Explaining the Seemingly Self-Interpreting Character of a Formula

Kai-Yuan Cheng, Min-Hsiung, Taiwan

I. The Puzzle About the Self-Interpreting Character of a Formula

In the notable discussion of rule-following, Wittgenstein (1974) sharply and rightly brings our attention to the fact that meaning is essentially normative, in the sense that the meaning of a linguistic term sets a standard of correctness and incorrectness for the application of the term. It is only because a person uses a particular term to mean a particular thing that the person's linguistic performances with regard to the application of the term can be evaluated as correct or incorrect. Take the term "add 2" as an example where it is most clear and uncontroversial that there is a uniquely correct answer for each step of application. To say that a person understands "add 2" to mean the mathematical rule *add 2*, is tantamount to saying that the person's performances, past, present, and future, are subject to the regulation of an infinite number of instances sanctioned by the *add 2* rule. Hence, the meaning of a term, by its nature, contains a normative element in it, and a person's understanding of the meaning of a term constrains the person's linguistic performance with regard to the application of the term in a certain way in each particular case.

However, a philosophical conundrum arises when we begin to ask what meaning or understanding consists in. What is it that is capable of delivering the normative force, and of bringing about the evaluative effect regarding the application of a term, such as "add 2"? One of the proposals that Wittgenstein considers is a natural suggestion that when a person understands the meaning of "add 2", the person has a verbalized formula, e.g., f (x) = 2x + 2, occurring in her introspective consciousness. It is this verbalized formula that guides and constrains a person's performance in each particular case when one continues a certain numerical series.

This formula proposal is *prima facie* implausible, given the fact that not all of the speakers who understand "add 2" and are capable of continuing a numerical series accordingly, have a formula in mind. Nevertheless, the formula proposal is tempting. Suppose that those who understand the term "add 2" do have a certain formula in mind. Doesn't it seem natural to them that the formula is self-interpreting, in the sense that it determines the procedure for each step that they should take in order to be correct? A formula, say f (x) = 2x + 2, clearly interprets itself in such a way that a certain series is specified as follows: f (x) is 2 given that the value of x is 0, f (x) is 4 given that the value of x is 1, f (x) = 6 given that the value of x is 2, and so on. Hence, this intriguing self-interpreting character of a formula *seems* sufficient to determine the rule referred to by the term "add 2", and hence *seems* sufficient to constitute a person's understanding of the term.

Wittgenstein has convincingly shown us, however, that the formula is unable to determine the rule being followed, because the signs in the formula, such as "+", can be interpreted in different ways. Any chosen interpretation would require further interpretation to fix its content. This thus leads to an infinite regress. Kripke's (1982) discussion of the counting case is illuminating in this regard. Consider the proposal that one's following the *plus* rule consists in one's being guided by a basic counting rule or algorithm which specifies a finite set of simple procedures. For example, to add 2 and 3 one should first count out a group of two things, one at a time, and then count out a group of 3 things, also on at a time, and finally put the two groups together, and count the totality, again one at a time. Nobody would deny that this procedure of counting is a legitimate way of adding. Some of us may use this procedure to add on some occasions. However, Kripke rejects it on the ground that it leads to an infinite regress, because the set of instructions involves the term "counting", which is in need of a further interpretation to determine its meaning. As a result, a set of instructions such as the counting procedures does not *fix* the rule being followed. This result appears counter-intuitive, but it is a solid point.

The puzzle that is the main concern of this paper can be made clear: a formula seems to be self-interpreting, but is in fact insufficient to determine the rule to be followed. We may put the puzzle in another way. There appears to be no gap in our phenomenology, between a verbalized formula and its interpretation, because of the *seemingly* self-interpreting character of the formula. However, in reality, a formula is insufficient to *fix* the rule to be followed. How may this puzzle be explained? In what follows, I explain why the formula seems self-interpreting, when it is not.

II. Explaining the Puzzle

In my account, a person's understanding the term "add 2" as meaning the f (x) = 2x + 2 function does not consist in her having the formula in mind. Rather it consists in the person's *possessing a disposition* to perform in a certain way. Furthermore, it is by virtue of the special way in which the person perceives her disposition, that she does not feel in her phenomenology any gap, between the formula and the interpretation of the formula.

My account has two parts. The first part is to characterize a disposition in terms of a functionalist account, which is regarded by McLaughlin (1995) as the leading theory of dispositions today. Functionalism is well known as a theory of mind in which states of mind and mental properties are construed as functional states and properties. Though there are a variety of functionalist theories, the basic and common idea of a state's possessing a functional property is for the state to occupy a certain causal role, relative to other states. A mental property or state has to be analyzed not merely in terms of its input cause and output effect, but also by reference to other mental properties or states. These other mental states in the network, such as believing that P and desiring that Q are to be in turn characterized by reference to their causal relations with other states in the network.

We may equally view a disposition to follow a rule as being a functional state, which is defined by its causal role in the person's functional organization. As a result, a rule-following disposition can be viewed as a complex functional property of a person, having a potentially infinite pairing of input-stimuli and output-responses, and having interrelations with other inner dispositional or functional states.

The second part is to show, given that a rule-following disposition is functionally characterized, how its possessor comes to know about such a functional state. How does a subject gain conscious introspective access to the complex causal network that she embodies? Dennett (1978) offers a plausible functionalist account of introspection in terms of a computer model that will help illuminate this issue. Roughly put, Dennett's account is that introspection is the series of processes by which Control, a higher executive component in the brain, directs questions to the buffer memory, gets the answers after some possible editions and transformations, and sends them to the speech center, which if so commanded, publishes them in a written or verbal form or in an internal voice. The process of introspection is best described as asking questions and receiving answers, rather than a direct monitoring or tracking of one's first-order inner processes. Dennett proposes that what a person is conscious of depends on what the person can introspect, and what the person can introspect is explained as a routine by means of which he gain access to and reports on the contents of his buffer memory.

Taking the two parts together, we may now explain the seemingly self-interpreting character of a formula. In my account, a person's use of "add 2" is determined by the person's disposition to behave in a certain way. More specifically, the person is disposed to answer f (x) = 2 given x = 0, f (x) = 4 given x = 1, f (x) = 6 given x = 2, and so on. Moreover, given Dennett's account of introspection, in which the process of introspection is depicted as asking questions and receiving answers, the objects of introspection when a person exercises his dispositional power of following a rule are the numerical series that figures in the contents of the person's utterances or written sentences or inner voices such as f (x) = 2 given x = 0, f (x) = 4 given x = 1, f (x) = 6 given x = 2, and so on. These contents are exactly "rooted" in a disposition which one perceives through directly expressing them or having thoughts about them. In our introspective consciousness, therefore, the formula makes us feel that it commands us to perform in a certain way in which a certain numerical series is generated. So, phenomenologically, there seems to be no gap, between the verbalization of the formula and its interpretation.

In the Dennettian picture, we may not know exactly how we get, say, f (x) = 4 given x = 2, which is achieved by a complex first-order process that is not accessible to our introspective consciousness. In a functionalist account of cognitive processes and introspection, what we introspect is that *there* is a mechanism which produces 4 when 2 is put in. However, we cannot introspect what that mechanism is. The formula is the hypothesis we form about our own inner workings when we reflect on the output that a dispositional state gives when given input. The hypothesis is the result of our taking the interpretive stance towards the behaviors or the disposition. Therefore, Wittgenstein is right that the rule is not the mechanism. The verbal report of the introspected formula is merely the report that there is some mechanism.

What is important to note is that, in my account, a verbalized formula is a mere hypothesis that I am following a certain rule, and that hypothesis may not be correct. It might not correctly capture my disposition or behaviors. The mathematical function that I am *actually* following has to be determined by the underlying disposition that I possess. I do not interpret a formula when I follow a rule. I simply act on my disposition. The verbalized formula is the appearance that my disposition takes on to my introspective consciousness. This explains why a formula seems to be self-interpreting.

III. Conclusion

A dispositional account can nicely explain the seemingly self-interpreting character of a rule or formula, by showing why the gap, between a verbalized formula or rule and its interpretation, is not phenomenologically apparent. We never noticed that a formula fails to be self-interpreting, because we don't interpret it when exercising our linguistic competence. When we exercise our linguistic competence, we simply act on our dispositions. This account does justice to our intuition that rule-following and linguistic meaning are transparent to the speaker, and do not need interpretation. At the same time, I preserve Wittgenstein's insight that a verbalized formula or algorithm does not constitute rule-following.

Literature

Dennett, D.C. 1978 *Brainstorms: Philosophical Essays on Mind and Psychology*, Bradford Books.

Kripke, S. 1982 *Wittgenstein on Rules and Private Language*, Oxford and New York: Blackwell.

McLaughlin, B. P. 1995 "Dispositions", in J. Kim and E. Sosa (eds.), *A Companion to Metaphysics*, Oxford: Blackwell, 121-4.

Wittgenstein, L. 1974 Philosophical Investigations, Oxford: Basil Blackwell.