

Semantic Dispositionalism and the Rule-Following Paradox

Elek Lane

PREPRINT. PLEASE CITE PUBLISHED VERSION.

Metaphilosophy 53 (5): 685-695 (2022)

Abstract

In virtue of what does a sign have meaning? This is the question raised by Wittgenstein's rule-following considerations. Semantic dispositionalism is a (type of) theory that purports to answer this question. The present paper argues that semantic dispositionalism faces a heretofore unnoticed problem, one that ultimately comes down to its reliance on unanalyzed notions of repeated types of signs. In the context of responding to the rule-following paradox—and offering a putative solution to it—this amounts to simply assuming a solution to the problem in one domain and using it to solve the same problem in another. Given, moreover, the level at which the rule-following paradox undercuts dispositionalism—the level of the notion of a sign's repetition—the objections made to the view also rule out causal/informational theories of meaning as well as communitarian/assertion-theoretic ones as potential solutions to the rule-following paradox.

Keywords

Kripke, rule-following paradox, semantic dispositionalism, theory of meaning, Wittgenstein

Since any set of causes whatsoever will have endless properties in common, we must look to some recurrent feature of the gatherer, some mark that he or she has classified cases as similar. This can only be some feature or aspect of the gatherer's reactions ... in which case we must once again ask: what makes these reactions relevantly similar to each other? Wittgenstein's problem once again.

—Donald Davidson (2001, 4–5)

1 Overview

There are many noises, some produced by humans. A subset of those have meaning. In virtue of what does the meaningful noise differ from the meaningless? I take this to be the question that leads to Ludwig Wittgenstein's rule-following paradox (see Wittgenstein 1953, § 201). There is little agreement on the significance of the rule-following paradox, both for understanding Wittgenstein's corpus and for philosophy in general. One thing that philosophers do tend to agree on, however, is that the paradox poses a challenge to the very possibility of a subject's meaning something by a sign.

Semantic dispositionalism is a (type of) theory about what grounds the act of meaning something by a sign. According to the theory, it is in virtue of having certain dispositions to deploy a sign in particular ways that one means something determinate by an actual deployment of that sign. It is, for example, in virtue of the fact that I am disposed to respond with the sum when prompted by sentences that fill out the schema “ $x + y = ?$ ” that by “+” I mean the plus function.¹

While much of the debate around semantic dispositionalism has centered on *ceteris paribus* clauses, I think this distracts from the real problem the theory faces. The rule-following paradox undercuts dispositionalism before anything like a *ceteris paribus* clause might be usefully invoked by the theory.²

In order to home in on the trouble with dispositionalism, I center the discussion here on versions of the view that do away with *ceteris paribus* clauses. I call these the “hard-nosed variants” of dispositionalism. This should not, however, be taken to indicate that the problem for the hard-nosed variants is special to that version of the view. The problem is perfectly general, applying to dispositionalism’s soft-nosed variants, which employ *ceteris paribus* clauses as well.

Ultimately the problem for semantic dispositionalism comes down to its reliance on unanalyzed notions of repeated sign-types. In the context of responding to the rule-following paradox—and offering a putative solution to it—this amounts to simply assuming a solution to the problem in one domain and using it to solve the same problem in another. (The first domain is that of the repetition of a *sign*; the second, that of the repetition of a *symbol*. I understand a sign to be the external aspect of a symbol, and I use “symbol” to mean a sign in a recognizable logical use.³ So—in English—“bank” is one sign but potentially two symbols.) This might prove illuminating in some limited sense, but such a strategy cannot solve the rule-following paradox and thus ought to be considered dissatisfying by the dispositionalist’s own lights.

Given, moreover, the level at which the rule-following paradox undercuts dispositionalism—at the level, that is, of the notion of a sign’s repetition—the objections made to the view also rule out causal/informational theories of meaning as well as communitarian/assertion-theoretic ones as potential solutions to the rule-following paradox. So, while this paper’s discussion centers on a rather special kind of theory of meaning—namely, that of semantic dispositionalism’s hard-nosed variants—the lessons to be drawn are quite general.

¹“+” and the plus function have been staples of the literature since Kripke 1982.

²Semantic dispositionalism has faced numerous different objections. Many of these (though not all) turn out to be issues that fall under the *ceteris paribus* clause debate (for example, objections that center on sincerity and lying, understanding, personality quirks, finitude, and so on—basically any issue that raises the specter of so-called aberrant dispositions). This is why I talk about *ceteris paribus* clauses in general instead of naming the myriad objections encompassed by the discussion of such clauses. As stated, my goal is to render that debate moot.

³This terminology comes from Wittgenstein’s *Tractatus* (see 1922, 3.32, 3.321, and 3.326). It is no accident that the two terms are inter-defined, so that comprehension of one relies on comprehension of the other. An upshot of the present paper is that these terms *must* be inter-defined.

2 The Traditional Problem for Semantic Dispositionalism

In his 1982 book, *Wittgenstein on Rules and Private Language*, Saul Kripke argues that there are no “straight” solutions to Wittgenstein’s rule-following paradox.⁴ One candidate for a straight solution that Kripke considers and dismisses is semantic dispositionalism. Let us begin by reviewing his objection.⁵

Kripke offers the following statement of dispositionalism (1982, 26):

The referent φ of ‘f’ is that unique binary function φ such that I am disposed, if queried about ‘f(m, n)’, where ‘m’ and ‘n’ are numerals denoting particular numbers m and n, to reply ‘p’, where ‘p’ is a numeral denoting $\varphi(m, n)$.

There is, of course, much that would require tweaking and filling out if one were to propose a serious semantic dispositionalism. For both our purposes and Kripke’s, however, this rough schema will suffice. Actually, the fact that Kripke offers a *schema* for dispositionalism is useful, as it makes it easier to attend to structural features of the theory. Notice in particular two of those features: First, in order to invoke the relevant dispositions (out of the indefinitely many that any subject has), Kripke’s statement of dispositionalism employs the notion of a *stimulus-type*. In this case, it is that of being “queried about” a particular sentence schema, but the important point is that there is some such stimulus that any theory of dispositionalism will have to invoke. Second, the statement employs the notion of a *response-type*, and this response-type is necessarily the sort of thing that can be identified as repeating across many instances. In Kripke’s schema, it is the notion of a particular numeral, “p.”

Kripke points out that the statement of dispositionalism above is false. Consider that although a speaker—call her Gertrude—may well be disposed to respond with the sum when prompted by *some* sentences of the form “ $x + y = ?$ ” this is not true for *all* sentences of that form. Gertrude is in other cases disposed to respond with the number that is one greater than the sum (because she has a knack for making little mistakes), roll her eyes (because she does not like simple arithmetic), or even just sneeze (because hearing “57” makes her do that). All such responses must be ruled out by the proponent of dispositionalism because none of these gives the correct output to the plus function in the relevant cases. If Gertrude is disposed to do such things, then the set of meaning-determining dispositions must be a proper subset of the things she is disposed to

⁴Kripke defines a straight solution as follows: “Call a proposed solution to a sceptical philosophical problem a *straight* solution if it shows that on closer examination the scepticism proves to be unwarranted; an elusive or complex argument proves the thesis the sceptic doubted” (1982, 66). A straight solution is opposed to a “skeptical solution” that accepts the conclusion of the rule-following paradox—that is, that there can be no such thing as meaning anything by any sign—but nonetheless finds some way of rehabilitating our talk of meaning and rule following.

⁵In fact we will review just one of Kripke’s objections. Another of his objections, which will go undiscussed beyond this footnote, is that there is a supposed mismatch between *what I am disposed to do* and *what I should do* and that dispositionalism’s plausibility turns on a conflation of these (see Kripke 1982, 23). The dispositionalist has an obvious response, though, since practical considerations about consistency—something that is surely necessary if one is to be a communicator—seem sufficient to bridge any gap between one’s dispositional profile and the set of actions they ought to take, all things considered.

do in response to the stimulus-type “ $x+y=?$ ”. I call this dispositionalism’s Scylla: the danger of failing to sufficiently limit the set of meaning-determining dispositions and thus allowing in aberrant dispositions that distort the intuitive meaning of, for example, “+”.

Dispositionalism’s Charybdis is the danger of successfully limiting the set of meaning-determining dispositions only at the cost of circularity. To rule out the aberrant cases, one might consider adding a *ceteris paribus* clause to the statement of dispositionalism. Instead of saying that Gertrude means *plus* by “+” if she is disposed to respond with the sum whenever suitably prompted by that sign, the dispositionalist can say that Gertrude means *plus* by “+” if, *ceteris paribus*, she is disposed to respond with the sum when suitably prompted. The *ceteris paribus* clause is intended to rule out those cases in which she is disposed to err, sneeze, or have any other untoward reaction to the stimulus.

Yet with the addition of this clause, the account looks problematically circular. For how can the dispositionalist spell out what it is for not all else to be equal? Given the presumed diversity of any subject’s aberrant dispositions, it would seem that the only real specification available is that they are those dispositions that incline Gertrude not to respond with the sum. (Or so we might suspect.) If this is right, then what was supposed to be an analysis of what it is to mean something by a particular deployment of a sign is hardly illuminating. It appears we must employ the notion of a correct response in order to explain what constitutes correctness of response.⁶

Dispositionalism must rule out cases of aberrant dispositions if it is to offer even a possibly true analysis of what it is to mean something by a sign. Yet this must be done without simply appealing to the function/rule/meaning in question if the account is to avoid being merely circular.⁷ It is the attempt to chart such a course that much of the post-Kripkean debate around dispositionalism has focused on.⁸ It is this literature that has generated the bulk of dispositionalism’s hard-nosed variants.

3 Warren’s Hard-Nosed Dispositionalism

Jared Warren (2020) provides an example of a hard-nosed variant that is instructive, recent, and sophisticated. His general strategy is to introduce distinctions between various kinds of dispositions without making use of intentional or semantic notions. Then he defines the meaning-determining dispositions in terms of those.

⁶Perhaps on a charitable reading, semantic dispositionalism here gives us some kind of platitude about meaning: “If Gertrude means *goat* by ‘goat,’ then she’ll use ‘goat’ to speak of the goats, all else being equal.” This may highlight an aspect of the very concept of meaning and, so taken, appears to have the merit of not being false. Still, it does not provide the form of illumination that the dispositionalist is after.

⁷Some, though, argue that *ceteris paribus* clauses had better be O.K., since they are used in science all the time (see Fodor 1991; 1992, chap. 4). (These philosophers defend versions of dispositionalism that constitute its soft-nosed variants.) Others argue that dispositionalists’ *ceteris paribus* clauses are not really like those of science at all, and so they demur (see Kusch 2005 and see Cheng 2009 for a reply). I ignore this debate here.

⁸I take this all to go both for realist views about dispositions and for views on which dispositions are to be analyzed counterfactually (see Heil and Martin 1998 for a version of the former sort of view about rule-following dispositions and Ginet 1992 for a version of the latter).

We can pass over the details and ask what the final form of his view looks like. Since he considers Kripke's objection to bifurcate into two distinct challenges, his answer comes in two parts (Warren 2020, 268, 272).⁹

Ludwig's composite dispositions, specified in terms of his singular dispositions, determine the answer he is "disposed" to give when presented with any query about "+", thus allowing us to completely fill out Ludwig's disposition table.

Let's bring everything together: to fill out Ludwig's disposition table, say that a numeral "m" is entered at row *n*, column *k*, just in case Ludwig has an *M*-general disposition to stably reply with "m" when queried about "*n* + *k*" in normal situations (that is "m" is Ludwig's stable answer to the query in the overwhelming majority of normal situations).

In the statement of his dispositionalism, Warren fills out Kripke's basic schema for the theory. The subject's meaning-determining dispositions are singled out in part by the fact that they are responses to a particular stimulus-type, namely, being queried about "+" or "*n* + *k*". And once again, a response-type is presumed to be identifiable as (potentially) repeating on different occasions.¹⁰

It is no accident that dispositionalism takes this form, that is, identifies both response-types and stimulus-types as simply *given* for the purposes of the theory. A stimulus-type must be invoked in order to (begin to) single out the relevant dispositions, because any human has countless dispositions, and the dispositionalist cares about only a portion of those. The invocation of a stimulus-type is what allows the dispositionalist to single out that portion.

The notion of a response-type is equally necessary for a statement of the theory. Once again, think of Gertrude, who is an English speaker grown to the age of reason. She uses the word "goat" to pick out the caprine things. Thus, according to dispositionalism, her meaning *goat* by "goat" is grounded in her dispositions to so call members of the species *aegagrus*. "Goat," then, is singled out as a response-type. It is assumed to be identifiable independently of its meaningful deployment to refer to goats. Were it not, the account would be circular, since the conditions of the term's meaningful deployment are precisely what dispositionalism seeks to explain. This is why the term has to be taken as identifiable independently of its (correct) usage.

4 Kripke/Wittgenstein's Rule-Following Paradox

I introduced semantic dispositionalism as a response to a certain threat, namely, that of meaning skepticism. This is what Wittgenstein's rule-following paradox threatens and what is expounded so forcefully in Kripke's

⁹In Warren's paper, to "fill out Ludwig's disposition table" is equivalent to isolating the set of an arbitrary subject's meaning-determining dispositions.

¹⁰Specifically, Warren writes that a numeral—"m"—is to be singled out as Ludwig's answer to the query "just in case Ludwig has an *M*-general disposition to stably reply with 'm'."

book. So let us review what Wittgenstein and Kripke say about rule following and the conceptual difficulties surrounding it.

I do not intend to run through Kripke or Wittgenstein's *arguments* so much as the *steps* each takes in presenting the paradox. The goal is to achieve an overview of how the rule-following dialectic progresses while abstracting away from the finer details.

In Kripke's dramatization, he asks us to imagine a skeptic who challenges us to explain what makes it the case that our previous and current uses of the plus sign conform to each other, that is, that we are using the sign consistently over time (Kripke 1982, 8). It is assumed (for purposes of exposition) that we mean *plus* by "+" in the present case, but what made it the case that in the past we used "+" for the plus function and not, say, the "quus" function?¹¹

Suppose that nothing in my past history—actions, thoughts, dispositions, or anything else—determined that my uses of "+" were deployments of the *plus symbol* as opposed to the *quus symbol* (which we can stipulate to be written the same). If that is so, then it is hard to see what could ground my meaning *plus* (not *quus*) by "+" in the *present* case. The dialectic thus turns from being focused on consistency with past use to the possibility of making sense of *current* use.¹²

The result is that if the skeptic's argument is unanswerable, there is no such thing as meaning *plus* by "+"—or, for that matter, meaning *anything* by *any* sign (see Kripke 1982, 19). And this means that there is no correct (or incorrect) response to " $58 + 67 = ?$ " or, in general, a correct or incorrect way of using any given string of signs at all. ("Sentence" ceases to feel as if it is an appropriate description of such a string.) Thus, the initial epistemological guise of the skeptic's challenge—in which it is asked how we know our current use of a particular symbol is consistent with our past uses of that same symbol—turns out to be only a veneer.¹³ We are ultimately faced with the impossible, *metaphysical* conclusion that there is no such thing as meaning anything by any word.

Wittgenstein's exposition is worth considering too, since it brings out the extent to which the problem is one of sameness and repetition.¹⁴ Wittgenstein asks his reader to consider a pupil who is learning to count by twos (1953, § 185). The student seems to have gotten the hang of it. Around the five-hundredth iteration of the *add two* operation, however, he continues thus: "998, 1000, 1004, 1008. ..." The pupil is stopped and his mistake pointed out. He is bewildered. He thought that *that* was what he was *supposed* to do; that he

¹¹Kripke (1982, 9) defines the "quus" function as follows: x quus y equals x plus y if x and y are less than 57. Otherwise x quus y equals 5.

¹²This shift is explained at Kripke 1982, 13. Concurring with Gary Ebbs (1997, 19–20), I take Kripke's primary interest to be in making sense of current uses of language, *punkt*. Thus, I think the dialectical purpose of his focus on past use is to sidestep issues about speaking "inside/outside" language. The purpose of this paper, however, is not Kripke exposition, and I think all my philosophical points go through even if my reading of him is wrong.

¹³Kripke is explicit about this at Kripke 1982, 21.

¹⁴Though note that Hannah Ginsborg (2020, 13–14) challenges the equation of issues of sameness/repetition with the (most fundamental) problem of rule following.

was continuing in the same way as before; that he was merely *repeating* the same step over again; and so on. It would appear that he had connected “add two” not with the *add two* operation but with some deviant one instead, one in light of which his actions *would* have been correct had “add two” expressed that deviant operation.

The challenge faced is basically the same as that described by Kripke. One wants to know what makes “1004” an incorrect continuation after “1000” (and “1002” the right one). If no answer presents itself, it will appear that there is no such thing as the correct or incorrect use of any sign. As Wittgenstein (1953, § 185) puts it, “[T]here would be neither accord nor conflict here”: we will apparently lose the notion of rule-governed action or behavior altogether.

Now, the problem goes deeper than a skeptical attack on the notion of a *symbol* (that is, a sign with conditions of correct use). Wittgenstein brings the point out by asking his reader to consider the series “2, 2, 2” and observing that there exists a question of what it would be to continue *that* series just as much as there exists one regarding the continuation of the series governed by the operation *add two* (Wittgenstein 1956, I-3):

“But you surely know for example that you must write the *same* sequence of numbers in the units: 2, 4, 6, 8, 0, 2, 4, etc.”—Quite true: the problem must already appear in this sequence, and even in *this* one: 2, 2, 2, 2, etc.—For how do I know that I am to write “2” after the five hundredth “2”? i.e. that ‘the same figure’ in that place is “2”?

Wittgenstein’s point is that the same type of rule-following or normativity that governs the *add two* operation is at play when it comes to determining what counts as the bare repetition of a sign.¹⁵ If we are compelled by the skeptic to conclude that there is no such thing as rule following, we give up not only the idea of there being a correct way to deploy any given sign but also the very idea that a sign can repeat. And I think this is just tantamount to losing the idea of a sign altogether.

Thus, there are two facets to the problem of rule-following. I say these are *two* facets of *one* problem because each facet hinges on the same issue, namely, that of the need to view rule followers as acting, *qua* rule followers, self-consciously or with a special self-understanding. For an object to merely pattern in a particular way—to trace a path through space and time that happens to instantiate certain identifiable regularities—is not yet for that object to follow a rule. In order to follow a rule, that object must trace that particular path through space and time in part *because* of its awareness of the rule, and its awareness that that rule determines that path as correct, so that deviations from it count as mistakes.

¹⁵The quote may seem ambiguous between being a claim about what counts as the bare repetition of a sign and what counts as the five-hundredth instance of the *rule* governing the series “2, 2, 2...” (If it is the latter case that Wittgenstein has in mind, then concerns about deviant mathematical functions are relevant. If the former, then not.) I favor the former reading due to the use of the term “the same figure” (*die gleiche Ziffer*). Wittgenstein begins this passage by considering what it means to continue the same “sequence of numbers” (*Zahlenfolge*) and then contrasts this with continuing a sequence of the same *numeral* (or “figure”). He moves, in other words, from considering units that bear both orthographic and semantic properties (numbers) to merely orthographic units (numerals).

The first facet of the problem occurs at the level of the *symbol*, that is, a sign with a determinate standard of correct use: one can look at a subject's action—writing down “998, 1000, 1004,” say—and ask what makes each of those actions correct in the light of the rule expressed by a particular symbol. If there is no answer, it seems, we are threatened by a loss of the distinction between going on in the same way as before (that is, in accordance with the rule *then* and in accordance with that same rule *now*), as opposed to going on in any which way with no conceivable standard of correctness. Thus, rule followers no longer appear able to deploy *symbols*, since we are no longer able to regard the use of a *sign* as having a normative dimension.

Second facet: in writing down “998, 1000, 1004,” one tokens three action-types: those of writing down “998,” “1000,” and “1004.” Yet now we can ask: In virtue of what are the actions I performed the first time I wrote down “998, 1000, 1004” the same as the actions I performed in writing down the ostensibly same sequence the second (third, et cetera) time? In the same way the rule-following paradox causes us to lose grip of what it is to repeatedly perform the *add two* operation (and so challenges the intuitive idea that there is such a thing as a symbol), it also causes us to lose our grip on what it is to produce the same *sign*—that is, a *merely* orthographic unit—twice. Perhaps instances of “1000,” properly understood, are any instances of the numeral “1” followed by any number of “0”s, or a single “1” and three “0”s written in any order, or et cetera. To regard “1000” as a sign is to understand that only a “1” followed by three “0”s counts as a repetition of it and thereby to recognize that the following token—“1000”—counts as the same sign but that “10000” and “0010” do not.

With respect to any two utterances (or instances of writing down), there are both similarities and differences between them (acoustic/phonetic, geometric/orthographic, spatiotemporal, and so on). We would ordinarily think there to be a rule that identifies “Thanks-GIV-ing” and “THANKS-giv-ing,” or “COO-pon” and “KYOO-pon,” as tokens of the same type of sign—but not “PRO-duce” and “pro-DUCE” or (the standard pronunciations of) “coup” and “queue.” This is not a matter of identifying distinct components of a sign and apprehending which elements of their composition are significant to the determination of the sign they make up (as was the case of “1000” above). It is rather a matter of identifying which elements of the sign's physical realization are significant to the sign's identity. There is, after all, tremendous variation in how the same word is pronounced/written not only within a given language but also in dialects and even idiolects, depending on the sentential context in which a sign occurs. To be able to recognize which variations in, for example, the phonetic field are significant (and which are not) is to be master of a set of rules governing what counts as a repetition of a sign and what counts as the production of a different sign altogether. Thus, if we accept the “skeptical conclusion” of the rule-following paradox, we are no longer able to see “O” as a token of a type (a type such that what counts as an instance of it is determined by a rule). The result is that nothing can fund our distinguishing between “O” and “0,” or indeed between “O” and “Q” or

“D” or even “CHÈVRE,” as instances of different signs.¹⁶

At the first stage, the rule-following paradox appears to rob us of the notion of a particular action’s being an instance of a rule-governed activity. In the linguistic case, this is to lose the notion of a symbol, that is, a sign that is capable of being used in a way that is subject to normative constraints. Yet at this first stage we maintain the illusion of being able to identify the production of particular signs *as such*, that is, as repeatables or tokens of a type that may occur again on occasions distinct in both space and time. At the second stage, however, we lose even that. The loss of normativity visited upon us by the rule-following paradox equally infects our ability to view an action—the production of a particular sign, say—as itself being an instance of a repeatable kind (regardless of whether any given instance of the kind might be used *correctly* because part of some rule-governed activity). Now if I respond to the question “ $58 + 67 = ?$ ” by saying “125,” it will be impossible to say not only whether what I have said is correct (there being, at this point, no distinction between correctness and incorrectness) but also whether the sounds I emit can truly be regarded as an attempted answer at all. For we can no longer understand those sounds to be the sort of thing that can so much as be repeated (and any answer must be, in principle, repeatable). In fact, the same considerations that apply to my so-called “answer” apply to the “question.” At this point, “speech” seems to be no more than noise, and “linguistic” “actions” appear formless.

5 The New Problem for Semantic Dispositionalism

All dispositionalist accounts rely on taking sameness of stimulus- and response-types as simply *given* for the purposes of the theory. But exactly such a notion is put under threat by the rule-following paradox. For this reason, it is philosophically illicit to presuppose it in the context of the skeptical dialectic on rule following.

In dispositionalists’ pursuit of an explanation of Gertrude’s capacity to use a symbol, they establish an expansive dispositional profile centered on how she inclines to respond to a certain prompt—something with the plus sign, say. In order for the responses she is disposed to give to be relevant at all, we must think of Gertrude as regarding the repeated prompts that her dispositional profile centers on as all involving the same sign, namely “+”. She must (somehow) register the fact that that sign is common to the various stimuli that prompt her. If she does not, then the fact that she is disposed to respond in different ways (sometimes with the sum, other times the sum-plus-one) is irrelevant. By the same token, if she were disposed to respond consistently with the sum, that too would be irrelevant—were this to happen, it would be pure happenstance, unrelated to the fact that she had been prompted by the plus sign. That the subject regards all the prompts (in terms of which her dispositional profile is defined) as being the *same* (kind of) prompt is an assumption

¹⁶Both Silver Bronzo and James Conant discuss a similar issue (see Bronzo 2017, 1356–58; Conant 2020, 12–14).

relied upon by the basic schema for dispositionalism.¹⁷

I cry foul. As I argued in the preceding section, the capacity to recognize the same (type of) sign across various tokenings is a perfect instance of what we are seeking to explain. If Gertrude is to be prompted by various verbal tokenings of “plus,” these will all have different acoustic (and countless other) properties.¹⁸ Thus there needs to be some rule, recognized by Gertrude, that binds together the disparate tokens into a unified type. In other words, Gertrude must already be a rule follower in order to recognize these prompts as tokens of the same stimulus-type; and without that assumption, dispositional analyses are worthless as part of an account of what it is to mean something by a sign.

The fact that Gertrude has certain specifiable dispositions, then, does not come in below the level at which she is recognizable as a rule follower; it cannot ground the fact that she is able to engage in rule-governed activity in general or the meaningful deployment of signs. It is, rather, the other way around: Gertrude’s capacity to follow a rule is something that the dispositionalist relies on in making any (plausible) claim about her being disposed to respond in specifiable ways to a given stimulus-type. If dispositionalists say that Gertrude is disposed to respond with the sum when queried with “+”, they implicitly count on her capacity to recognize distinct tokens of “+” as being of a single type.

These considerations apply *mutatis mutandis* for response-types. Gertrude has many dispositions to respond when prompted by “+”. In those responses, she is disposed to frequently deploy the numeral “0,” by itself and in compounds. We regard this as so many tokens of the *same* numeral. Does Gertrude? If she does not, the dispositional table based on her responses is once again worthless. It is a necessary assumption that she take her responses to deploy repeatable tokens of “0” (or whatever else). So once again the dispositionalist strategy presupposes an instance of what it is seeking to explain.¹⁹

Here is a quick summary of the dialectic thus far.

Skeptic: What grounds Gertrude’s capacity to follow a rule in the case of the meaningful deployment of a sign?

Dispositionalist: Gertrude’s dispositional profile.

¹⁷An illustration: Imagine there are some Martian scientists who determine that we English speakers tend to have the following disposition: when prompted by a certain sign, we consistently respond with the sum. They conclude that we mean *plus* by that sign. It turns out, however, that the Martian plus sign strikes us as being a highly gerrymandered kind. We have to revert to a bizarre, disjunctive description of it: either “+” or a concatenation of noises too varied to bear a convenient description in English. I think that the Martians are wrong and that we anglophones do not mean anything by their plus sign. Their mistake, it seems, is in assuming that their subjects regard each prompt as containing instances of a particular type of sign (namely, the Martian plus sign) when in fact they do no such thing.

¹⁸Though it is unrealistic to suppose that any two human vocalizations of “plus” have *exactly* the same acoustic properties, even if they did, they would have different spatiotemporal properties. Thus, to recognize them as tokens of the same type is thereby not to regard them as tokens of a Goodmannian acoustic-property-shifting type.

¹⁹A further illustration: Suppose another Martian—call him Goodman—comes along and sees Gertrude’s dispositional profile. Goodman sees all the responses she is disposed to give as tokens of a single type, and the type just happens to be the Martian numeral zero. Being a dispositionalist, Goodman thus concludes that by “+” Gertrude means the function that always returns zero, no matter what the input is. Goodman, of course, is wrong about what Gertrude means by “+”, and his mistake lies in assuming that Gertrude would regard all the responses she is disposed to give as tokens of a single type (the Martian numeral for zero), as opposed to tokens of many types (generally corresponding to the sum of the arguments).

Skeptic: What grounds Gertrude's capacity for rule following, implicitly relied on in the construction of her dispositional profile?

Dispositionalist: ...

I am open to suggestions about how the dispositionalist might fill in the ellipsis, but it should be clear that it cannot be filled in by any version of the dispositionalist's first answer. Some other *kind* of answer would have to be proffered.

One might wonder whether we cannot define Gertrude's dispositional profile in terms of *causally* relevant features of stimulus-types and the responses they beget. This would (appear to) be tantamount to grounding Gertrude's rule-governed actions in a nonnormative base of facts, and we would thus sidestep the need for her *recognition of sameness* upon which I relied so heavily in the foregoing arguments. Such a strategy seems promising for understanding (genuinely dispositional) properties such as fragility and solubility. Why should it not work in this case too?

My answer, in short, is that the difference between being fragile and being a rule follower is that to be the latter implies consciousness of oneself as such, which is in turn connected to the rule follower's being an appropriate subject of normative evaluation. If in counting by twos Gertrude says "998, 1000, 1004," she makes a mistake. But not so if she fails to break when dashed upon the floor, despite her fragility. So the question we ought to ask is this: Why in the counting case is Gertrude doing something wrong, as opposed to merely failing to instantiate some particular regularity definable in terms of her dispositional profile? I think she is wrong in the former case (but not the latter) because she intends to conform to the *add two* rule. Her act is a self-conscious one. Self-consciousness cannot be eliminated in favor of a merely causal property, since doing so makes us lose sight of the fact that Gertrude's rule-governed actions occur in a normative nexus.

What goes for the symbol goes for the sign. Just as the *add two* rule determines what actions count as correct iterations of it, so too there is a rule that determines "O" and "O" to be repetitions of the same sign, while "CHÈVRE" is not. Imagine a pupil learning to write: we would count the first two signs in the preceding sentence (presented in quotation marks), but not the third, to be legitimate repetitions of "O." But if to write "CHÈVRE" is to constitute a mistake on the pupil's behalf, it is because the pupil intends to write "O," that is, to produce a token of that type. If the pupil is merely caused to produce "CHÈVRE," then her action may no more be viewed as incorrect than may a fragile vase's failing to break when dashed upon the floor.

Now it might seem that even if my arguments are right, there is still something of potential philosophical interest in dispositionalism. Perhaps the theory appears to promise a semi-reductive analysis of what it is to mean something by a sign. Why not (the thought goes) take recognition of repeatable response- and

stimulus-types as primitive and then ground our more complex capacities to engage in rule-governed activity in that?²⁰

There are two things to say about this. The first is that such a hypothesis would have to be vindicated by empirical evidence. In essence it is a claim about how various human cognitive capacities relate to one another. This is an interesting topic (and not an intractable one), but progress in that area is more likely to issue from the fields of cognitive science, linguistics, and so forth, than from armchair philosophy.²¹

The second thing to say is that even if a semi-reductive analysis were established via empirical means, it would (obviously) not be a *full* reduction of our capacity to follow a rule. That is, should the research project gestured at above be developed *even to the point where it has by its own lights no more questions to answer*, it would nonetheless not be a complete answer to the rule-following paradox. This is because it would still rely on unanalyzed notions of at least one aspect of our capacity for rule following, that is, our capacity to recognize sameness across various tokens of whatever the theory's favored stimulus- and response-types turn out to be.

I earlier narrowed down our discussion of dispositionalism to its hard-nosed variants: versions of the view that steer clear of dispositionalism's Charybdis by eschewing *ceteris paribus* clauses. In fact, though, my critique applies to dispositionalism's soft-nosed variants, which do employ *ceteris paribus* clauses, equally well. Statements of dispositionalism, even when equipped with *ceteris paribus* clauses, fill out Kripke's original schema; that is, they make use of stimulus- and response-types that are taken as simply given for the theory. Similarly, for the purposes of the rule-following dialectic, causal/informational theories of content count as versions of dispositionalism.²² The reason is that the same basic schema of a particular stimulus-type being correlated with a particular response type is at play in causal/information theories of meaning. They thus fall to the same general objection I brought to bear against other forms of dispositionalism.²³

6 Conclusion

At the beginning of this essay, I said that asking a certain question leads to the rule-following paradox. That question (presented here in three different guises) was: What differentiates *meaningful* noise from *mere* noise? What is required for a noise or mark to be an element of a language? Or, in the terminology of the *Tractatus*,

²⁰Ginsborg (2011, 2020) articulates such a view.

²¹An example of what I have in mind is Chomskyan generative grammar. This provides us with a good candidate for a basic capacity for human rule following, but it is a matter for linguistics and biology (broadly construed) to investigate, not philosophy.

²²See Paul Boghossian: "[I]n all essential respects, a causal theory of meaning is simply one species of a dispositional theory of meaning" (1989, 527). See also Warren (2020, 258).

²³This is not exactly to say that causal/information theories of meaning (or for that matter dispositional ones in general) are false. It is to say that they are inadequate as responses to the rule-following paradox. In the context of other inquiries, they may yet be illuminating.

what turns a *sign* into a *symbol*?

Accepting that these questions must be answered in the terms in which they are posed leads to the rule-following dialectic and ultimately to skeptical paradox. This is because they all invoke a conception of a sign as a bare repeatable, the production of which is nonetheless subject to normative evaluation. What I think the rule-following paradox shows is that there is no such coherent conception of a sign. To recognize something as a sign at all requires the exercise of a capacity for rule following. For this reason, any fundamental account of meaning that starts with the notion of a sign—ostensibly a sign *as such*—and asks what it is that must be *added* to it in order that it achieve the status of *symbol*—that is, that it count not just as a “bare repeatable” but a repeatable that has conditions of appropriate (true, felicitous, what-have-you) use—is bound to be undercut by the rule-following paradox.²⁴

Acknowledgements

Drafts of this paper were presented at the University of California, Berkeley, dissertation seminar as well as the Berkeley-London graduate philosophy conference. In addition, Sonya Bennett-Brandt, Hannah Ginsborg, Jared Warren, and an anonymous referee provided draft comments. This all helped a lot, and I am grateful to everyone.

References

- Boghossian, Paul. 1989. “The Rule-Following Considerations.” *Mind* 98, no. 392: 507–49.
- Bronzo, Silver. 2017. “Wittgenstein, Theories of Meaning, and Linguistic Disjunctivism.” *European Journal of Philosophy* 25, no. 4: 1340–63.
- Cheng, Kai-Yuan. 2009. “Semantic Dispositionalism, Idealization, and Ceteris Paribus Clauses.” *Minds and Machines* 19, no. 3: 407–19.
- Conant, James. 2020. “Wittgenstein’s Critique of the Additive Conception of Language.” *Nordic Wittgenstein Review* 9: 7–36.
- Davidson, Donald. 2001. “Externalisms.” In *Interpreting Davidson*, edited by Petr Kotatko, Peter Pagin, and Gabriel Segal, 1–16. Stanford, Calif.: CSLI.
- Ebbs, Gary. 1997. *Rule-Following and Realism*. Cambridge, Mass.: Harvard University Press.
- Fodor, Jerry. 1991. “You Can Fool Some of the People All of the Time, Everything Else Being Equal; Hedged Laws and Psychological Explanations.” *Mind* 100, no. 1: 19–34.
- Fodor, Jerry. 1992. *A Theory of Content and Other Essays*. Cambridge, Mass.: MIT Press.
- Ginet, Carl. 1992. “The Dispositionalist Solution to Wittgenstein’s Problem About Understanding a Rule: Answering Kripke’s Objections.” *Midwest Studies in Philosophy* 17, no. 1: 53–73.
- Ginsborg, Hannah. 2011. “Primitive Normativity and Skepticism About Rules.” *Journal of Philosophy* 108, no. 5: 227–54.
- Ginsborg, Hannah. 2020. “Wittgenstein on Going On.” *Canadian Journal of Philosophy* 50, no. 1: 1–17.
- Heil, John, and C. B. Martin. 1998. “Rules and Powers.” *Philosophical Perspectives* 12: 283–312.

²⁴I argue that communitarian/assertion-theoretic accounts of meaning fall to this objection in Lane 2017, 70–73.

Kripke, Saul. 1982. *Wittgenstein on Rules and Private Language*. Cambridge, Mass.: Harvard University Press.

Kusch, Martin. 2005. "Fodor v. Kripke: Semantic Dispositionalism, Idealization and Ceteris Paribus Clauses." *Analysis* 65, no. 2: 156–64.

Lane, Elek. 2017. "Excursus on Wittgenstein's Rule-Following Considerations." *Nordic Wittgenstein Review* 6, no. 1: 53–83.

Warren, Jared. 2020. "Killing Kripkenstein's Monster." *Noûs* 54, no. 2: 257–89.

Wittgenstein, Ludwig. 1922. *Tractatus Logico-Philosophicus*. Translated by C. K. Ogden. London: Kegan Paul.

Wittgenstein, Ludwig. 1953 [1958]. *Philosophical Investigations*. Translated by G. E. M. Anscombe. Oxford: Basil Blackwell.

Wittgenstein, Ludwig. 1956 [1967]. *Remarks on the Foundations of Mathematics*, edited by G. E. M. Anscombe, Rush Rhees, and G. H. von Wright. Translated by G. E. M. Anscombe. Cambridge, Mass.: MIT Press.