. In other words, it is difficult to see how we can reject Source-Resp.

Unfortunately, the resulting position is inconsistent. We must either reject naturalism, or we must reject Source-App or Source-Resp. If we reject naturalism, we have failed to understand rule-following as a natural phenomenon (unless we suggest a plausible way of understanding “naturalism” according to which naturalism is compatible with Auto-App or Infinity, in contrast to my stipulative use of “naturalism” in this paper). And if we reject Source-App or Source-Resp, it is unclear how we can account for responsibility-source acts, which are a necessary part of any genuinely cultural normative practice.

The arguments in this paper do not suffice to choose a particular one of the options to avoid this inconsistency. I hope to have shown, however, that reconciling a naturalistic (at least in my stipulative sense of this term) and a cultural perspective on rule-following is challenging. The challenge is a specifically philosophical one. Whatever surprising facts empirical research might reveal, the inconsistencies that were my topic here cannot be addressed by empirical research. The question how normative practices can arise by natural or cultural evolution may have fascinating answers; but they presuppose a consistent concept of the normative practices at issue. It hence seems to me that before we can address the empirical and etiological questions about norms and rule-following that are popular today, we need a philosophical account of how we can combine a cultural and a natural perspective on rule-following. If what I said in this paper is correct, we are currently not in possession of such an account.

References

- Boghossian, P. A. (2012). Blind rule-following. In Coliva, A., editor, *Mind, meaning, and knowledge: themes from the philosophy of Crispin Wright*, pages 27–48. Oxford University Press.
- Boyle, M. (2011). ‘Making up your mind’ and the activity of reason. *Philosophers’ Imprint*, 11(17).
- Brandom, R. B. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press, Cambridge, Mass.
- Heyes, C. (2024). Rethinking norm psychology. *Perspectives on Psychological Science*, 19(1):12–38. PMID: 37439763.
- Hieronymi, P. (2009). Two kinds of agency. In O’Brien, L. and Soteriou, M., editors, *Mental Action*, pages 138–162. Oxford University Press, Oxford.
- Kant, I. (1900 ff.). *Gesammelte Schriften*. Bände I–XXII hg. v. d. Königlich Preussische Akademie der Wissenschaften, Berlin 1900ff., Band XXIII hg. v. d. Deutschen Akademie der Wissenschaften, Berlin 1956, ab Band XXIV hg. v. d. Akademie der Wissenschaften zu Göttingen, Berlin 1966ff.
- Nieswandt, K. (2019). What is conventionalism about moral rights and duties? *Australasian Journal of Philosophy*, 97(1):15–28.
- Peregrin, J. (2023). *Normative Species: How Naturalized Inferentialism Explains Us*. Taylor & Francis.
- Pippin, R. B. (2005). Brandom’s hegel. *European Journal of Philosophy*, 13(3):381–408.
- Raleigh, T. (2024). The emptiness of naturalism. *Philosophy*.
- Reiland, I. (2024). Rule-following ii: Recent work and new puzzles. *Philosophy Compass*, 19(5):e12976.
- Satne, G. (2017). Brandom and the second person. *International Journal of Philosophical Studies*, 25(2):189–209.
- Scanlon, T. (1990). Promises and practices. *Philosophy and Public Affairs*, 19(3):199–226.
- Schmidt, M. F., Rakoczy, H., and Tomasello, M. (2011). Young children attribute normativity to novel actions without pedagogy or normative language. *Developmental science*, 14(3):530–539.
- Sellars, W. (1969). Language as thought and as communication. *Philosophy and Phenomenological Research*, 29:506–527.
- Sellars, W. (1997). *Empiricism and the philosophy of mind*. Harvard University Press.
- Spiegel, T. J. (2022). Why naturalism cannot (merely) be an attitude. *Topoi*, 42(3):745–752.
- Sripada, C. S., Stich, S., Carruthers, P., and Laurence, S. (2005). A framework for the psychology of norms. *The innate mind*, 2:280–301.

- Stroud, B. (1996). The charm of naturalism. *Proceedings and Addresses of the American Philosophical Association*, 70(2):43–55.
- Sultanescu, O. (2023). Meaning, rationality, and guidance. *Philosophical Quarterly*, 73(1):227–247.
- Tomasello, M. (2016). *A natural history of human morality*. Harvard University Press.
- Wittgenstein, L. (1953). *Philosophical investigations*. Blackwell, Oxford.
- Wright, C. (2004a). Intuition, entitlement and the epistemology of logical laws. *Dialectica*, 58(1):155–175.
- Wright, C. (2004b). Warrant for nothing (and foundations for free)? *Supplement to the Proceedings of The Aristotelian Society*, 78(1):167–212.
- Wright, C. (2007). Rule-following without reasons: Wittgenstein's quietism and the constitutive question. *Ratio*, 20(4):481–502.
- Wright, C. (2012). The rule-following considerations and the normativity of meaning. In Coliva, A., editor, *Mind, meaning, and knowledge: themes from the philosophy of Crispin Wright*, pages 379–401. Oxford University Press.