Saul Kripke

Bryan Frances

In Barry Lee, ed., ***Key Thinkers in the Philosophy of Language***, Continuum.

We are concerned with two main issues: meaning, and linguistic understanding. Saul Kripke has investigated several fundamental philosophical questions concerning these, but in this chapter I will focus on just two themes of his work: what he said regarding the paradigm set up by Gottlob Frege (which touches on both meaning and linguistic understanding), and what he said about the metaphysics of meaning. I will focus on his "A Puzzle About Belief" and *Wittgenstein on Rules and Private Language* more than on his *Naming and Necessity*, as the latter has already received much competent treatment at introductory and intermediate levels.1

By the late 1960s Fregean views dominated the philosophy of language, and with good reason. But in the 1970s Saul Kripke reduced that dominance and repositioned research into meaning and linguistic understanding. I begin below by reconstructing key elements of the Fregean paradigm, ones relevant to belief sentences. Then I explain how Kripke challenged those elements. I'll finish my discussion of the Kripkean challenges to the Fregean paradigm by sketching some of Kripke's positive theses regarding meaning and reference.

Then I move to Kripke's work on the *metaphysics* of meaning. In order to have anything like a systematic theoretical account of meaning it will probably be helpful to have an idea as to what meaning *is.* Kripke is interested in the underlying basis of facts such as 'Jones, like many of us, means addition by his use of "+"'. What facts about Jones make it true that when he uses '+' he is using it to express addition rather than something else (e.g. some other mathematical function)? Are there facts about Jones' mind or brain or linguistic behaviour that determine that he means addition by '+'? Kripke has no published answers to these questions, but in his landmark book *Wittgenstein on Rules and Private Language* he presents (without endorsement)

1 In the philosophy of language Kripke has also done significant work on theories of truth (1975), speaker's reference vs. semantic reference (1977), logical form and Russell's theory of descriptions (1976, 2005), presupposition and anaphora (2009), the social character of naming (1986), and the interpretation of Frege's philosophy of language (2008). Most of these essays demand more of the reader than the works examined in this chapter. A number of his unpublished essays will soon be published in several volumes under the title *Collected Papers* by Oxford University Press.

an argument inspired by his study of Wittgenstein that has the shocking thesis that "[t]here can be no fact as to what I mean by 'plus', or any other word at any time" (1982, 21). As we will see below, what this claim amounts to, even in the context of Kripke's discussion, is a difficult matter. We will also see that he provides the beginnings of a positive account of how to understand putative truths such as 'Jones means addition by "+"' without running afoul of the previous thesis.

[A]1. A reconstruction of some Fregean views concerning meaning

In this section I outline a few key views on meaning that came out of Frege's work. You might want to read or re-read the Frege chapter before going on. When I was a graduate student I had a philosopher of language tell me that to say one needs to read Frege is like saying one needs to breathe; Kripke is not far behind Frege in that respect. However, a reading of Frege isn't *strictly* necessary here, as my goal is not to elucidate what Frege thought but rather to describe just five important ideas that in the 1960s were commonly endorsed by philosophers of language *inspired* by Frege.

The Fregeans we have in mind start with the innocent observation that there is more to the meaning of a proper name than its referent—at least, this is true for one important sense of 'meaning'. The two names 'Superman' and 'Clark Kent', for instance, differ in many psychologically relevant ways: we associate different notions with the two names. With 'Superman' we associate 'Superhuman powers', 'Can bend steel in his bare hands', and 'Wears a brightly coloured cape'. With 'Kent' people associate 'mild mannered', 'unassuming', and 'clumsy'. Even today just about everyone agrees that in a suitably broad sense of 'meaning', the two names differ in meaning despite having the same referent (for the purposes of this essay I am pretending that they have the same real referent—pretending, that is, that the Superman/Kent story is true).

The Fregean's second step is to observe that it's intuitive that 'Lois Lane believes that Superman flies' is true while 'Lois Lane believes that Kent flies' is false. After all, if you ask her to evaluate 'Superman flies' and 'Kent flies' she will understand you and say that the first is true and the second is false.

The Fregean didn't leave it to pure intuition that (A) 'Lois believes Superman flies' is true while (B) 'Lois believes that Kent flies' is false. On my reconstruction, the third Fregean step is an *argument* that backs up the intuition mentioned in the second step. The argument could be articulated in various ways, but here is a particularly clear and perspicuous formulation of the argument for (A):

(A1) *Disquotation*: If a normal English speaker assents to a sentence '*p*', then she believes *p*.

(A2) Lois satisfies the antecedent of (A1) when the sentence in question is 'Superman flies'.

(A3) Thus, by (A1) and (A2) Lois believes that Superman flies.

(A4) If Lois believes that Superman flies, then 'Lois believes that Superman flies' is true.

(A5) Therefore, by (A3) and (A4) 'Lois believes that Superman flies' is true.

As it is formulated in premise (A1), Disquotation is implausible, since one needs to add all sorts of qualifications to it to make it reasonable. For instance, I might assent to 'Dogs are insane' while acting in a play even though I don't believe that dogs are insane. Since Lois is an ordinary English speaker who is assenting to 'Superman flies' in the ordinary kind of circumstance where those qualifications don't matter (e.g. she isn't acting in a play or being insincere or anything like that), we can omit the qualifications.

Now here's the Fregean argument for (B):

(B1) *Disquotation*: If a normal English speaker assents to '*p*', then she believes *p*.

(B2) Lois satisfies the antecedent of (B1) when the sentence in question is 'Kent doesn't fly'.

(B3) Thus, by (B1) and (B2) Lois believes that Kent doesn't fly.

(B4) *Consistency*: If someone has the belief *p* and the belief not-*p*, then she is irrational. Any pair of beliefs of the form *p* and *not-p* are called *contradictory* beliefs.

(B5) Lois is rational. (She is *confused* about 'Superman' and 'Kent', thinking that they pick out different men, but that doesn't mean she's *irrational*.)

(B6) Thus, by (B4) and (B5) Lois doesn't have the belief Kent flies as well as the belief Kent doesn't fly.

(B7) Thus, by (B3) and (B6) Lois does not have the belief that Kent flies.

(B8) If Lois doesn't believe that Kent flies, then 'Lois believes that Kent flies' is not true.

(B9) Thus, by (B7) and (B8) 'Lois believes that Kent flies' is not true.

As with Disquotation, the principle Consistency in premise (B4) would need to include a bunch of qualifications in order to be plausible. For instance, I could be rational in having beliefs *p* and not-*p* if I have them at different times. More interestingly, I could have multiple personality disorder and have those beliefs "in" different personalities. Or, I could have the pair of contradictory beliefs deep in my subconscious but have never reflected on them; in such a case the charge of irrationality might be a bit too extreme. But just as before, none of those qualifications will matter to Lois's case.

By arguing that 'Lois believes Superman flies' is true while 'Lois believes that Kent flies' is false, the Fregean has argued against the principle *Substitutivity* for belief sentences, which says that when coreferential names are interchanged (substituted for one another) in the '*p*' part of 'S believes *p*' no change in truth-value results.

The Fregean's fourth step is to notice that the phenomenon of associated notions can be used to give an *account* of the alleged fact that 'Lois believes that Superman flies' is true while 'Lois believes that Kent flies' is false. The main idea here is that if the two names 'Superman' and 'Kent' had the same meaning, as they would if the entire meaning of a name was its referent, then the two 'Lois' sentences would attribute the same belief to Lois, and so they couldn't differ in truth-value. But they do differ in truth-value. If we have the two names differ in meanings, as they would if the associated notions constituted parts of their meanings, then we can see how the two sentences can have different meanings and truth-values.

Thus, the Fregean is saying that this difference in notions associated with two names makes for a *truth-conditional* difference: a difference that generates a difference in the truth conditions for sentences like 'Lois Lane thinks that Superman/Kent flies', so that the 'Superman' sentence is true while the 'Kent' sentence is false. This goes beyond the "innocent observation" noted in the first Fregean step that we associate different notions with different names: the Fregean is saying that these associative differences make a *semantic* difference, since (she claims) they cause differences in *truth conditions*. For instance, I personally associate different ideas with 'I love catsup' and 'I love ketchup', but I don't think that this difference affects the aspects of the meanings of 'catsup' and 'ketchup' that matter to truth conditions. It's just a quirk of mine that has no semantic significance. The Fregean has argued that in cases of the kind we've been looking at the differences in associated notions *do* make semantic differences.

The fifth step is the observation that the difference in notions associated with the two names can also be used to account for how names refer to things. Fregeans called the truth-conditionally relevant notions we associate with names *senses*. They also typically thought that

senses were the meanings of *definite descriptions*, or perhaps clusters of such descriptions. For instance, just about the only thing some philosophers know about Giuseppe Peano is that he was a 19th century mathematician who was the first person to come up with the axioms for arithmetic. They associate both the indefinite description 'a 19th century mathematician' and the definite description 'the first person to come up with the axioms for arithmetic' with the name 'Giuseppe Peano'. But if the definite description gives part of the meaning of the name, as the Fregean hypothesized in the fourth step, then we can see how the name refers to the person: since the name 'Giuseppe Peano' has a meaning that includes what is expressed by 'the first person to come up with the axioms for arithmetic', and the latter refers to Peano in virtue of the fact that he alone satisfies that description, the name must refer to Peano. So, the name does its referring job via the description we associate with it.

In sum, the Fregean had several good points regarding the meanings of proper names (as well as some other terms):

1. We often associate different notions with different names, even when the names are coreferential.
2. It is intuitive that 'Lois believes Superman flies' is true while 'Lois believes that Kent flies' is false.
3. There is a good argument backing up that anti-Substitutivity intuition in (2), relying on Disquotation and Consistency.
4. We can use the differences in associated notions in order to indicate the facts in virtue of which the sentences in (2) have different truth-values.
5. If the associated notions are the meanings of definite descriptions, then we can also account for how names get linked to their referents. (The referent is just the one and only thing that is picked out by the description.)

All in all, it was a nice package of views. But as we are about to see, Kripke challenged the Fregean on (2)-(5).

[A]2. Kripke's challenge to the Fregean argument against substitutivity

I start with Kripke's alleged refutation of the conjunction of Disquotation and Consistency. Kripke's Paderewski story (1988, 130-1) seems to prove that if Disquotation is true, then Consistency is false. And if he is right about that, then of course the Fregean argument for (A) and (B) fails due to at least one false premise.

In the Paderewski story the protagonist Peter has for a long time thought that there are two famous people named 'Paderewski,' one a pianist with musical talent and one a politician without musical talent. In reality the politician is the pianist: there's just one relevant person named 'Paderewski' and Peter is just plain confused, although understandably so, since Paderewski keeps his two lives largely separate in the public eye. Here is a reconstruction of Kripke's main argument.2

(P1) Assume for the sake of argument that Disquotation is true: if a normal English speaker assents to '*p*', then she believes *p*.

(P2) Peter is a normal English speaker who assents to both 'Paderewski is musical' and 'Paderewski isn't musical'. (He assents to the first on many occasions, such as when listening to a recording of Paderewski; he assents to the second on many occasions as well, such as when discussing how politicians are usually liars without any interesting talents.)

(P3) So by (P1) and (P2) Peter believes that Paderewski is musical and that Paderewski isn't musical.

(P4) Peter is rational (like Lois Lane he is confused but still rational).

(P5) Therefore, by (P3) and (P4) someone is rational and has beliefs *p* and *not-p*.

(P6) Thus, by (P5) Consistency is false.

(P7) Thus, by (P1)-(P6) if Disquotation is true, then Consistency is false. Alternatively, if Consistency is true then Disquotation is false. In any case, they can't *both* be true.

At this point in the dialectic the Fregean can try to adjust his arguments for (A) and (B) by slightly revising Disquotation and Consistency. The alleged problem with this move is that the Kripkean argument (P1)-(P7) can be revised in response to refute the revised Fregean principles. I won't go through these moves here.

Although Kripke claims to have undermined the Fregean argument against Substitutivity for belief sentences, Kripke holds that Substitutivity fails for other kinds of sentences. For instance, Kripke would probably accept that 'It is a priori known that if Superman exists, then Superman is Superman' is true while 'It is a priori known that if Superman exists, then Superman is Kent' is

2 It should be noted that Kripke didn't set out the argument this precisely; neither did he make Consistency explicit anywhere in his long and detailed article. Nevertheless, most philosophers think Consistency did figure in his arguments in an essential way.

false (he argues for this in *Naming and Necessity*). So Kripke rejects Substitutivity for certain *epistemic* sentences (i.e. sentences about epistemological matters like a priori knowledge). However, he never explained *how* Substitutivity fails for such sentences; that is, he didn't tell us how the differences in 'Superman' and 'Kent' could create different truth conditions for the two sentences 'It is a priori known that if Superman exists, then Superman is Superman/Kent'.

*[A]**3. Kripke's challenge to semantic intuitions***

In his essay "A Puzzle About Belief" Kripke did not merely attempt to refute the Fregean argument against Substitutivity for belief sentences. He also argued that the Paderewski story presents us with a deep philosophical puzzle: although both Disquotation and Consistency are very reasonable (when the qualifications are added on), at least one of them is false (even with the qualifications added on).

Just because we might not be sure how to solve the puzzle (i.e. figure out which principle is false and why) doesn't mean that we can't draw any interesting conclusions from it. Although Kripke didn't go on to state it explicitly, I think he would say that the *main* lesson of the puzzle, before we proceed to any solution, is that we should put much less trust in our intuitions about semantic matters. The idea here would be that Disquotation and Consistency were about as intuitive as semantic principles ever get; so if we're wrong about *them*, well, then we need to take seriously the idea that our intuition could be wrong concerning just about any semantic principle. Thus, Kripke took the puzzle to have sceptical consequences for intuition-based theorizing in the philosophy of language.

This lesson is important when addressing a typical Fregean reply to Kripke's apparent refutation of the conjunction of Consistency and Disquotation. The Fregean might say that although the Fregean argument has failed, we're still left with the *highly intuitive* judgment that Substitutivity is false for belief sentences: just because (3) (from the end of section A) is out doesn't mean that (2) is gone too. Since (2) is still in place, we should still reject Substitutivity for belief sentences. Thus, Kripke's refutation of (3) has only limited consequences.

However, this Fregean reply neglects the primary lesson of Kripke's puzzle: the point about semantic intuitions. Once we have seen that the extremely intuitive Disquotation and Consistency can't both be true, why on earth should we trust our intuition that Substitutivity is false for belief sentences? The puzzle's sceptical lesson undermines (2) just as much as the refutation undermines (3), or so I think Kripke would argue.

It's worth noting that the anti-Fregean typically *accepts* the Fregean argument (A1)-(A5) for the claim that 'Lois believes Superman flies' is true. And of course that means accepting its first premise, Disquotation. As Kripke noted, 'Taken in its obvious intent [i.e. with the qualifications added on], after all, the principle appears to be a self-evident truth' (1988, 113). Thus, many anti-Fregeans have gone on to reject Consistency.

Let me try to give a brief idea of why some philosophers have rejected Consistency (other than the reason that Disquotation seems true and at least one of the two principles has to be false). You understand the name 'Aristotle' because you've read some philosophy. You may also recall that U.S. President John F. Kennedy's widow Jackie married a rich Greek guy, Aristotle Onassis. So there are two people named 'Aristotle' that you know about. In some contexts you will happily assent to 'Aristotle was a philosophical genius', as you know of great works of his; and on other occasions you will happily assent to 'Aristotle was not a philosophical genius', as you have no reason to think that Onassis was any good at philosophy, let alone a genius at it. You have something like two "files" in your mind, both with the label 'Aristotle'. In one of them you have things like 'wicked smart philosopher' and in the other you have things like 'rich guy who married Jackie'. Naturally, you think that when you say 'Aristotle was a philosophical genius' and 'Aristotle was not a philosophical genius', in appropriate contexts, you are not contradicting yourself.

Ahh, but did you know this: the rich shipping guy just is the ancient Greek philosopher! The philosopher found the secret to eternal life eons ago, and has been around for thousands of years. Just for fun, he uses the name 'Aristotle' in some centuries before faking his death and taking on a new identity. (Just pretend with me here.)

When you say things like 'Aristotle once taught Alexander the Great' you refer to the very same person you refer to when you say things like 'Aristotle once was married to JFK's widow'. Now it seems as though your two beliefs ascribed with 'Aristotle was a philosophical genius' and 'Aristotle was not a philosophical genius' really do contradict each other—even though you are rational and couldn't discover the inconsistency without empirical help. If that's right, then Consistency is false.

When a philosopher tells you that a highly intuitive principle is false, she usually goes on to try to account for the intuitiveness of the principle—often by arguing that what we are finding intuitive isn't the false principle in question but a closely related true one that we are confusing with it. That's the case here: the philosophers who reject Consistency sometimes say that a principle closely related to Consistency is untouched by these considerations. This might be such a principle: elementary reflection and memory are sufficient to know whether you attach

different notions to two uses of a name. But the thesis that elementary reflection and memory are sufficient to know whether you attach different *truth-conditionally relevant* notions to two uses of a name—notions that effect the truth conditions of belief sentences—is a different and more ambitious thesis, one that might have to be rejected along with Consistency.

It's hard to quarrel with the premises that your many assents in normal philosophical contexts to 'Aristotle was a philosophical genius' show that you do indeed believe that Aristotle was a philosophical genius, and that your many assents in normal political gossip contexts to 'Aristotle was not a philosophical genius' show that you do indeed believe that Aristotle was not a philosophical genius. Notice that when *I* write 'You believed that Aristotle was a philosophical genius and you believed that Aristotle was not a philosophical genius' I am the one using 'Aristotle' and since I know perfectly well of the identity of the philosopher and the rich shipping guy, my use is univocal. So Disquotation really is being used to attribute contradictory yet rational beliefs.

A Fregean has the resources to say that the two beliefs don't really contradict each other even though only one of them is true: the real content of the belief you have with 'Aristotle was a philosophical genius' is given by 'The guy who wrote the *Nicomachean Ethics*, taught Alexander the Great, etc. was a philosophical genius' whereas the content of the belief you express with 'Aristotle was not a philosophical genius' is given by 'The rich guy who married JFK's widow Jackie, etc. was not a philosophical genius', and those two meanings don't contradict each other in the sense of logical contradiction relevant to Consistency. One problem with this approach is that it doesn't look as though we have a decent argument or trustworthy intuition that senses are truth-conditionally relevant to belief sentences.

Now for a potential lesson regarding not meaning but linguistic understanding. If we give up on Consistency, then we are saying that a rational person could "gaze" at two of her occurrent thoughts, perhaps expressed with 'Superman flies' and 'Kent doesn't fly', and not be able to tell by reflection alone that they contradict one another. That has consequences for our view of the nature of linguistic understanding. After all, now we are saying that a person can *understand* two sentences that express her beliefs and yet she doesn't know that one is the negation of the other. Thus, understanding a sentence—or grasping a thought—has serious limitations: just because you have a thought and express it with a sentence doesn't mean that you understand its truth-conditional structure, as Lois understands 'Superman flies and Kent doesn't fly' and she thinks it has structure 'a is F and b isn't F' but really it has the structure 'a is F and a isn't F'. The truth-conditional structure of the thought is not immediately open to view—even though, as we noted above, we do seem to know, without any empirical investigation, the notions we associate with various concepts and words.

In the wake of Kripke's work on belief sentences philosophers of language have tried to figure out the logical structure of belief sentences. Often, although not always, they have accepted Substitutivity for belief sentences. For additional thoughts on this topic first study Salmon 1986 and Salmon and Soames 1988.

D. Kripke's Challenges to the Fregean Theory of Truth-Conditional Senses

Kripke's work in this area has already been competently and thoroughly presented and evaluated in many places in the literature, so I will have little to say about it. For in-depth treatment see Devitt 1980 and Soames 2002.

The central Fregean idea, which is the basis for themes (4) and (5) (from the end of section A), is this:

(*) The meaning of a name as used by a specific person is expressed by a definite description (or perhaps a group of such descriptions) in the person's mind.

Thesis (*) has a couple interesting consequences:

(**) Since a person uses 'Superman' with the meaning expressed by a definite description such as 'The superhero who flies and has x-ray vision', it follows that the meaning of 'If Superman exists, then Superman is the superhero who flies and has x-ray vision' should have the meaning had by 'If the superhero who flies and has x-ray vision exists, then the superhero who flies and has x-ray vision is the superhero who flies and has x-ray vision'. And the latter is both a necessary truth and knowable a priori. So the former is as well, as they have the same meaning.

(***) Since a person uses 'Superman' with the meaning expressed by a definite description such as 'The superhero who flies and has x-ray vision', it follows that 'Superman' in their mouth refers to whatever is referred to by 'The superhero who flies and has x-ray vision'.

Kripke offered several arguments against (*)-(***) and thus (4) and (5).

First, in many cases people understand a name and use it to refer to its standard referent even though they have no definite description in their minds that picks out the referent. For instance, many people know that Richard Feynman was a famous physicist, but they might not

know of any description that picks him out uniquely. They know he is *a* famous physicist, but they don't know anything that picks him out *uniquely.* Thus, it's not the case that each person who uses a name to refer to something has in her mind a definite description that picks that person out uniquely. Thus, (*) is false.

Second, in some cases people understand a name and use it to refer to its standard referent even though the definite descriptions they associate with the name pick out some other entity or nothing at all. Earlier in this essay I mentioned that some people associate both the indefinite description 'a 19th century mathematician' and the definite description 'the first person to come up with the axioms for arithmetic' with the name 'Giuseppe Peano'. But although Peano did come up with those axioms, he was not the *first* person to do so. That was an honest mistake people made, as it was later discovered that Richard Dedekind found the axioms first. Thus, I might use a name to refer to X even though the definite descriptions I associate with the name might actually pick out not X but Y. Thus, (***) is false. In addition, if someone uses 'the first person to come up the axioms for arithmetic' to give the meaning of 'Peano', then 'If Peano exists, then Peano is the first person to come up with the axioms for arithmetic' would have to be both necessarily true and knowable a priori, as per (**). But it's not even true, let alone necessary and a priori! Thus, (**) is false.

For the most part philosophers of language have tended to agree with Kripke that the truth-conditional part of the meaning of a name is not also expressed by a definite description (or group of such descriptions). Even so, philosophers of language have explored various Fregean options in response to Kripke's challenges. For instance, there is the question of whether the Fregean can repair matters by appealing to more clever definite descriptions, perhaps even ones that the user of the name doesn't know about consciously; alternatively, perhaps senses for proper names aren't expressed by definite descriptions but some other means (for these ideas see Evans 1973, Evans 1982, McDowell 1977, Chalmers 2002, and Jackson 1998).

E. Kripke's Positive Theses Regarding Names

Kripke has a positive view regarding names: they are "rigid designators". In order to understand how rigid and non-rigid designation works it is helpful to start with a non-rigid definite description (and then we'll move back to names). Consider 'The best male tennis player in the world in the 1990s'. As a matter of fact, this definite description denotes Pete Sampras. We can put it this way: *when evaluated at the actual world*, the bit of our language 'The best male tennis player in the world in the 1990s' refers to Sampras because he *actually is* the best male tennis player in the world in the 1990s. Now consider three possible scenarios.

> Possible world W1: Pete Sampras breaks his arm at tennis camp as a kid. This makes him hate tennis so he plays baseball instead. He's no good at it and ends up an investment banker.
>
> Possible world W2: Pete Sampras starts out as a fantastic tennis player. Then at 17 he first reads Hume's *Dialogues on Natural Religion*. He becomes a highly intelligent plumber.
>
> Possible World W3: Pete Sampras wins every Grand Slam tennis tournament in the 1990s.

When *evaluated at* W1 or W2, 'The best male tennis player in the world in the 1990s' doesn't designate Sampras (perhaps it designates Andre Agassi if he's the best male tennis player in the 1990s in those worlds); when evaluated at W3 or the actual world it does designate Sampras. That is, when we consider W1-W3 and ask who would be picked out by the meaning we actually attach to 'the best …' we don't get the same person every time. That's what we mean in saying that the definite description designates different things "in" or, perhaps better, "when evaluated at" those possible worlds; the description is non-rigid. On the other hand, 'the positive square root of nine' designates the very same thing when evaluated at every possible world; so it's rigid. And according to Kripke 'Pete Sampras' designates the very same thing in each world (e.g., it never designates Agassi); so it's rigid.

Kripke also had a positive view about how the reference of a name is transmitted from one person to the next, sometimes a bit misleadingly called the "causal theory of reference" (it's misleading because it's not a theory of reference—it's a theory of reference *transmission* from person to person—and it relies on a lot more than causality). He never developed the view into a fully-fledged theory that one can rigorously evaluate. Perhaps the best that can be said by way of introducing his view is what he said himself (for evaluation first study Stanford and Kitcher 2000 and chapter 4 of Devitt and Sterelny 1999):

> Someone, let's say, a baby, is born; his parents call him by a certain name. They talk about him to their friends. Other people meet him. Through various sorts of talk the name is spread from link to link as if by a chain. A speaker who is on the far end of this chain, who has heard about, say Richard Feynman, in the market place or elsewhere, may be referring to Richard Feynman even though he can't remember from whom he first heard of Feynman or from whom he ever heard of Feynman. He knows that Feynman is a famous physicist. A certain passage of communication reaching ultimately to the man himself does reach the speaker … [A] chain of communication going back to

> Feynman himself has been established, by virtue of his membership in a community which passed the name on from link to link … (1980, 91)

F. Kripke's Thoughts on the Metaphysics of Meaning

Kripke's book on Ludwig Wittgenstein is meant to be an *exposition* of some of Wittgenstein's primary thoughts on rules and private language. It may strike you as surprising that most philosophers have not bothered to figure out if Kripke's exposition is accurate—whether it really fits what Wittgenstein said. This lack of interest makes sense though for several reasons, the most important of which is that Kripke's Wittgenstein—often called 'Kripkenstein'—is interesting regardless of whether he's the real Wittgenstein. (Another reason: Wittgensteinians tend to be very picky, even prickly, about what Wittgenstein "really said".)

We tend to think that some facts are at a "higher level" than other facts; and the "low level" facts somehow determine the "high level" facts. For instance, the water in my teakettle just started boiling. This happened in virtue of, or because of, a whole bunch of facts about what was happening to the individual water molecules in the kettle. The high-level and macroscopic fact that my kettle started boiling right then is determined or fixed by the various low-level and microscopic facts about the water molecules in the kettle.

It's natural to think that facts about meaning are high-level. Suppose we visit an alien culture and we are trying to figure out what meanings their words have. We pick an individual alien, call him Odd, and ask ourselves 'What does Odd mean when he uses the symbol "*"?', where '*' is a symbol from his language. One would think that the answer to our question might be found in a number of places. Perhaps what he means by '*' is determined by what is going on in his brain, the various physical facts about his brain. So the low-level determining facts are facts about his brain. Or maybe the facts about how he *uses* '*'—what he applies it to, what he refuses to apply it to—fix the meaning of his uses of '*'. Or maybe some combination of those and other low-level facts fix the meaning of his uses of '*'. Let's say that Odd uses '*' to pick out property or substance P (e.g. P might be the property red, so his '*' and our 'red' are synonymous, or the substance water, in which case his '*' and our 'water' are synonyms).

Kripke's, or should I say Kripkenstein's, answer to 'What facts make it true that Odd uses "*" to express P?' seems to be 'None at all'. But interpretation is difficult here. On the one hand, Kripke writes in many places things like "[t]here can be no fact as to what I mean by 'plus', or any other word at any time" (1982, 21). He writes that there is no 'condition in either the "internal" or the "external" world' that 'constitutes my meaning addition by "plus"' (1982, 69). Again, 'no facts, no truth conditions, correspond to statements such as "Jones means addition

by '+'"' (1982, 77). Passages like those seem to make it clear that Kripkenstein is saying that *it's not a fact and not true* that I mean addition by 'plus' (or that Odd means P by '*'). And yet there are all sorts of passages that prove that such an interpretation of Kripkenstein is superficial at best:

> We do not even wish to deny the propriety of an ordinary use of the phrase 'the fact that Jones meant addition by such-and-such a symbol', and indeed such expressions do have perfectly ordinary uses. We merely wish to deny the existence of the 'superlative fact' that philosophers misleadingly attach to such ordinary forms of words, not the propriety of the forms of words themselves. (1982, 69)

> Do we not call assertions [such as 'Jones, like many of us, means addition by "+"'] 'true' or 'false'? Can we not with propriety precede such assertions with 'It is a fact that' or 'It is not a fact that'? [Wittgenstein's way with these challenges is this: we] *call* something a proposition, and hence true or false, when in our language we apply the calculus of truth functions to it. That is, it is just a primitive part of our language game, not susceptible of deeper explanation, that truth functions are applied to certain sentences. (1982, 86)

These crucial passages require us to take great care in figuring out what Kripkenstein's thesis is. For he now seems to be admitting that it really is *true* that Jones means addition by 'plus'; it's a *fact* that he means addition by 'plus'—which is precisely what it looked like he was denying above.

Perhaps Kripkenstein's thesis is something along this line: although it's true that Jones means addition by 'plus', there is nothing that *makes* that truth true. There are no underlying or low-level facts ('condition[s] in either the "internal" or the "external" world', 'superlative fact[s]') in virtue of which the Jones truth is true (that 'constitute' the Jones truth). This is unlike the teakettle case in which various facts about the water molecules underlie the high-level fact that the water is boiling. It's also inconsistent with various popular and influential theories of meaning determination put forward by Fred Dretske, Jerry Fodor, and others.

This interpretation has the merit of making a good deal of Kripke's discussion make sense, as most of his energy in the book seems to be spent arguing why various candidate lower-level facts cannot be the truthmakers for truths such as 'Odd means redness by '*''. For instance, it might be thought that what makes it the case that Odd means redness by '*' is some special mental state that he instantiates when and only when he grasps the concept of redness. But how did that mental state get the property of being a grasp of *that* concept instead of some

other concept? Again, we are looking for some lower-level facts in virtue of which the mental state in question gets its specific meaning. But there is nothing in Odd's mind or brain that shows that he meant redness rather than some closely related property such as scarlet or crimson.

There is a great deal to say about Kripke's arguments, but I want to close with an analogy meant to give a rough idea of Kripkenstein's *positive* view about the metaphysics of meaning.3 We agree on 'Cats have four feet' because we see that it's true: the agreement *follows* the truth. But 'Bishops move diagonally in chess' is true because we agree on it: the truth *follows* the agreement. So the relation between social agreement and truth is reversed. Roughly put, Kripkenstein would put meaning facts in with the chess fact and not with the cat fact. In some sense, we just decide to adopt a rule that bishops move diagonally; similarly, we just decide to adopt a certain rule regarding meaning 'Jones will mean addition by '+''. The latter case is much more complicated, and Kripke attempts to flesh it out in his Wittgenstein book.

I think Kripkenstein would agree that there are underlying facts in virtue of which 'Jones means addition by '+'' is judged by us as true: we see that certain conditions are met, and then we think that that means that the Jones sentence is true. We could discover the same thing about Odd and his linguistic community. But Kripkenstein would say that such facts aren't what *make* the 'S means X by Y' sentences true. Instead, those lower-level conditions are what we go by in *attributing* various meanings to Odd.

Bibliography

Boghossian, Paul (1989), 'The rule-following considerations'. *Mind*, 98, 507-49.

Byrne, Alex (1996), 'On misinterpreting Kripke's Wittgenstein'. *Philosophy and Phenomenological Research*, 56, 339-343.

Chalmers, David (2002), 'On sense and intension', in James Tomberlin (ed), Philosophical Perspectives 16: Language and Mind. Oxford: Blackwell, pp. 135–82.

Devitt, Michael (1981), Designation. New York: Columbia University Press.

Devitt, Michael and Sterelny, Kim (1999), Language and Reality: An Introduction to the Philosophy of Language, 2nd edition. Cambridge, MA: MIT Press.

Evans, Gareth (1973), 'The causal theory of names', *Aristotelian Society Supplementary Volume*, 47, 187–208.

Evans, Gareth (1982), The Varieties of Reference. John McDowell (ed), Oxford: Clarendon Press.

3 Boghossian 1989, Byrne 1996, Soames 1998a, and Soames 1998b are central articles devoted to working through Kripkenstein's arguments and theses; additional central essays are collected in Miller and Wright 2002. I recommend the Soames articles as the best for an initial study.

Jackson, Frank (1998), 'Reference and descriptions revisited', in James Tomberlin (ed), Philosophical Perspectives 12: Language, Mind, and Ontology. Oxford: Blackwell, pp. 201–18.

Kripke, Saul (1972/1980), 'Naming and necessity', in Davidson and Harman (eds), Semantics of Natural Language. Dordrecht: Reidel, pp. 253-355 and 763-769. Reprinted in 1980 with a new preface as Naming and Necessity. Cambridge: Harvard University Press.

Kripke, Saul (1975a), 'Outline of a theory of truth', *Journal of Philosophy*, 72, 690-716.

Kripke, Saul (1976). 'Is there a problem about substitutional quantification?' in Evans and McDowell (eds), Truth and Meaning; Essays in Semantics. Oxford: Oxford University Press, pp. 324-419.

Kripke, Saul (1977). 'Speaker reference and semantic reference', in French, Uehling, and Wettstein (eds), Contemporary Perspectives in the Philosophy of Language. Minneapolis: University of Minnesota Press, pp. 6-27.

Kripke, Saul (1979/1988), 'A puzzle about belief', in A. Margalit (ed), Meaning and Use. Dordrecht: Reidel, pp. 239-283. Reprinted in 1988 in Salmon and Soames (eds), Propositions and Attitudes. New York: Oxford University Press, pp. 102-148.

Kripke, Saul (1982), Wittgenstein On Rules and Private Language. Cambridge, MA: Harvard University Press.

Kripke, Saul (1986), 'A problem in the theory of reference: the linguistic division of labor and the social character of naming,' in Philosophy and Culture (Proceedings of the XVIIth World Congress of Philosophy). Montreal: Editions du Beffroi, Editions Montmorency, pp. 241-247.

Kripke, Saul (2005), 'Russell's notion of scope', *Mind*, 114, 1005-1037.

Kripke, Saul (2008), 'Frege's theory of sense and reference: some exegetical notes', *Theoria*, 74, 181-218.

Kripke, Saul (2009), 'Presupposition and anaphora: remarks on the formulation of the projection problem', *Linguistic Inquiry*, 40, 367-386.

McDowell, John (1977), 'On the sense and reference of a proper name', *Mind*, 84, 159-85.

Miller, Alexander and Wright, Crispin (eds) (2002), Rule Following and Meaning. Montreal: McGill-Queen's University Press.

Salmon, Nathan (1986), Frege's Puzzle. Atascadero, CA: Ridgeview Publishing Company.

Salmon, Nathan and Soames, Scott (eds) (1988), Propositions and Attitudes. New York: Oxford University Press.

Soames, Scott (1998a), 'Facts, truth conditions, and the skeptical solution to the rule-following paradox', in James Tomberlin (ed), Philosophical Perspectives, 12, Language, Mind and Ontology. Oxford: Blackwell, pp. 313–348.

Soames, Scott (1998b), 'Skepticism about meaning: indeterminacy, normativity, and the rule following paradox', *Canadian Journal of Philosophy*, Supplementary Volume 23, Ali A. Kazmi (ed), Meaning and Reference, pp. 211–249.

Soames, Scott (2002), Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity. New York: Oxford University Press.

Stanford, P. Kyle and Kitcher, Philip (2000), 'Refining the causal theory of reference for natural kind terms', *Philosophical Studies*, 97, 97-127.