5. In passages like §437, Wittgenstein discusses the way intentional concepts like that of wish and thought are affected by versions of the problems about making room for the notion of accord that make for difficulties in our thinking about ‘grasping a rule’.
8. My “Wittgenstein on Following a Rule” is too hospitable to this kind of reading, I now think.
9. I have changed the punctuation in the second paragraph, in a way that brings out what I take to be the dialectical flow of the passage.

**The Dispositionalist Solution to Wittgenstein’s Problem about Understanding a Rule: Answering Kripke’s Objections**

CARL GINET

Wittgenstein’s problem about understanding a rule relates to a central preoccupation of his throughout his philosophical life, the question of how it is possible for there to be such a thing as linguistic meaning. How are we able, with our various sorts of signs, to represent or refer to facts or things in the world, and even to nonexistent possibilities? I think Wittgenstein always felt a deep mystery here and in his later work this feeling took the shape of a puzzle that questions the very possibility of meaning. Kripke characterizes this puzzle as “a new form of philosophical skepticism... the most radical and original that philosophy has seen to date...”

What is the puzzle? Let us take it that understanding the meaning of a descriptive term or predicate can be thought of as grasping a rule for a certain function whose domain contains all the cases about which it can be asked whether the predicate applies to them and which assigns to each such case one of two values: applies or does not apply. If one knows what the predicate, ‘x is red’ means, then one knows a rule for such a function; one knows what determines for any particular member of the domain of visible objects whether or not it satisfies the predicate. To understand the predicate ‘x + y = z’ is to grasp a function that assigns a value to every one of the infinitely many different pairs of integers.

For each of us there are enormously many such functions for which we think that we have come to know a general rule that determines the value of the function for every argument to it; and we think we grasp such a thing even though we have not yet explicitly considered all the arguments to the function, even when it would be impossible in principle for us to consider them all. When we do explicitly consider an argument that we have not hitherto considered, we confidently judge that the rule we have all along had in mind dictates such-and-such a value for that argument.
Now comes the problem: there are many incompatible rules that agree on all the cases we have hitherto explicitly considered but disagree on this new case. What exactly was it in our mind that determined that the rule we understood was a particular one of these incompatible rules rather than another? To take Kripke’s favorite example (7–9): suppose that I have never previously added any numbers larger than 56. When I compute 57 + 68 and confidently get the answer 125, I suppose that this result conforms to the function I have all along meant by the plus sign. But what in my mind could have determined that it was the ordinary plus function that I meant rather than, say, the “quus” function—which is defined this way: quus equals x plus y if x and y are less than 57, and otherwise it equals 5—if I never before considered any numbers as large as 57?

Take a different sort of example. The first time I confront a red patch that is surrounded by an area that is turquoise in color, I confidently judge it to be red and I suppose that my doing so accords with the meaning of ‘red’ I have long understood. But what that was previously in my mind determined that the rule I intended to follow in applying ‘red’ was the ordinary one rather than the following one: if a surface area looks like this [pointing to a red surface] and is not surrounded by turquoise then call it ‘red’, but if it does not look like that or it is surrounded by turquoise then say ‘not red’? The more we ponder such questions, the more it can seem that no sort of fact about me at the earlier time could have determined which rule I intended.

But aren’t we perfectly familiar with the sort of fact in question? The understanding of a general rule for an “application” function is the sort of thing we often acquire “in a flash.” When, watching “Sesame Street,” I suddenly see which thing does not “belong,” I instantaneously become confident that I could tell for any new thing whether or not it “belongs.” I realize, of course, that the general rule I have suddenly grasped is not uniquely identified by the description “a rule that these cases I have so far considered conform to.” But, I want to say, I do somehow have in mind something that determines the value of the function for all the cases I have not yet considered, for I do understand a general rule that does determine the whole function.

But, says our puzzler, consider. How could there be something in your mind (or anywhere) which, on the one hand, does not involve explicitly exhibiting a particular argument and its value, but, on the other hand, does fix it that that pair does belong to the function? How could something that is not there in your mind be thus fixed or determined by something that is there? Isn’t that a remarkable feat? What sort of thing could pull this off? Any ordinary thing we can think of to suggest—a formula that states the rule, a picture of what all the same values have in common—is just an expression of the rule that also needs to be understood and thus seems merely to put off answering the hard question: understanding that expression of the rule involves with respect to the symbols involved in it the very same sort of grasping of general rules for their application that we are asking about.

Wittgenstein uses the term “cube” as an example (Philosophical Investigations 139–41). Suppose we try saying that to understand the general rule that determines whether a given visible object is a cube or not is to have in mind a picture that shows what anything that deserves the positive judgment must look like. This, too, is just to suggest another way of expressing the rule: “A thing is a cube when it looks like this, otherwise it isn’t.” Wittgenstein asks: what tells me how to apply this to a new case? There are alternative ways in which it is possible to apply it. And one need not have in mind here that the picture may include features (such as color) that are irrelevant to a thing’s being a cube. Even if the irrelevant features are ignored, or one’s mental picture is indeterminate with respect to them (if that is possible)—even if, per impossible, one could contemplate an object that is a cube and that has no irrelevant features, the perfect abstract template—one could, Wittgenstein points out, use a method of comparison different from what we normally think of as determining what counts as looking like a thing. (The beauty of Wittgenstein’s move here is that it shows that even if Platonic Forms or Lockeian abstract ideas were metaphysically possible, they would not do the job that seems to be required.)

There are further suggestions we could try, but it becomes clear that no explanation of how to decide whether a new case comes under the rule is going to be able to avoid this sort of reply from the puzzler. As Wittgenstein puts it (Remarks on the Foundations of Mathematics 113), “However many rules you give me—I give a rule which justifies my employment of your rules”—that is, which justifies an application in a particular case that is different from the one you would make. Concerning any categorical fact we seem able to suggest for the sort of fact that constitutes the grasping of a general rule, the puzzler points out that exactly the same kind of thing could have been in the mind of one who thinks she grasps a general rule but judges a new case differently from the way we would. Thus no such fact can be metaphysically sufficient for having in mind something that fixes the values for hitherto unconceived cases in one way rather than another. There is no sort of fact that can constitute understanding a general rule as we need to think of it, as something that somehow “contains” the values for all the cases it does not explicitly exhibit. “This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule” (Philosophical Investigations 201).

In explaining how this conclusion is reached, we have focused on examples of functions that we can identify only by general rules and not by enumeration. But this is not essential. The same problem can be generated for small finite functions that we can give by complete enumeration and that we must initially learn from a complete enumeration. The alphabetical ordering of the letters of our alphabet can be thought of as a function of 25 ordered pairs: <a', b'>, <b', c'>, etc. We know the alphabetical ordering when we know this function and we learn all the ordered pairs in it by enumeration. Typically, when I speak of the alphabetical order of the letters and mean this function, I do not have the whole thing exhibited in my mind. And yet on those
occasions I mean that very function and not some other. How do I do it? How can anything short of exhibiting the whole alphabet to myself constitute my then meaning that order rather than some other?

The crux of the problem is that one does not have all of a number of things *explicitly* in mind at a time when one does, as one wants to say, have them all in mind *implicitly*—by having in mind a rule or just the thought, "Yes, I know what all that is." How can anything be in my mind only implicitly? How can anything short of actually having a particular thing present to my mind constitute having that particular thing "in mind," meaning it, rather than something else instead? Meaning seems typically to entail my being somehow in contact with things that I do not then have explicitly in mind. The problem is that, when we ask what sort of current fact about me could put me in contact with those absent things, no sort of fact seems capable of doing the job.

That is the problem. What is the right response? Kripke (66–67) divides the possible ways of trying to deal with the problem into two kinds: straight solutions and skeptical solutions. A straight solution tries to make intelligible some kind of fact or other that, it alleges, is capable of doing the job that needs to be done. A skeptical solution concedes that no kind of fact can meet the demand and says that we must therefore reinterpret all of our talk that seems to imply the mysterious containment of absent cases in such a way that it no longer has that implication. Kripke attributes to Wittgenstein a skeptical solution, but he also considers and criticizes several straight solutions, usually claiming to be expounding Wittgenstein’s criticisms of them. The solution I find most appealing is the straight solution that Kripke spends the most space criticizing (22–37). He calls it *dispositionalism.*

The dispositionalist’s idea is that the sort of fact that can do the job we want done is a *counterfactual conditional* fact. Roughly, the idea is that what I meant or intended at a particular time dictated a certain result for a case that was not then before my mind if it was true of me then that I would have reacted in an appropriate way to the case if I *had* then considered it: I had at the time a disposition to react appropriately in the relevant circumstance. My meaning at a given time a certain function by a sign I used then was my having then a multiplicity of dispositions to assign the various values of the function to its various arguments, each disposition in the set determining one of the ordered pairs in the function. Any later claim I make that a particular ordered pair belongs to the function I then meant is correct or not according to whether or not the corresponding disposition is a member of that set of dispositions I then had. It is this counterfactual conditional sort of fact that provides an intelligible way of clarifying the sense in which my present state of meaning or intending can "contain" something, even a great many things, that are not presently before my mind.

At *Philosophical Investigations* 195, Wittgenstein puts in quotation marks what he is tempted to say—"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"—and responds:

"But of course it is, 'in some sense'! Really the only thing wrong with what you say is the expression 'in a queer way'." A bit earlier, at 187, Wittgenstein remarked:

- When you said "I already knew at the time..." that meant something like: "If I had then been asked what number should be written after 1000, I should have replied '1002.'"

Perhaps dispositions provide a non-queer way in which all the uses (applications) are present at one time. The dispositionalist says that, if the reach to the absent cases strikes us as something very "queer" and "mysterious," that is because we are looking for what constitutes it in the wrong "dimension." We are thinking of it as a *categorical* fact rather than as one expressible only in the subjunctive, a counterfactual conditional fact.

Let me now be more specific about the content of these counterfactual conditionals. According to the dispositionalist, the fact that I meant a certain application by a certain term I used at a certain time is (for a basic class of cases: more on this qualification later) the fact that there was true of me a counterfactual conditional for each ordered pair in the application function that I meant, that is, for each truth of the form "the application function I meant assigns this value to that argument". As a first shot, it is natural to think of this conditional as having something like the following form:

If I had been asked whether the rule (application function) I meant (understood) assigns this value (applies or does not apply) to that argument,
I would have answered that it does.

But this is too crude. It makes the presence of the disposition depend on whether or not I would hear and understand the question asked, and whether or not I would be able or motivated to respond, and should not depend on such things. For if it did, then it would not be plausible for the dispositionalist to claim (as I think the dispositionalist must claim: more on this point later) that I can later know that the disposition was there only by remembering it, in the same direct way that I can know what a past intention of mine was. So let us refine the form of the counterfactual conditional to read as follows:

If I had considered the question whether the rule I meant assigns this value to that argument and given a relevant and sincere response to it, my response would have been that it does.

The antecedent and consequent here do imply that I understand some meanings and hence some rules of application. But this is not objectionable, for the dispositionalist’s purpose of giving a non-mysterious fact that contains decisions on the absent cases. What the puzzler wants contained and claims cannot be contained are *judgments* about the absent cases. To explain satisfactorily how these can be contained we do not need to reduce them to something of a non-intentional nature.

My object in the remainder of this essay is to defend the dispositionalist idea against Kripke’s objections to it. I divide these into three.
1. The objection on which Kripke spends the most space (22–25, 37) is that the dispositionalist account leaves out the normative aspect of a rule. A collection of dispositions gives us a function only as a mere collection of ordered pairs. But the functions meant by the plus symbol and by descriptive words generally are not arbitrary collections, but are bound together by a rule. A set of dispositions determines the function only extensionally, whereas there is something more when there is a rule we grasp. There is a reason why the function includes just the pairs it does. Kripke suggests (23) that the mere truth of a counterfactual conditional about me cannot make the response it says I would give "justified in terms of instructions I gave myself, rather than a mere jack-in-the-box unjustified and arbitrary response."  

This normative aspect of a rule is certainly something that dispositionalism must get into the picture if it is to provide an adequate solution to the puzzle; and nothing I have said so far brings it in.  

The challenge to the dispositionalist is this: You have given us a disposition, a counterfactual conditional, for each ordered pair in the function; but you have not given us anything that makes this collection of dispositions nonarbitrary, determined by a general rule, as must be the case for the application functions determined by the meaning of "plus" and by the meanings of most of our descriptive terms.

It seems to me that the first thing the dispositionalist should say in order to provide the account needed here is this: a subject's meaning a function involves understanding it to be determined by a rule if (but not only if) the subject is disposed to recognize some non-trivial expression of the rule as correctly stating the rule determining the function the subject means. Such a disposition is surely always present, for instance, in the case of meaning addition by "+"; one who knows what addition is disposed to recognize as correct some description of some procedure for finding the sum of any two numbers. There is, for instance, the sort of procedure we all learned in school when we got beyond single digit numbers, that of "roughly, writing one numeral under the other (preserving the order of digits and aligning on the right), adding zeros on the left of the shorter number to make both numbers the same length (when necessary) and then adding corresponding digits by referring to the table, carrying 1 and entering only the right digit when the table gives a two-digit answer."  

Of course, this raises the question of what the subject's understanding of the expression of the rule consists in and how that determines (contains) all the values of the function. And one must admit that, if the subject's understanding of the various terms in the expression of the rule is unpacked in terms of her having further dispositions to recognize expressions of rules that determine their application functions (e.g., a rule she knows for addition involves reference to counting and she knows a rule that determines when that term is correctly applied), then the same question arises, until we get to terms such that the subject's grasp of a rule that determines their applications does not require explanation in that way (in terms of a disposition to accept some informative expression of a rule for how to apply the term) but can be explained in some other way.

We must get down to what might be called a brute level of rule grasping where, as Wittgenstein puts it, justification comes to an end—that is, there is no more justification in terms of further expressions of rules.

There do seem to be terms such that, although the application function is not arbitrary, one would be at a loss to describe any general procedure for finding the application value given the argument, or at any rate it would not seem to be crucial to one's understanding the term, to one's knowing how to apply it correctly, that there be something one would recognize as an informative expression of the rule one follows. Cases in point would, I think, be "x is red," "x is hot," "x is sweet," "x is shrill (in sound)," "x contains an edge" (applied to visual fields), "x is the same in color as y," "x has the shape of the numeral 3," "x feels warmer than y," "x sounds higher in pitch than y"—terms for directly sensed properties and relations.

The application of the term is nonarbitrary—indeed, it is rule-governed in the sense that it is always the same thing that makes it correct to apply it—but one's grasp of what makes all the cases the same in the respect one means by the term need not be expressible in other terms in any informative way. (Of course, one can always use or invent a synonym for the term in question or say, "They are all the same in the respect I mean"). In such cases one's grasp of a rule for applying a term does not (at any rate need not) consist in one's being disposed to recognize some informative definition of the rule, but just in one's being disposed to apply the term correctly to particular cases. As Wittgenstein says in Philosophical Investigations 201, "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." This way of grasping a rule, which is exhibited in appropriate judgments about particular cases, could be just one's possessing dispositions to make such judgments.

But, one still wants to ask, what is it in such cases that makes all the applications one is disposed to make correct, in accord with a single rule? Taking a clue from some of Wittgenstein's remarks, one might try to get nonarbitrariness into a collection of dispositions by putting into the consequent of the conditional something about the subject's attitude toward her response, producing the following:

If I had considered the question whether or not the rule I meant assigns this value to that argument and given a relevant and sincere response to it, then I would have responded that it does and I would have regarded this response, not as a new decision made for this case, but as just doing the same thing I have done (or would do) for other arguments where I assign the same value.

Here the consequent of the conditional specifies a two-part response: a judgment about the case and an attitude toward that judgment, the attitude that this judgment is not an arbitrary decision for just this particular case.

But, of course, that attitude toward the judgment implies something about the content of the judgment, namely, that what the subject is judging is that
the term applies (or does not apply) for the same reason it does (does not) in other cases and that this reason has to do with the nature of the case. So we might make this explicit in the consequent of the conditional, producing the following:

If I had considered the question whether the rule I meant assigns this value to that argument and given a relevant and sincere response to it, then I would have responded that it does and I would have regarded this response, not as a new decision made for this case, but rather as a judgment that there is present in this case that makes it right to assign this value (or there is absent in it that whose absence makes it right to assign this value) and it is the same thing that is present (absent) in any other case to which it is right to assign this value.

Note that the proposition that the subject makes (or would make) a judgment naturally expressible in the terms suggested here does not imply that the subject does grasp any rule that determines positive and negative cases, that makes the collections nonarbitrary. For it is possible for the subject to be disposed to make judgments with this content that are all false, to be under an illusion that she goes by a rule in judging whether or not the term applies, to think she grasps a general rule (and to have the dispositions to back up this thought) when she in fact does not. This means that the dispositionalist has not completed the required task for these brute level cases. We still need a noncircular account of what distinguishes really grasping a general rule in these brute level cases from only thinking one does.

Before pursuing that question, let me introduce one more refinement of the content of the counterfactual conditional that I think the dispositionalist should make. The antecedent should include something about the subject’s being adequately informed about the particulars of the case being judged. We want to rule out the possibility of the subject’s giving the wrong judgment or “don’t know” because of her having a false impression of, or being ignorant about, crucially relevant elements in the case. So the standard-form counterfactual should, finally, be something like this:

If I had considered the question whether or not the rule I meant assigns this value to that argument, been fully aware of all relevant particulars of the case, and given a relevant and sincere response to the question, then I would have judged that there is present in this case that which makes it right to assign this value (or there is absent in it that whose absence makes it right to assign this value) and it is the same thing as is present (absent) in any other case to which it is right to assign this value.9

Now, our question is: What distinguishes pseudo-rule-following from the real thing? The real thing is where the rule-follower’s dispositions are not an arbitrary collection but are all correct because they are in accord with a single rule and the subject has the dispositions because she grasps that rule. What distinguishes this from a case where a subject is disposed to react as a genuine rule-follower would (that is, a large set of counterfactual conditionals of the sort most recently specified are true of her) but there is in fact no rule which makes all those dispositions correct, with which they all accord, and so the subject has only a delusion of following a rule?

Two rather different sorts of possible answers come to mind. (1) The community agreement answer: the dispositions constitute genuine following of a rule if and only if other subjects in a suitably defined community would, after suitably defined training, have the same dispositions. (2) The objective property answer: there is a property that is present in all the cases on which the subject is disposed to make the positive judgment, and absent in all those on which she is disposed to make the negative judgment, and the subject is disposed when suitably placed to recognize the presence or absence of that property.

The first answer is nominalist, subjectivist, anti-realist with respect to the properties or universals our discerning of which many have thought (and it is natural to think) is the basis for our application of descriptive terms. The second answer is realist, objectivist about properties or universals. Can dispositionalism choose the second sort of answer? I do not see why not. As far as I can see, dispositionalism does not need to be nominalist about universals in order to achieve its aim. What dispositionalism aims at is to provide a non-mysterious fact about the subject that has the property of “reaching to” or “containing” all the absent cases covered by the rule the subject grasps. This is the feature of understanding a rule that puzzles us (and the feature that the puzzler says is impossible). What dispositionalism offers is a set of counterfactual conditional facts, each dealing with a different absent case. I cannot see any incompatibility between, on the one hand, offering that to question that and, on the other hand, taking seriously and literally the idea that what the grasped rule dictates is that the term applies to just those cases that have a certain objective property in common.

To say that a property is objective is to say that it is (or can be) possessed by things independently of any subject’s dispositions to classify particular cases. The property could still be there in things even if there were no subjects capable of classifying things or there were none who noticed or discerned this property. Even if no entities with minds ever existed there nevertheless might exist various shapes in common to many objects. Even if no one ever discerned or thought of it, there could exist in things the property of being negatively charged.

On an objective, realist view of properties, the problem dispositionalism addresses (and proposes to solve) could be put this way: How can a present fact about the subject collect just the actual and possible instances of this property rather than those of some other property, since all that the subject can have before her mind at one time, in the way of positive and negative instances, is a small number of cases which are also positive and negative instances, respectively, of a great many other properties? What constitutes the subject’s picking out the right property, her grasping the particular respect of resemblance, that the term stands for? It seems that no occurrent phenomenon
1. Watch a slow-moving fire, in high excitement follow the progress of the burning and its approach to the explosive. Engagement, I don't think we're expecting anything at all or a multitude of disconnected thoughts.

2. For several moments before the explosion occurs I have been expecting to see an explosion very soon. If I see an explosion I regard it as the way an explosion at last second had been killed instantly by a bolt of lightning a couple of seconds before the explosion, then I have to make a decision whether or not I ever actually reached in that case that began been. It's not bad to be that, it's in my mind any sort of representation of one, not more. Is it possible to suggest that what I did have in mind sort of representation of one, not more. Is it possible to suggest what I was expecting an explosion without having those words, just quoted an explosion. As I think, this against an expectation of an explosion in a way that I think against an expectation of an explosion, it is the subjective conditional fact that if an explosion occurred just then I would have had a reaction, but it is not expressible by this. This is the sort of thing I would have been.
I've been expecting. When the explosion actually occurs I know with direct certainty that this is the sort of thing I have been expecting. But to know this, it seems to me, is to know that I would have had the same reaction if an explosion had occurred at some earlier point during the period I have been expecting it.

(2) I am working at home and I think, "I need to take this book to the office with me today." Later when I get to the office, I do not have the book: I forgot to stick it in my briefcase. But I am sure that, had I thought of it when I was packing my briefcase, or when I was leaving the house, I would have stuck it in. This is a fact about my past of which I can be as certain as I am of the fact that I did form the intention earlier to bring the book. I am no more forced back to mere hypothesizing in the one case than in the other.

In such cases, there is a marked and profound contrast between the way I am justified in believing the counterfactual conditional about myself and the way I can be justified in believing a similar conditional about someone else. Forming a belief about what another person would do, or would have done, is to form a hypothesis (based perhaps on my knowledge of their past behavior, or perhaps on the assumption that the other person is very much like me) in a way that my belief about what I would do, or would have done, is not a hypothesis. The contrast here strikes me as quite as great as the contrast between the way I can know about my own past conscious states and processes (sensations and the like) and the way I can know about those of others.

I conjecture that behind Kripke’s assumption that one’s beliefs about one’s past dispositions must be hypotheses there lies something like the following reasoning:

(a) It is a necessary condition of the truth of the counterfactual conditional about me that there was some categorical basis for it. (That is, there were independent intrinsic properties of me that would causally explain why the condition or occurrence described in the antecedent would produce the occurrence described in the consequent, that would be sufficient for the truth of the counterfactual conditional.)

(b) If (a), then I could not know that the counterfactual conditional was true of me without knowing that there was some categorical basis for it.

(c) I could know that there was a categorical basis only as an hypothesis (that is, something confirmed for me by evidence as to what specific sorts of mechanisms there are in me or by evidence as to how I and others like me have reacted in similar situations in the past).

(d) Therefore, I could know that the conditional was true of me only as an evidence-based hypothesis (and not in the sort of direct way I have memory knowledge of my recent past experience or have knowledge of my current intentions).

Of the three premises here, (c) alone seems quite acceptable: at any rate, I shall not question it. (b) seems doubtful. It is not the case in general that one must know everything entailed by what one knows. For many instances where p entails q, one may know p without knowing q if one fails to know that p entails q. Is this such a case? I find it difficult to address this question, since I doubt that in this case p does entail q.

That is, I doubt (a), that the existence of a categorical basis is a truth-condition for every sort of counterfactual conditional. In particular, I doubt that it is a truth-condition for the sort of conditional that the dispositionalist account of meaning needs or for the sort illustrated in my recent examples. And it appears that Wittgenstein eventually had doubts about this, too. In notes written in 1937 he says:

Think of two different kinds of plant, A and B, both of which yield seeds; the seeds of both kinds look exactly the same and even after the most careful investigation we can find no difference between them. But the seeds of an A-plant always produce more A-plants, the seeds of a B-plant, more B-plants. In this situation we can predict what sort of plant will grow out of such a seed only if we know which plant it has come from. Are we to be satisfied with this; or should we say: “There must be a difference in the seeds themselves, otherwise they couldn’t produce different plants; their previous histories on their own can’t cause their further development unless their histories have left traces in the seeds themselves.”

But now what if we don’t discover any difference between the seeds? And the fact is: It wasn’t from the peculiarities of either seed that we made the prediction but from its previous history. If I say: the history can’t be the cause of the development, then this doesn’t mean that I can’t predict the development from the previous history, since that’s what I do. It means rather that we don’t call that a ‘causal connection’, that this isn’t a case of predicting the effect from the cause.

And to protest: “There must be a difference in the seeds, even if we don’t discover it”, doesn’t alter the facts, it only shows what a powerful urge we have to see everything in terms of cause and effect.

In Zettel 608 this example of the seeds is introduced with

No supposition seems to me more natural than that there is no process in the brain correlated with associating or with thinking; so that it would be impossible to read off thought-processes from brain-processes. I mean this: if I talk or write there is, I assume, a system of impulses going out from my brain and correlated with my spoken or written thoughts. But why should the system continue further in the direction of the center? Why should this order not proceed, so to speak, out of chaos?

In Zettel 609 he says

It is thus perfectly possible that certain psychological phenomena cannot
be investigated physiologically, because physiologically nothing corresponds to them.

And in **Zettel** 610 he comments, "If this upsets our concept of causality then it is high time it was upset." Commenting on Wittgenstein’s example of the seeds, Budd (1984, 311) supposes that an A-plant seed and a B-plant seed fail to germinate because of lack of water and says:

All the same, we can truly say of the one seed that if it had been watered it would have become a plant of kind A, and of the other that if it had been watered it would have become a plant of kind B. And we can truly say this even though there was no intrinsic, nonrelational difference between the physical states of the two seeds and they were in identical circumstances...Wittgenstein sees no difficulty in the supposition...that counterfactual differences need not be based in intrinsic, nonrelational differences...

If Wittgenstein did indeed think that counterfactual conditionals could be true without a categorical basis, and if he thought that ascriptions of understanding and meaning could be cashed in terms of such counterfactual conditionals, then Wittgenstein thought that, as Budd puts it (313), “the condition of someone who suddenly understands how to continue a series [or means or intends or expects a certain thing] need not be distinguished from that of someone who does not understand [or mean or intend or expect]...either by (i) something that was present to or went through his mind but which was not present to or did not go through the mind of the other person...or by (ii) the condition of his mental apparatus.” I am inclined to agree with Wittgenstein’s remarks, as Budd interprets them, and therefore I am inclined to think, with respect to the argument I attributed to Kripke, that there is no compelling basis for premise (a) or, therefore, for the conclusion (d).

Is there an argument against the claim that the truth of a counterfactual conditional always entails a categorical basis? It is notoriously difficult to argue for the denial of an entailment, but one can make the following points.

What argument is there for the entailment? The entailment is not self-evident and until we are given a convincing argument for it we should hesitate to believe in it.

If a categorical basis were entailed, this should be reflected in our intuitions. Concerning Wittgenstein’s hypothetical example of the seeds, for instance, we might (if the entailment held) be expected to feel intuitively compelled to deny that it is even in principle possible that (i) it is true of an A-plant seed but not of a B-plant seed that, if it had been watered, it would have grown into an A-plant, and (ii) there is no intrinsic difference between the seeds that would causally explain why the counterfactual conditional holds for the one but not for the other. But I find myself quite willing to allow that, however much I might initially resist believing both of these things to be actually true, it is nonetheless possible that they should be: if I were confronted

with all the facts of Wittgenstein’s example, it would seem to me reasonable, not incoherent, to begin to suspect that both things are true.

Consider another example, a variant of one Wittgenstein gives in *Philosophical Investigations* 666. Suppose there are two people in the situation of simultaneously having a toothache and hearing a piano being tuned and both say, “It will stop soon.” A short time later one says that if she had been asked what she meant by “It” in that remark, she would have said that she meant the toothache, and the other says that if she had been asked she would have said that she meant the piano tuning. If I know that both speakers speak sincerely then I have a basis for confidently accepting both of those counterfactual conditionals that would not be undermined by my finding that there was no relevant, independently ascertainable intrinsic difference between the two people at the time the conditionals are about, nor any relevant difference in their histories prior to that time, that would explain why the one conditional holds for the one person and the other holds for the other person. Though there may have been such differences, why must there have been? My intuitions about such cases do not reflect any sense that the counterfactual conditional simply cannot be true unless there existed independent conditions that would explain why, if the antecedent had been true, then the consequent would have been true.16

If this does not demonstrate that a counterfactual conditional does not entail the existence of a categorical basis, it at least shows that one can fail to know that there is such an entailment. And that is enough to open up the possibility of knowing a counterfactual conditional to be true without knowing that there is a categorical basis for it. So I think the dispositionalist has a viable response to Kripke’s objection that if meaning intentions were a matter of dispositions then one could not know one’s own meaning intentions with direct certainty (as we do) but could only hypothesize about them. The dispositionalist can respond that the objection incorrectly assumes the truth of premises (a) and (b) in the argument above.

3. The two remaining objections that Kripke makes (26–32) I will consider together because I think there is a move the dispositionalist can make that deals with both difficulties at the same time. They are difficulties the dispositionalist faces in trying to explain our understanding of computable functions like addition. One problem is that there are numbers so large that I lack the capacity to consider them and, therefore, it seems to make no sense to suppose that I am disposed to respond in any determinate way if I should consider them; yet the function I mean by “plus” determines sums for those very large numbers, too. The other problem is that people are sometimes disposed to make systematic mistakes in computing sums; yet what they mean by addition is not different from what is meant by people who are not disposed to make those mistakes: they mean the standard function, not some bizarre function that includes what would be incorrect results on the ordinary understanding.
Regarding the first problem Kripke asserts (26–27) that
not only my actual performance, but also the totality of my dispositions,
is finite. It is not true, for example, that if queried about the sum of any
two numbers, no matter how large, I will reply with their actual sum,
for some pairs of numbers are simply too large for my mind—or my
brain—to grasp.

How is it that “some pairs of numbers are simply too large for my mind—
or my brain—to grasp”? In the sentence following the one just quoted, Kripke
suggests that there are numbers that would take too long to mention—one would
lie of old age first. Couldn’t a suitable notation help with this difficulty? We
would use exponents, special names for very large numbers ("googol"), and
have special rules for adding numbers expressed in these special ways. Are
there any integers so large that special notation and techniques could not make
it possible to add two of them in a reasonable length of time?

Kripke’s likely response to this would be that it misses the point of the
objection. The point is that there are numbers so large that I am not now
cognizable with notation and techniques that would make me able even to
mention them, much less add them; therefore, my current dispositions and
capacities with respect to doing sums do not reach beyond some finite limit.
That does seem right. We need to find a better response to this too-big-numbers
objection.

Suppose we grant that it is a feature of the addition function that, although
we grasp it and frequently mean it by what we say, we do not have dispositions,
of the sort we specified, for all the arguments to it. That is, for some trio of
integers $x$, $y$, and $z$ such that $x + y = z$ it is beyond my capacity to consider
the question whether $x + y = z$ and therefore the counterfactual conditional ‘If
I had considered whether $x + y = z$ and...etc., I would have judged that it
does’ is not true of me: it is either false or lacks a truth value altogether. There
are two important points to note.

First, this feature of the addition function is not a feature of every function
we mean by a descriptive term we use, not even every one that is defined
on an unlimited set of actual and possible cases and determined by a rule.
There are application functions for meaningful descriptive terms where no case
to which the function applies is more difficult to consider, as far as one’s
cognitive capacities are concerned, than any other. For color terms, shape terms,
‘one side of the visible rectangle is longer than the other’, ‘the noted change
pitch downward’, and the like the too-big-numbers sort of problem just does
not arise.17

Second, those application functions where this problematic feature (of
cases too “big” to consider) is present will always be ones such that a person
means the function by a term if and only if the person knows some correct
and informative definition of a rule for the function. That is, the person is
disposed to recognize as correct some correct expression of the rule and the
person possesses understanding of the terms used in that expression sufficient
to enable her to understand it. For these sorts of functions, having a capacity to
understand some correct expression of the rule and a disposition to recognize
it as correct when understood is a necessary as well as sufficient condition
of meaning the function. Whatever dispositions to respond to questions about
particular applications of the addition function a particular subject may have
or lack, this capacity and disposition to recognize a correct general definition
of addition is the overriding criterion of whether or not the subject grasps or
means the appropriate rule and function.

These points suggest a different and less crude dispositional account
of meaning such functions and rules from that which Kripke seems to be
considering, one to which his criticism does not apply. The account is that
understanding a term to mean an unlimited function determined by a rule is
always either a case of the sort where the cases-too-big-to-consider feature is
not present or it is a case where the understanding is built on the basis of
dispositions to recognize correct definitions plus understandings of the terms
of the definitions.

Let me elaborate. Consider the case of meaning the addition function by
the plus symbol, where, I have claimed, the overriding criterion of a person’s
meaning this is her knowing some informative correct definition of the function.
With respect to the subject’s understanding of each term in that definition, either
this condition of her knowing some correct definition of the term is satisfied
for it, too, or it is not. If it is satisfied then she must understand the terms of
some correct definition of this term. If it is not satisfied, then it must be a case
(such as that of understanding a color term) where the criterion of whether she
grasps the appropriate rule is just her having the appropriate dispositions with
respect to all the actual and possible cases to which the rule applies.

This second sort of understanding a rule for the application of a term is
what I earlier called the “brute level” way of doing so. And the distinction
between the first and the second sorts of understandings is, to borrow
Wittgenstein’s words, the distinction between a way of grasping a rule that
is an interpretation and a way that “is not an interpretation, but which is
exhibited in what we call ‘obeying the rule’ and ‘going against it’ in actual
cases” (Philosophical Investigations 201).

A higher level sort of understanding unpacks into various understandings
of terms in expressions of rules, at least some of which may also be higher
level understandings. This unpacking continues until brute level sorts of
understandings are eventually reached: there is a hierarchy of understandings built
up by understandings of rule expressions, or definitions, on a foundation of brute
level sorts of cases where understanding a rule is not a matter of understanding
an expression of it (not a matter of understanding a definition of the term).

The unpacking of any particular higher level understanding can be repre-
sented by a downward branching tree. However many branches the tree has,
they all eventually terminate in brute level understandings of terms, where,
for any case in the domain of the application function, there is no difficulty
about $S$’s having the capacity to consider it. For example, among the many
brute level understandings at the base of my understanding the rule for adding two numbers quoted earlier from Graeme Forbes would, perhaps, be sets of dispositions to recognize marks as instances of the numerals "0" through "9," to recognize when two numeral marks are in vertical alignment, and the like.

The dispositionalist thesis is, then, that understanding the meaning of a term is always either a brute level case or else a higher level case for which such an understanding tree can be constructed on the basis of brute level cases. This dispositionalism avoids Kripke's objection regarding the addition function (and the like) that, for any person who does understand and mean the function, there are arguments to the function beyond the capacity of that person to consider. This dispositionalism admits this point, but says that its dispositional account of grasping the addition function does not require that every argument to it be within the grasper's capacity to consider. It requires this only for rule-grasping of the brute level sort, and for these every argument to the function is such that there is no difficulty regarding the subject's having the cognitive capacity to consider it.

This dispositionalism avoids equally well, I think, the other objection mentioned above (presented by Kripke, 28–32), which is based on the fact that people can have dispositions to make systematic errors in calculating sums even when they correctly understand what addition is and do mean the right function by "+
h} (similarly for other functions). Thus there might be someone of whom are true a great many counterfactual conditionals of our prescribed form, where the antecedent has him considering a pair of numbers as argument for the plus function and the consequent has him giving the wrong value for the sum. This might be so despite his correctly understanding what addition is—as evidenced by, say, his being able to give a correct informative general account of it—because, for example, he is disposed to forget to carry when calculating the sum of numbers longer than two digits. Our dispositionalism can allow that such dispositions to make errors in applying terms are compatible with correct understanding of them, provided that they occur for terms that are at levels in the structure of one's understanding that are above the brute level. At these higher levels there is a criterion of one's correctly understanding the application rule that overrides one's dispositions to make erroneous judgments in particular cases, namely, one's understanding a correct definition of the term (which entails possessing an understanding tree for it that goes down to brute level dispositions).

On the other hand, there do seem to be terms with respect to which there cannot be both systematic errors in application and also correct understanding of the term. If I am disposed to errors in judging the application of "red" to particular objects when I have optimal visual awareness of them, then I do not grasp the correct rule for applying "red." There is not available another explanation of my errors that is compatible with my grasping the correct rule: simply do not understand the meaning of that term. I think the same goes for any of the terms for which one can grasp a rule of application without knowing any informative expression of the rule. These are the brute level cases on which.

On this theory, the explanation of what a person's understanding of a particular rule or meaning consists in depends on where it is in the structure of a person's understanding. If it is above the brute level then it will consist partly in the person's understandings of other rules or meanings. The account tells us how every grasping of a term's meaning can be given a finite analysis that does not appeal to her understanding of this term's meaning (though, if the understanding in question is not at the brute level, it will appeal to the person's understandings of other terms' meanings). And, most important for the problem we have been concerned with, it does not appeal to any mysterious state that "contains" all the applications in an unexplained way. For each particular understanding of a term's meaning, the sense in which it does contain all the applications is explained in terms of a tree of understandings which has at its terminal nodes sets of dispositions to apply or withhold terms in particular cases.18

Notes

1. Kripke (1982), 7; all references to Kripke are to this work.
2. I always use "disposition" to refer to the counterfactual conditional fact itself, never to a categorical fact that is its causal basis (if there is one).
3. It cannot be said, however, that there is conclusive evidence that dispositionalism was Wittgenstein's own solution to his puzzle. I do doubt that his solution was the "skeptical" one that Kripke (chapter 3) attributes to him.
4. Note that this objection applies only to cases where the application function I mean is a collection of ordered pairs that is, in fact, not arbitrary, but determined by a general rule. It is the general rule that determines the function and makes nonarbitrary, justifies, the inclusion of each ordered pair. In cases where the function I mean by a term is in fact arbitrary and not determined by a general rule—as, for example, the function we usually mean when we speak of the alphabetical order of the letters of the alphabet—the question Kripke is raising here does not apply; for here there is not the sense that all the different assignments of value to argument are justified by a single rule.
5. But we have already made some significant progress with the puzzle. Dispositions show how a fact about me now can reach an object or state of affairs not present to me now and thus provide a non-qua sense in which my present state of meaning (intending) a certain application function can "contain" all the applications not presently before my mind.
6. To try to help make clear how insufficient a merely arbitrary collection of dispositions must be, Kripke introduces (on p. 24) the example of the star symbol. He writes:

Assuming determinism, even if I mean to denote no number theoretic function in particular by the sign 'a', then to the same extent as it is true for 'a', it is true here that for any two arguments \( m \) and \( n \), there is a uniquely determined answer \( p \) that I would give. (I choose one at random, as we would normally say, but causally the answer is determined.) The difference between this case and the case of the 'a' function is that in the [latter] case, but not in the [former], my uniquely determined answer can properly be called 'right' or 'wrong'.

And he comments in a footnote to this passage:

I might have introduced 'a' to mean nothing in particular even though the answer I arbitrarily choose for 'a' is, through some quirk in my brain structure, uniquely determined independently of the time and other circumstances when I am asked the
in terms of the meaning I assigned to 'a', as I will for 'b', since there is no such meaning.

The example is not spelled out as fully as one would like. What is the question about 'm ≠ n' to which I am disposed to give an answer? If it is "What is the value of 68 + 57?", why shouldn't my response be simply, "I don't know; the stars mean nothing to me?"

Kripke wants the case to be one where (1) it is plausible to suppose that I am disposed to come up with a number when asked the question (whatever it is) about any instantiation of "m ≠ n"—so my dispositions determine a function on all pairs of integers—but (2) my having this collection of dispositions does not constitute my understanding the stars symbol to mean that function (or anything else). Perhaps both conditions are satisfied if the question I am disposed to respond to with a value is "What value do you (now) arbitrarily choose to assign m ≠ n?" The word "arbitrarily" needs to be in the question or, otherwise, my response is likely to be "How am I to decide what value to assign?"

If this is the sort of case Kripke has in mind, then the dispositionalist might be tempted to deal with it by saying that the counterfactual conditionals involved here are importantly different from those the dispositionalist posits in the case of meaning the addition function by "+". For an application of "+" the conditional would read: "If I had been asked to arbitrarily choose a number to associate with '68 + 57', I would have chosen 125." For an application of "≠" the conditional (as I most recently formulated it) reads: "If I had considered whether the value assigned to '68 + 57' by the function I mean by '≠' is 125 and had given a relevant and sincere response, I would have said that it is." In the difference between these, the dispositionalist might say, lies the difference between meaning a function in the one case and failing to do so in the other.

It is fairly clear what Kripke's response to this move should be. It is to image a person who has an unlimited set of dispositions to respond to star symbol pairs, each of which dispositions is specifiable by the sort of counterfactual conditional the dispositionalist gives for the plus symbol. Such a thing is possible. Now, one of two things must hold of such a person. (1) Her set of dispositions is an arbitrary collection and not determined by her grasp of a rule; in that case, despite the nature of the reactions she is disposed to have, it is doubtful that her having that set of dispositions constitutes the subject’s meaning something by the sign. Why should we think that merely having all those dispositions is meaning that completely arbitrary function? (2) Her set of dispositions is not arbitrary but is determined by the subject's grasp of a rule, in which case the dispositionalist still owes us an account of this latter, crucial aspect of the matter; we need not only the collection of dispositions, but also some non-mysterious fact about the subject that makes them all in accord with a single rule she grasps.

Forbes (1984), 224. The table referred to is one which gives the sums of all pairs of single digits. To carry 1 is to add to the digit heading the next column to the left, unless (a) that digit is 9, in which case one changes it to 0 before adding that column and carries 1 to the next column to the left of it, or (b) there is no column to the left, in which case one adds a leftmost place to the sum and puts 1 in it.


9. It should be noted that, for simplicity's sake, we are ignoring the fact that among the relevant sorts of responses to which the dispositionalist needs to posit dispositions in brute level cases is "This is not a case to which that question applies," which are needed to handle cases for which the application function determined by the rule is not defined. And we are ignoring the fact that for vague terms (which may include all empirically descriptive terms) there are needed dispositions to respond in certain appropriate cases with "That is not a clear case either way; it's on the borderline."

10. And a case of pseudo-rule-following will be one where there is no property in common to all the possible cases to which the subject is disposed to have the specified reaction.

ANSWERING KRIPE'S OBJECTIONS

11. On this issue of whether one can know such counterfactual conditionals about oneself in the direct sort of way, it is not clear that Wittgenstein always sides with me rather than with Kripke. On p. 142 of the Brown Book Wittgenstein seems to take Kripke's line on how I know these conditionals about myself; they are hypotheses my confidence in which should be based on observation and experiment (and, presumably, well-confirmed theory about how things work). In the same passage, Wittgenstein appears to doubt the view Kripke and I share, namely, that I do know with direct certainty, and not merely as a hypothesis, what my meaning intentions were.


13. Independent because the sort of disposition we are concerned with could itself be regarded as an intrinsic property.

14. We are concerned here only with cases where the truth of the counterfactual conditional about me is not guaranteed by properties of things external to me.

15. Wittgenstein (1976), 410.

16. Some may say that, however it may be with the conceptual requirement, we in fact have plenty of reason in what we know about how nature works to be justified in being confident that the truth of any contingent counterfactual conditional about a macroscopic object supervenes on the microphysical properties of the object and its circumstances, and, hence, that the truth of any counterfactual conditional of the sort we are concerned with does have a categorical basis. Perhaps so, but if we substitute this claim for (a) in the reasoning I attributed to Kripke, then the revised premise (b) looks very doubtful indeed. (b) becomes: If in fact the truth of any such counterfactual conditional does have a categorical basis, then I could not know that it was true of me without knowing that there is a categorical basis for it. I see no more reason to suppose that this is so than to suppose that one must know that sensations supervene on brain processes in order to know that one feels pain.

17. To be sure, for a perceptible object remote from the subject, or one not actually in existence, it may be that it would be practically impossible for the subject to perceive the object in the way required by the antecedent of the counterfactual conditional (one cannot easily get rocks on Mars within one's field of view, one does not have the resources to manufacture certain chemical compounds). But this is not the sort of difficulty regarding the subject's mental capacity to consider them that Kripke cites for very large numbers. One still knows quite well what it would be like to have a good look in good light at rocks now lying on the surface of Mars or to have a good look at a certain chemical concoction that has not been and never will be created. The counterfactual hypothesis that one does these things has definite truth-conditions which do not entail one's having any cognitive capacities one now lacks.

18. I have discussed the issues in this essay in several courses at Cornell and have received a lot of help with them from people attending those courses. I am grateful to Sydney Shoemaker for a number of helpful comments.

References


