A CRITIQUE OF KRIPKE'S FINITUDE ARGUMENT

In *Wittgenstein on Rules and Private Language* [WRPL], Kripke interprets Wittgenstein as wrestling with the following problem about meaning: *Is there any fact about me that constitutes what I mean by a word?* This problem is developed by Kripke in WRPL in terms of an example from arithmetic. Suppose, the example runs, that I am given a computation that I have never performed before – say, '68+57=?'. I then perform the computation, obtaining '125' as my answer, and after checking my work, I am confident that '125' is the correct answer. Now suppose, Kripke continues, a 'bizarre skeptic' comes along who questions my confidence. 'What fact about you makes it the case that, as you used "+" in the past, the answer you meant or intended for "68+57" should have been 125 rather than, say, 5?' he asks. [p. 8][1] 'Perhaps', he continues, 'as you used the term "+" in the past, you meant some nonstandard function', which he calls 'quus', and defines as follows:

$$x \oplus y \quad = \quad x+y \text{ if } x, y < 57$$
$$= \quad 5 \text{ otherwise}$$

If I meant *quus* by '+' in the past, then the answer I intended for '68 + 57' should have been '5', so that, if I am to accord with my past intentions, the answer I should now give is '5'. What fact makes it the case (makes it true) that I meant *plus* and not *quus*?

For future reference, I will summarize the skeptical problem as follows:

---

[1] All bracketed page number references are to WRPL, unless otherwise indicated.

[SKEP]   Is there any fact about me that constitutes my meaning *addition* by 'addition' (or, more generally, my meaning *w* by 'w', for any expression w of our language)?

The skeptical argument developed by Kripke in Chapter 2 of WRPL proceeds by canvassing a range of candidate straight solutions to [SKEP]. One of these is the so-called dispositional theory of meaning, which Kripke formulates as follows for the term '+':

[KD^{plus}]   S's meaning *addition* by '+' consists in the fact that for any p, q, S is disposed to answer r to the query 'p+q=?' iff r is the *sum* of p and q.

Kripke raises four objections against [KD^{plus}]: (1) the finitude objection, (2) the rrror objection, (3) the circularity objection, and (4) the justification objection. In this paper, I will focus on the first of these objection: the finitude objection.

Kripke's finitude argument against the dispositional theory of meaning emerges in the following:

The dispositional theory attempts to avoid the problem of the finiteness of my actual past performance by appealing to a disposition. But in doing so, it ignores an obvious fact: not only my actual performance, but also the totality of my dispositions, is finite. It is not true, for example, that 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. When given such sums, I may shrug my shoulders for lack of comprehension; I may even, if the numbers involved are large enough, die of old age before the questioner completes his question. [WRPL p. 27-28]

One problem Kripke raises for [KD^{plus}] in the quoted passage is that it is not true that if queried about the sum of *any* two numbers, no matter how large, I will answer with their sum: some pairs of numbers are simply too large for my mind to grasp. In this sense, my additive dispositions are

finite – they can only extend so far. But surely I can mean *addition* by '+' despite the fact that I

don't have dispositions to handle huge sums.

This objection shows that the *right to left direction* of the analysans in [KD^{plus}] fails, i.e., that:

$\forall$p,q, S is disposed to answer r to the query 'p+q=?' $\nleftarrow$ r is the *sum* of p and q.

and forms the key premise in the finitude argument. As I see it, the argument (which can be

extracted from the passage) takes the form of a *reductio*, and may be summed up as follows:

*Finitude argument against the dispositional theory (applied to '+')*
i Suppose I mean *addition* by '+'.
ii According to the dispositional theory [KD^{plus}], I mean *addition* by '+' if and only if I am disposed to give answers that covary with the set of ordered triples that lie in the extension of '+'.
iii Since '+' applies to objects that are *epistemically inaccessible* to me, there will be triples <p,q,r> in the extension of '+' over which I will not be disposed to answer r to the query 'p+q=?'.
iv Therefore, according to the dispositional theory [KD^{plus}], I do not mean *addition* by '+', or understand it.
v But this contradicts the supposition (i) that I do mean *addition* by '+'. So the dispositional theory [KD^{plus}] must be wrong.

It will not help the dispositional theorist to reformulate [KD^{plus}], Kripke adds, by appealing to

how I would respond under *ideal conditions* that overcome the finiteness:

[KD$_I$^{plus}] To mean *addition* by '+' is to be disposed, *under ideal conditions*, to answer r
to the query 'p+q=?' iff r is the *sum* of p and q.

For how are we to flesh out the ideal conditions? The proposal that we characterize them as

conditions where 'I am given the means to carry out my intentions' will succeed, Kripke says,

only if we 'presuppose a prior notion of my having an intention to mean one function rather than another by "+"'. [p. 28] We cannot confidently say that, if I am given the means to carry out my intentions, I will give answers that accord with the addition function, for how am I to be given these means? If my mind is stuffed with extra brain matter, or my life prolonged by some magic elixir, we would have no idea how I would be disposed to apply '+'. Stuffing my brain with extra matter or prolonging my life with magic elixirs might lead me to go insane! And even if I am not led to go insane, there is no guarantee that I would be disposed to give answers that accord with the addition function as opposed to a function – let us call it skaddition -- that coincides with the addition function on some finite (surveyable) initial segment of the addition function, but that diverges from it thereafter
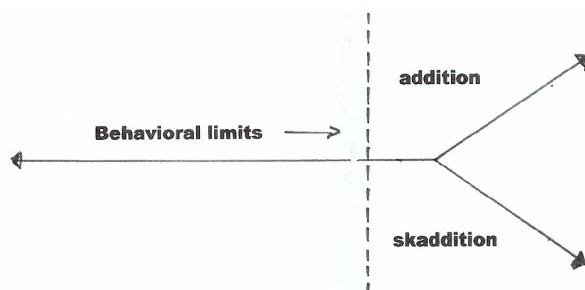
unless, Kripke says, we presuppose that I have formed an intention *to give answers that accord with the plus function*. But, he maintains, building this clause into the ideal conditions would render [$KD_I{}^{plus}$] circular.

Critique of the Finitude Argument

Where, if anywhere, has the finitude argument gone wrong? Premise (iii) assumes that if I mean *addition* by '+', then, I can't be disposed to apply '+' in any old way -- I must apply it correctly. This much seems correct. As Michael Dummett points out, using language is not like doing

one's hair and taking a bath. One may do either of these two things as one likes, and still be doing it. But if there cease to be right and wrong uses of a word, the word loses its meaning. [1991, p. 85] But, as Kripke presents the dispositional theory ([KD^{plus}]), I am disposed to apply '+' **correctly**, given that I mean *addition* by '+', if and only I am disposed to give answers that covary with the set of ordered triples that lie in the **extension** of '+'. This generates a problem because the extension of '+' involves terms that are ungraspably huge, so that (as premise iii states) the *right-to-left direction* of the biconditional in [KD^{plus}] breaks down, i.e.:

$$\forall p,q, \text{ S is disposed to answer r to the query 'p+q=?'} \nleftarrow p+q=r$$

Some terms apply to objects that are **epistemically inaccessible** to us, and hence objects about which we do not have any normal dispositions. 'Plus' is one such term, for a certain range of values. Thus for some trio of integers p, q, and r, such that p+q=r, it is beyond my capacity to consider the question whether p+q=r, and so the statement 'If <p,q,r> is in the extension of "+", I would be disposed to answer r to the query p+q=?' is not true of me: it is either false or lacks a truth value altogether. But, we may reasonably ask, why should I have to get the sum right for numbers that are *too large for my mind to grasp* to succeed in meaning *addition* by '+'? This requirement places *too great an epistemic burden* on the speaker – indeed one that is *impossible* to meet. Given that no human can meet this requirement, no human can succeed in meaning *addition* by '+', or in understanding the term, according to [KD^{plus}]. Thus, by characterizing the dispositional theory in a way that leaves this requirement in place, Kripke creates a *straw man*.

If we leave the requirement in place, we can try to meet it by adopting a proposal like that offered by Simon Blackburn, which characterizes the answer I *would* give for indefinitely large

sums as the answer I would give by 'reiterat[ing] procedures I am disposed to use a number of times'. [1984 p. 289].

I think a simpler solution, though, is simply to restrict the biconditional in [KD^{plus}] for terms like '+' whose extensions are determined by a computational or algorithmic procedure.

> S's meaning *addition* by '+' consists in the fact that **for any p, q, r that S is capable (in principle) of grasping**, S is disposed to answer r to the query 'p+q=?' iff r is the *sum* of p and q, and *addition* is the simplest function corresponding to these behavioral limitations.

What this version of the dispositional theory entails for '+' is that I need not possess dispositions that get indefinitely large summation queries right – queries concerning numbers that are too large for even the fastest computer to handle -- to succeed in meaning *addition* by '+', a demand which no human can meet, but only an *initial segment* of the function involving the longest computations that my mind is (in principle) capable of grasping. Thus reformulated, the dispositional theory would be invulnerable to the finitude objection.

So it seems, but Paul Boghossian, among others, has offered an example which purports to show that the finitude objection applies to non-mathematical terms, like 'horse', whose extensions are not determined by a computational or algorithmic procedure. Thus, he says: 'if it is indeed the property *horse* that I am disposed to apply the term to, then I should be disposed to apply it to *all* horses, including horses so far away and so far in the past that it would be nonsense to suppose I could ever get into causal contact with them. Otherwise, what is to say that my disposition is not a disposition to apply the term to the property *nearby horse*, or some such? But no one can have a disposition to call all horses 'horse', for no one can have a disposition with respect to inaccessible objects.' [1989 p. 528-9]

In order to determine whether Kripke's formulation of the dispositional theory is vulnerable

to this objection, we first need to determine what a more general version of the theory is supposed to look like. Here we run into some difficulties, because Kripke never describes the more general version.

In his discussion of Kripke's dispositional theory, Paul Horwich offers the following schematic generalization of '[KD^{plus}]':

[KD]   S's meaning F by w consists in the fact that S is disposed to apply w to x iff x is an f. (where 'f' is to be replaced by a predicate (e.g. 'dog') and 'F' by a name of the meaning of that predicate (e.g. 'DOG')

He then argues [1998, 1995, 1990] that [KD] is implausible and should be rejected because it presupposes an *inflationary* theory of truth. According to Horwich [e.g. 1995 p. 363-364], [KD] is designed to provide a way of 'reading off' a predicate's extension from whatever property constitutes its meaning. In other words, [KD] implies that we can determine the extension of w from how S is disposed to apply w:

w is true of x iff S is disposed to apply w to x

This implies that there are non-semantic necessary and sufficient conditions for being true of – some account of the form

w is true of precisely the f's iff R(w,f)

where R(w,f) is deducible from whatever non-semantic property constitutes the meaning of w. But the existence of any such theory is plausible, Horwich maintains, only if *being true of* is a

substantive relation, one that has some non-semantic underlying nature, some naturalistic analysis. And that, he maintains, is precisely what deflationism, which he upholds, denies.

We can use some terminology developed by Crispin Wright to see Horwich's point more vividly. Consider the following statement:

(i) For any act x, x is <u>pious</u> iff x is loved by the Gods.

In *Truth and Objectivity* (Appendix to Ch. 3: "The Euthyphro Contrast: Order of Determination and Response-Dependence", pp. 108ff), Wright points out that (i) is common ground between Socrates and Euthyphro. So what is the difference between them? Developing an idea deployed by Mark Johnston[2], Wright proposes that the difference lies in the fact that Socrates accords a certain *order of priority to the left hand side* of the biconditional, which he terms '**detectivism'** about piety, while Euthyphro accords a certain *order of priority to the right hand side*, which he terms '**projectivism'** about piety.

Johnston[3] used this distinction to differentiate primary and secondary qualities. Consider, e.g., the following statements:

(ii) x is <u>square</u> iff x would look square to standard observers under standard conditions.
(iii) x is <u>red</u> iff x would look red to standard observers under standard conditions.

Johnston proposed that (ii) be read in a *detectivist* direction involving left to right priority, and (iii) in a *projectivist* direction involving right to left priority. The latter involves 'writing human

---

[2] See his 'Objectivity Refigured: Pragmatism without Verificationism' in J. Haldane and C. Wright (eds) *Reality: Representation and Projection* (New York: Oxford Unviersity Press, 1992).

[3] Wright notes that Johnston developed this view in a seminar delivered at Princeton in 1986, but has since moved

responses into the account of a substantial truth condition'. [*Truth and Objectivity*, p. 109] In other words, judgments about color are claimed by Johnston to be **response-dependent**:[4] the verdicts of standard observers under standard conditions **determine the extensions** of our color terms. On this reading, (iii) is true by definition: the biconditional stands for conceptual equivalence. Shape terms, on the other hand, according to Johnston, are response-independent: the verdicts of standard observers under standard conditions **track** an independently constituted extension. So (ii) is a contingent truth: the verdicts of standard observers under standard conditions *may* fail to track an independently constituted extension.

Now, consider again Kripke's version of the dispositional theory.

[KD]   S's meaning F by w consists in the fact that x is an f iff S is disposed to apply w to x.

In light of the terminology developed above, we can say that what Horwich finds problematic in [KD] is that it makes the meaning of our terms response-dependent, since their extensions are defined by how agents (it doesn't matter if we replace S with a community of agents C) are disposed to apply them, and this, he maintains, entails an inflationary theory of truth. It will not help, Horwich says, to introduce a ceteris paribus clause into [KD] (as we find in ii and iii above), for, he agrees with Kripke and, as we will later see, Boghossian, that such a clause cannot be described *non-circularly*. [See 1995 p. 361, 1998 p. 219]

According to Horwich [1998 p. 219], Kripke's imposition of the '**reading off' requirement** causes him to assume that the right dispositional analyses have the form [KD]. Recall that, as

---

away from it.

[4] This term is Wright's. I am using it interchangeably with what I referred to as the projectivist view above, without assuming any of the additional properties -- in particular those involving necessity and aprioricity, which I find

Kripke defines this requirement, the value that I (or my linguistic community) means '+' to have for m and n is, *by definition*, the answer I (or they) would give if queried about m and n. [WRPL p. 25] Horwich believes that we can in fact give a dispositional account of meaning, and that Kripke's argument to the contrary is fatally flawed because it wrongly presupposes that a dispositional theory must meet the reading-off requirement. What we need to do, according to Horwich, is reject this unreasonable constraint.

I agree with Horwich that [KD] is implausible and should be rejected. However, unlike Horwich, I do not think the fatal flaw lies in Kripke's 'reading off' requirement. In other words, it is not the *projectivist* reading of the biconditional in [KD] or its communal version that I find problematic. One can, along with Horwich, interpret the biconditional in [KD] or its communal version in a *projectivist* fashion, and perhaps this is how Kripke intended it to be read. However, one might also interpret it in a *detectivist* fashion, and it is this latter reading of [KD] or its communal version that I find problematic. For one thing, the standards for what counts as a correct application of a term, on this reading, would have to issue their requirements independently and in advance of communal verdicts for an open-ended range of situations. But how can they 'reach ahead of us', so to speak, and determine of themselves their every actual and counterfactual application? And how can we account for our ability to be appropriately sensitive to the specific requirements that they demand?[5]

I believe a more reasonable and defensible version of the dispositional theory is the following:

---

highly problematic -- that Wright builds into it in *Truth and Objectivity*.

[5] At least two (metaphysical and epistemological) difficulties emphasized by Wright [See, e.g., 2001, which collects many relevant articles] in his critique of detectivism, though he does not apply these objections to Kripke's

> [D_{CP}]   S's meaning by w what is meant by it in his linguistic community C consists in the fact that (1) in ceteris paribus conditions, S is disposed to apply w in ways that agree with how his linguistic community C is disposed to apply w, and (2) C's meaning F by w consists in the fact, in ceteris paribus conditions, they are disposed to apply w in such-and-such a way.

On this social theory of meaning, the norms that govern linguistic practice are determined by the community. Language mastery is *parasitic* on communal practice, on more or less uniform patterns of collective behavior. No appeal to mysterious Platonic entities – 'extensions' having a life of their own, so to speak -- is necessary (as the detectivist assumes) once we interpret the extensions of expressions as *projections* of communal use.

It should be noted that I am employing the notion of a '**community**' loosely in [D_{CP}]. It is an elastic notion, like, and not in any significant respect different from, Wittgenstein's concept of a *language-game*. Sometimes 'community' refers to S's linguistic community *as a whole*; sometimes to a *proper subclass* of his community (e.g., biologists, mathematicians, historians, art critics, etc.). For many technical terms, this subclass involves a class of specialists – people who have expert knowledge of the term. For terms where it makes no sense to speak of someone having 'expert knowledge' of them – e.g., 'chair', 'table', 'beautiful' – 'community' is to be interpreted as referring to the subject's linguistic community *as a whole*. 'Community' can also refer to a narrow use-context involving two members (or 'players', as Wittgenstein would call them). How small can this use-context be? Can S himself constitute a 'community'? If so, must his language be accessible to others? These questions form the substance of Wittgenstein's famous private language argument. Though they are interesting and important, I do not, of course, have the space to deal with them here. What I would like to note is that [D_{CP}] does not rule out the possibility of a private language, IF we permit an individual to constitute a

---

discussion of the dispositional theory as I do here.

'community'. Of course, the possibility of creating standards in this way assumes that the members of the 'community' have a propensity to *agree* in judgment; in absence of such agreement, standards could not emerge.

Now, how does [D$_{CP}$] fare with Boghossian's extension of the finitude objection? I stated that it is only terms whose extensions can be determined by a computational or algorithmic procedure that can be thought of as response-independent; it is only here where it makes sense to say that communal use may fail to track an independently constituted extension. Boghossian appears to be assuming that the extension of 'horse' is determined in advance of communal verdicts; that it somehow takes care of itself. But, as I have argued here, whether an object falls under the extension of 'horse' is a decision that is made by the relevant experts on the matter (here, biologists), and not something that is determined in advance. And so long as my use conforms with that of the experts, I mean what they do by it. So [D$_{CP}$] is not vulnerable to this objection.

References

Blackburn, Simon
    [1984] [a] 'The Individual Strikes Back', *Synthese,* 58, pp. 281-301.
Boghossian, Paul
    [1989] 'The Rule Following Considerations', *Mind,* 98, pp. 507-49.
Dummett, Michael
    [1991] *The Logical Basis of Metaphysics*, Cambridge MA: Harvard University Press.
        [Chapter 4]
Horwich, Paul
    [1998] *Meaning*, Oxford: Clarendon Press. [Chapter 10]
    [1995] 'Meaning, Use and Truth', *Mind,* 104.
    [1990] 'Wittgenstein and Kripke on the Nature of Meaning', *Mind and Language,* 5(2).
    [1984] 'Critical Notice: Saul Kripke: Wittgenstein on Rules and Private Language',
        Philosophy of Science, 51, pp. 163-171.
Wright, Crispin
    [2001] Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical
        Investigations, Harvard University Press.