Wittgenstein on Rule-Following

Roderick T. Long

[For Kelly Dean Jolley, ed., *Wittgenstein: Key Concepts*]

The Rule-Following Paradox

I shall begin by misdescribing the moral of Wittgenstein's rule-following paradox, because I take the misdescription to be a helpful one, more of a ladder than a stumbling-block (though it should be borne in mind that it is always possible to trip on a ladder).

The moral of the rule-following paradox, then, is that what rule one is following when one acts is (radically) underdetermined by anything in either one's actions or one's thoughts. Underdetermined by anything in one's actions, since every actual sequence of behaviour is finite, and so is capable of being extended in an infinite variety of ways, each one corresponding to a different rule, all such rules being equally consistent with the behaviour thus far exhibited. (Worse yet, even an *infinite* sequence of behaviour, were one possible, wouldn't settle the matter, since such a sequence would be consistent with both a flawless execution of rule A and a bungled attempt to apply rule B.) And underdetermined by anything in one's thoughts, because no matter what the agent may have before her mind's eye, as it were, it doesn't count as such-and-such a rule unless the agent reliably applies it in such-and-such a way:

> I cannot know what he's planning in his heart. But suppose he always wrote out his plans; of what importance would they be? If, for example, he never acted on them. … Perhaps someone will say: Well, then they really aren't plans. But then neither would they be plans if they were *inside* him, and looking into him would do us no good. (*Last Writings on the Philosophy of Psychology* I. 234-235.)

Nothing in the mind seems to settle what I mean; everything depends on how a given mental item gets expressed in practice. But no amount of external conduct settles what I mean either. A move in chess, for example, is not simply a matter of "moving a piece in such-and-such a way" (the machinelike option), but neither does it consist in "one's thoughts and feelings as one makes the move." (*PI* 33.)

Nor will it help to identify grasping a rule with the *combination* of some interior mental item and some sequence of bodily behaviour; we cannot impose specificity on one cloud of ambiguity by tying it to another such cloud. And obviously nothing the agent *says* will help either, since whatever the agent says will just be one more bit of variously-interpretable behaviour.

Yet surely we do succeed in meaning and intending things and in following rules – despite its apparently being impossible for us to do so. Hence the paradox.

The form of skepticism with which the rule-following paradox threatens us seems still more vertiginous than the familiar Cartesian variety. The latter merely cuts us adrift from the objective world while leaving our subjectivity intact; whereas Wittgenstein's paradox invades our subjectivity, casting into obscurity not merely other people's mindedness but our own. How can there be so much as a fact of the matter concerning what I mean or intend, if nothing in either my mind or my conduct settles what it is I'm doing? In facing the rule-following paradox, we seem to lose our grip on our own self-understanding.

Yet it is, of course, no part of Wittgenstein's aim to cut the ground away from under us; he is, on the contrary, always out to remind us of the ground on which we stand and have always stood. And in the present case, the point of the rule-following paradox is not to undermine our confidence in our ability to understand ourselves and one another, but rather to liberate us from a muddled picture of what such understanding is like.

Against Self-Applying Rules

Wittgenstein invites us (*PI* 185) to imagine a case where we have asked someone to continue a sequence of numbers in accordance with the rule "add 2 each time" and she seems to be doing so. The problem with the *spoken phrase* "add 2 each time" is that it can be interpreted in a variety of ways, can express a variety of rules, all equally compatible with the person's behaviour thus far. When we initially suppose that reading her mind would clear up our worries, what we're supposing is that there's some item in her mind that *can't* be interpreted or applied in different ways, something that carries its own interpretation or application with it. But that supposition is dissolved by imagining ourselves peering telepathically into the subject's mind and seeing, say, the thought "add 2 each time" inscribed there in big shining ectoplasmic letters – whereupon the subject cheerfully proceeds to do something else (*i.e.*, something *we* would not describe as adding 2 each time). We then see that what one means by the *thought* "add 2 each time" depends on how one actually applies it in practice, no less than what one means by the spoken words does.

The upshot is not that there is something mysterious or impossible about following a rule, but rather that there *would* have to be something mysterious or impossible about it if following a rule involved what we're tempted to think it involves: a *self-interpreting* or *self-applying* rule. This is in effect what we are hoping to find when, in imagination, we peer telepathically into the subject's mind – only to find, to our dismay, merely more stuff that requires interpreting and applying. As Wittgenstein puts it, we are tempted to suppose that "the act of meaning the order had in its own way already traversed all those steps: that when you meant it your mind as it were flew ahead and took all those steps before you physically arrived at this or that one." (*PI* 188)

If we think that what makes rule-following possible *must* be the rule's somehow having its application already built into it, then careful reflection on rule-following is bound to turn vertiginous, because – with Wittgenstein's guidance – we soon recognise that there's no such self-applying rule to be found: "any interpretation still hangs in the air along with what it

interprets, and cannot give it any support." (*PI* 198) But what Wittgenstein infers from this is not that grasping a rule is impossible or mysterious, but rather 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'." (*PI* 201)

It is thus a mistake to suppose that, having failed to find the magical meaning-determining item either in the agent's thought or in her conduct, we should start to look for it *somewhere else* – say, in the agent's *behavioural dispositions*, or in the practices of the agent's *linguistic community* (to pick two examples not exactly at random). It's true enough, of course, that for Wittgenstein the agent's ability to mean and intend as she does, and to engage in rule-guided activity, depends crucially on various facts about what he would call her "natural history," including her behavioural dispositions and linguistic community. But these can no more function as *independently specifiable* determinants of the agent's meaning than her thought or conduct can.

If a behavioural disposition is thought of as a disposition to exhibit various *bodily movements* in various situations, then it too will underdetermine which rule the agent is following, since a description in terms of mere bodily movements is going to have a hard time distinguishing between, say, a) an intention to add 2 most of the time but 3 occasionally, and b) an intention to add 2 in all cases, coupled with a tendency to make mistakes in calculation.

As for the practices of the agent's linguistic community, any ambiguity in specifying which rule an individual is following is simply going to be reproduced at the collective level as an ambiguity in specifying *which practice the linguistic community is following* – since, for the same reasons, the community's noises and movements are going to be consistent with an infinity of possible practices. To follow a rule is to engage in a certain kind of social practice, to be sure; but what practice that is cannot be identified independently of a reference to following *that rule*.

Wittgenstein warns us against supposing that "because only the actors appear in the play, no other people could usefully be employed upon the stage of the theatre." (*Remarks on the Foundations of Mathematics* VII. 18) The error against which he warns us is that of confusing a process's *presuppositions* with its *content*. It's true that our grasp of mathematical propositions, for example, depends upon facts about our natural history, but that does not mean that mathematical propositions themselves are "anthropological propositions saying how we men infer and calculate." (III. 65.) In the present context: dispositions, community practices and the like are stagehands in the theatre of meaning – indispensable stagehands, to be sure (and indispensable not just practically but conceptually) – but they are not the actors.

Of course if we describe the agent's disposition, or the community's practice, as a *disposition to add 2 each time*, or a *practice of applying the phrase "adding 2 each time" to adding 2 each time*, then the underdetermination problem vanishes; but it is equally true that if we describe the agent's thought as an *intention to add 2 each time*, or simply describe her conduct as *intentionally adding 2 each time*, then the underdetermination problem vanishes

once again. *And that in a way is the answer.* But in such cases we have not identified any factor *distinct* from her rule-following that determines which rule-following it is.

The fundamental mistake that Wittgenstein is trying to disentangle us from is precisely the assumption that in order to make sense of such rule-governed activities as understanding, meaning, intention, action, and the like, we must be able to analyse them in terms of something more basic – an assumption that leads us to make a mystery out of the ordinary and then to generate further mysteries in a vain attempt to dispel the first one. The rule-following paradox exposes as a confusion the familiar philosophic distemper of seeking the coherence of human life and practice in something external to such life and practice – from the epistemologist's search for the indubitable, self-certifying foundations of knowledge to the Hobbesian's search for a force or institution that will impose cooperative order upon society without presupposing such cooperative order for its own establishment and maintenance.

In the regular course of life, Wittgenstein reminds us, we don't generally find ourselves mystified at what we (or even others) mean or intend; we entangle ourselves in mystery only when we try to dig *beneath* our ordinary experience in order to uncover foundations for what needs no such foundations.

Is Meaning Arbitrary?

Yet if what we mean is not grounded in anything beyond itself, how does it escape being arbitrary? Wittgenstein might well say it doesn't, since he often uses the term "arbitrary" precisely to mean "not grounded in anything beyond itself." But there's nothing pejorative about this sort of arbitrariness; indeed, it's the kind of arbitrariness that *logic* is seen to have once we assimilate Frege's lesson that logic is not to be grounded in psychology – while avoiding Frege's mistake, or at least the mistake Frege's language might encourage, of attempting to ground logic in a metaphysical "third realm." (Talk of a third realm is innocent enough so long as it is understood as a description of various logical features rather than as a reference to a realm of entities purportedly underlying and explaining those features.)

Psychologism and Platonism both attempt to ground logic in something distinct and more basic; but for Wittgenstein it is incoherent to seek anything deeper than logic, since this could only be something to which logic does not apply, and so something we cannot so much as speak of or do anything with. "I must *begin* with the distinction between sense and nonsense. Nothing is possible prior to that. I can't give it a foundation." (*Philosophical Grammar* I. 6. 81)

Of course rule-following *can* be arbitrary in the more ordinary, voluntaristic sense too; it depends on the details of the case. Some practices are localised and dispensable; we can take or leave their rules as we will, for any reason or for none. (Though we ordinarily cannot keep the practices while dropping the rules.) Other practices are woven more deeply into the fabric of our lives; abandoning their rules, while possible, would mean a major disruption.

Still other practices may be so bound up with rational agency itself that no avenue of abandonment (short of a bullet to the head) is intelligible to us.

The thought that the rule-following paradox *must* make all meaning arbitrary (in the voluntaristic sense) can draw support from Wittgenstein's insistence that it is "no act of insight, intuition, which makes us use the rule as we do," and that it would be "would be less confusing to call it an act of decision." (*Brown Book* II, 5) But Wittgenstein attempts to forestall such a misunderstanding by immediately adding: "though this too is misleading, for nothing like an act of decision must take place, but possibly just an act of writing or speaking."

How is rule-following like and how is it unlike an "act of decision"? We can see how it is like a decision by reflecting on the following remark:

> In all language there is a bridge between the sign and its application. No one can make this for us; we have to bridge the gap ourselves. No explanation ever saves the jump, because any further explanation will itself need a jump. (*Lectures: Cambridge 1930-32*, p. 67)

Wittgenstein's thought here is closely akin to Lewis Carroll's parable in "What the Tortoise Said to Achilles." (*Mind* 4, no. 14 (April 1895), pp. 278-280) The Tortoise grants Achilles some premises from which a certain conclusion follows, but refuses to grant the conclusion. When Achilles points out that if the premises are true, the conclusion must be so as well, the Tortoise seemingly accepts Achilles' claim, *adding it to his premise set*, but still resisting the conclusion. And each time that Achilles insists that if the most recently expanded premise set is true, so must be the conclusion, the Tortoise responds by expanding his premise set once again to incorporate Achilles' latest insistence (without drawing the conclusion).

The Tortoise is in effect demanding to be provided with a self-applying rule, one that will all by itself bridge the gap from premises to conclusion without *his* having to do anything. And of course each new insistence from Achilles, *when interpreted as one more premise*, simply "hangs in the air along with what it interprets" and brings the conclusion no nearer.

Achilles would do better to answer the Tortoise with a remark of Wittgenstein's as recorded by Rush Rhees: "I don't try to make you *believe* something you *don't* believe, but to make you *do* something you won't do." (*Discussions of Wittgenstein*, New York: Schocken 1970, p. 43) No pile of premises, no matter how towering, can substitute for the *action* of actually drawing the conclusion. Rule-following is like an act of decision because our action when we follow a rule is free from logical determination by anything external to it; we "obey the rule *blindly*." (*PI* 219)

Yet the decision comparison is also, as Wittgenstein notes, misleading. Talk of "decision" makes it sound as though we are conferring meaning on something, just as talk of "insight" and "intuition" makes it sound as though something is conferring meaning on us. *Both* metaphors bifurcate meaning into a receiving and a bestowing element – dough on the one hand, cookie-cutter on the other. But what is this meaning-bestowing decision but

another attempt at a self-applying rule? If a "decision" is needed to specify which rule I'm following in my action, what specifies which decision I'm making? Or if the decision doesn't need its meaningfulness bestowed from without, why does the action need to receive its meaningfulness from the decision?

It is also misleading, then, to describe Wittgenstein as teaching that there is "no fact of the matter" as to what we mean or what rule we are following; the slide from "no *independent* fact" to "no fact" is unwarranted. Ontologically, what "makes it true" that I am following rule A rather than rule B is simply *my following rule A*. Epistemically, the feature of my conduct that others pick up on in order to detect that I am following rule A rather than rule B is, once again, simply my following rule A.

Of course their ability to see the rule in my actions – what Wittgenstein would call their "sane human understanding" (*Philosophical Remarks* 18) – depends on their sharing the right sort of natural history with me. But there is no neutral, regress-proof vocabulary in which to specify uniquely what that shared natural history is without invoking the very sorts of meaning-facts that the natural history was supposed to explain.

In the Beginning Was the Deed

I began by describing the moral of the rule-following paradox as follows: what rule one is following when one acts is (radically) underdetermined by anything in either one's actions or one's thoughts. But I also began with a warning that this description was not quite accurate.

We can now see where the inaccuracy lies. *Of course* there is something in one's thoughts that settles which rule one is following: namely, the intention to follow that rule. And likewise, *of course* there is something in one's actions that settles which rule one is following: namely, one's intentionally following that rule. But one can earn the right to these commonsensical banalities only by ceasing to think of thought and action as independently specifiable; and to win our way to that insight we need to work our way through the rule-following paradox.

A living, conscious being is neither a ghostless machine nor a machineless ghost; but neither cannot it be understood as a mere *gluing-together* of these two nonliving items, ghost and machine. Aristotle defines soul as the form of the organic body and organic body as body informed by the soul – neither specifiable independently of the other. (*De Anima* II. i. 412a19-b26) In similar spirit, Wittgenstein affirms that "the human body is the best picture of the human soul" (*PI* II, p. 178), while nevertheless denying that "the soul itself is merely something about the body" (*Remarks on the Philosophy of Psychology* II. 690; quoting Nietzsche, *Zarathustra* I.4). A living being is an integrated, organic whole – in Aristotelean terms, a *hylomorphic unity* – of which soul and body are distinguishable but inseparable aspects (not ingredients).

By extension, we cannot arrive at the notion of action by gluing together one ghostly item – a mental image with no behavioural import – and one machinelike item – mere bodily movement with no psychological import. Action too is an indivisible whole, of which thoughts and movements are aspects but not separable ingredients. "Thinking is not an incorporeal process which lends life and sense to speaking, and which it would be possible to detach from speaking rather as the Devil took the shadow of Schlemiehl from the ground." (*PI* 339) Action, in short, is more than the sum of its parts; the very identity of my thoughts depends on how I express them in action – but which action I am performing depends on what thought I am expressing in it. The difference between thoughtful action and mere bodily movement is thus not a *distinct ingredient*; the mistaken search for such an ingredient is in fact yet another quest for a self-applying rule.

The rule by which we play a musical score, Wittgenstein tells us, "is not contained in the result of playing, nor in the result plus the score (for the score might fit *any* playing by *some* rule)," but "only in the *intention* to play the score." (*Lectures: Cambridge 1930-32*, p. 40) Yet this claim that the rule is "contained … in the intention" might seem to be contradicted by Wittgenstein's later remark that "The rules are not something contained in the idea and got by analyzing it. They constitute it." (*Lectures: Cambridge 1932-35*, p. 86) But we should not assume that the same fixed sense of "contained" is in play in both passages; after all, an expression has meaning only in a proposition (*TLP* 3.314), *i.e.*, only in the context of significant use (3.326).

In the later passage, the sense in which the rule might be thought of as "contained" in the intention is being contrasted with Wittgenstein's preferred formulation that the intention is "constituted" by the rules. The contrast between "contained" and "constituted" suggests that "contained" as used here bears the implication that the intention is *distinct from* and *more than* the rules it carries within it – an implication that Wittgenstein is concerned to reject. In short, "contained" marks the ghostly while "constituted" marks the organic interpretation.

In the earlier passage, however, the rule's being contained in the intention is being contrasted with its being contained in something narrower – "the result of playing," or "the result plus the score"; here it is "contained" that marks the organic, this time in contrast with the machinelike. And so the inconsistency vanishes; both passages endorse the organic unity of action against mentalistic or behaviouristic alternatives.

The Redemption of Metaphysics

The claim that human agents, and human actions, are indissoluble, irreducible hylomorphic unities sounds suspiciously like a metaphysical thesis. Hasn't Wittgenstein set his face against such theses? If so, how can such a thesis be one of the upshots of his rule-following paradox?

As with "contained," so with "metaphysics," the meaning of the term depends on the context of significant use. Stanley Cavell has characterised Wittgenstein as seeking to de-psychologise psychology. ("Aesthetic Problems of Modern Philosphy," p. 91; in *Must We*

Mean What We Say? pp. 73-96) In like spirit we may say that Wittgenstein's project also seeks to de-metaphysicise metaphysics.

Part of what Wittgenstein customarily calls metaphysics, but which we may perhaps call *metaphysicism*, is the error of treating essentially *logical* or *grammatical* principles as though they were *descriptions* – contingent in form though not in intent – of some extramental reality. (For example, the tendency, frequently discussed by Wittgenstein, to treat logical constraints as though they were like physical constraints, only super-rigid.) As such, metaphysicism is the twin of psychologism, the error of treating logical or grammatical principles as descriptions of some *psychological* reality (*e.g.*, explaining the laws of inference in terms of psychological association). Indeed metaphysicism and psychologism are perilously entangled, for each tempts us to accept it as the remedy for the other.

Plato's Forms might be seen as an example of metaphysicism. On such a reading, Plato saw, rightly, that logical concepts are not reducible to anything physical or psychological or empirical, and his exaltation of the Forms is thus his attempt to convey the irreducibility of logic; but in describing the irreducibility of logic in terms of a realm of irreducible entities, he slid into treating logic as grounded in, and reducible to, the natures of these entities, and so lost his hold on the very position he was trying to defend. Another example might be Duns Scotus's theory of individuation. Scotus may be interpreted as wishing to claim that all objects are irreducibly particular and so do not need to derive their particularity from some added ingredient such as prime matter; but rather than expressing this idea by saying (as later Scholastics would) that objects do not have or need a principle of individuation, or even that each object is its own principle of individuation, he arguably slid into reifying the object's irreducible particularity, treating it as a special metaphysical ingredient – "thisness," *haecceitas* – and in effect treated the object's particularity as reducible after all, *i.e.* to the particularity of its *haecceitas*, thus depriving the object of its genuine *haecceitas* precisely by giving it a pseudo-*haecceitas* conceived in the manner of metaphysicism.

Just as much that passes under the name of psychology is in Wittgenstein's eyes mere psychologism, so much that passes under the name of metaphysics is doubtless mere metaphysicism. Still, Wittgenstein is trying to rescue and clarify our psychological concepts, not eliminate them; and the same is arguably true for metaphysics as well. When Wittgenstein remarks that "grammar tells us what kind of object anything is" (*PI* 373), and characterises his grammatical investigations as exploring "the 'possibilities' of phenomena" (*PI* 90), he is essentially *assigning to grammar the traditional task of metaphysics*. In doing this he is not so much *rejecting* metaphysics (as he would be if he were to insist that *nothing* can perform the traditional task of metaphysics) as he is *logicising* metaphysics. (Of course this involves reconceiving – logicising – the task as well.)

The irreducibility of logic may be called a metaphysical thesis, so long as such terminology is not misunderstood; metaphysical it may be, but it is precisely the antidote to most of the errors that have been termed metaphysical. Indeed, one function of the rule-following paradox is arguably to help us distinguish between nonsensical metaphysicism and sensible (because logicised) metaphysics. Wittgenstein writes:

> "But I don't mean that what I do now (in grasping a sense) determines the future use *causally* and as a matter of experience, but that in a *queer* way, the use itself is in some sense present." – But of course it is, 'in *some* sense'! Really the only thing wrong with what you say is the expression "in a queer way". … In our failure to understand the use of a word we take it as the expression of a queer *process*. (*PI* I. 195-6)

To Wittgenstein's mind there's nothing *inherently* wrong with a metaphysical-sounding statement like "When we grasp a sense, the future use is already present." It depends how we take it. When we take it "in a queer way," we are thinking of the presence of the future use as a fact that is *independent* of the ordinary business of rule-following, a transcendental process in which that business is grounded. This is to fall into metaphysicism. But if we take the presence of the future use as an illuminating description of rule-following itself, rather than of something else that serves as rule-following's ground, we are innocent of metaphysicism; we have recovered the proper grammar for such locutions. "The rules are not something contained in the idea and got by analyzing it. They constitute it." That means: the presence of the future use is not something in which grasping a rule is grounded; it is simply the grasping redescribed.

The idea of action as irreducible and basic, not decomposable into ghostly thought and machinelike movement, may be further explicated by means of the distinction Wittgenstein draws in the *Tractatus* between signs and symbols, where a sign is a mere mark or sound while a symbol is that same mark or sound employed with a particular meaning – so that "bank," meaning the edge of a river, and "bank," meaning a financial institution, would be the same sign but different symbols. For Wittgenstein, a symbol is *neither* a mere sign, *nor* a sign plus some ghostly accompaniment (like Schlemiehl's shadow); it is the sign *in significant use*. (*TLP* 3.326) Nor, on pain of a rule-following regress, can significant use itself be analysed either in sign-talk or in ghost-talk. What the symbol adds to the sign cannot be specified independently; hence the symbol is not built up from the sign *plus* something further. Rather, the symbol is basic, and the sign is a sort of abstraction from it, the "perceptible aspect of the symbol." (*TLP* 3.32). There is no getting *behind* or *beneath* the symbolic level. Because logic is basic, symbolising is basic; because symbolising is basic, the organic unity of action is likewise basic.