

Comments on Saul Kripke's *Philosophical Troubles**

THEODORE SIDER

Disputatio 5 (forthcoming)

1. Introduction

My comments will focus on some loosely connected issues from “The First Person” and “Frege’s Theory of Sense and Reference”.¹ There is much more in these rich and rewarding papers than I’ll have time to discuss.

2. ‘I’ and having a self-concept

David Kaplan’s (1989) semantics for ‘I’ is based on the following rule:

K For any speaker, *s*, the content of ‘I’ as used by *s* is simply *s*

“Content” here means propositional contribution. Thus when I say ‘I am a philosopher’, the contribution made by ‘I’ to the proposition expressed by this sentence is simply the person Ted Sider. But Ted Sider is the content of ‘I’ only when I use that word; when Saul Kripke uses it, the content is Saul Kripke. **K** gives the general rule; it tells us what the content of ‘I’ is, as used by an arbitrary speaker, or more generally, in an arbitrary context of utterance. (Kaplan calls this rule the “character” of ‘I’.)

In later work² Kaplan stresses certain aspects of this account of the semantics of ‘I’. First, instead of trying to give a synonym for ‘I’, it describes the rules governing the use of ‘I’. Second, the description of these rules is a “description from above”, a description from an external, austere perspective. The rules of use for contextual expressions can themselves be understood without possessing any particular bit of contextual information. (Similarly, the rules for use for expressive terms can be understood with a “cool head”—without oneself expressing anything.)

*Given at the December 2013 meeting of the Eastern Division of the American Philosophical Association. Thanks to David Braun and David Chalmers for helpful feedback.

¹Kripke (2011).

²The unpublished “The meaning of Ouch and Oops...”/ “What is meaning?”.

It's natural to take rule **K** as giving a complete account of the semantics of 'I'. But Kripke gives a fascinating argument against this claim. First, he argues that Kaplan's idea of giving the semantics of expressions "from above" has an "instructional" aspect: one ought to be able to use the rules to learn the language. Second, he argues that if someone doesn't yet understand the concept of 'I', you can't teach it to her by telling her the rule **K**:

...recall my remarks that the 'description from above' ought to be usable as an instruction manual for someone wishing to learn the language. Though Kaplan's explanation is all very well for some sort of descriptive anthropologist who may in fact have the concept of 'I', it would be very difficult to get it across to Frege (or anyone else who is presumed to lack this concept). So, for example, let Kaplan say to Frege or to anyone else (but if it is Frege, one should use German): 'If any person *s* speaking German attributes a property using the word "ich," then what *s* says or thinks is true if and only if *s* has that property.' But how can Frege use the word 'ich' on the basis of these instructions? Should he think, 'Hmm, so how am I going to use the word "ich" on the basis of this general statement? Well, any German should attribute, say, being in pain or being a logician to himself if and only if the German is in pain or is a logician, as Kaplan says. So I should do this.' Alternatively, Frege might remark, 'So Frege, or Dr. Gustav Lauben, should attribute a property to Frege, or respectively to Dr. Lauben, using "ich" if and only if Frege (or Dr. Lauben) has the property. But I am Frege, so I suppose that I should use the word "ich" if and only if Frege has the property.' Either formulation would presuppose that Frege already has the concept of himself, the concept he expresses using 'ich,' so here we really are going in a circle. (Kripke, 2011, 300-1)

To bring this point home, imagine beings who entirely lack the capacity for first-person thought; they lack "self-concepts". If they were then informed of rule **K**, they would merely learn that each person *s* refers to *s* when using 'I'. This wouldn't give them self-concepts.

Are such beings really possible? Perhaps. David Lewis (1979) imagines two gods, each of whom knows all the "impersonal" facts but is ignorant as to which of the two gods she is. The idea presumably is that these gods somehow have knowledge of the world that isn't from any particular perspective. Perhaps such beings could be imagined, not merely to be ignorant who they are, but to lack a concept of themselves altogether. Or perhaps we could imagine more mundane

creatures who lack a self-concept because of disinterest in themselves.³ Unlike the gods, these latter beings do experience the world from particular spatial perspectives; and they do have the ordinary concept of a person. But each of these beings is completely uninterested in herself, and has only impersonal desires. They conceive of the world impersonally. They think “Something red is nearby” or “someone is seeing red”, but never “I am seeing red”. They have general desires, such as the desire that people be happy; but they don’t desire things on behalf of themselves. Perhaps we can coherently imagine that such beings lack a concept of the self.⁴

I think that Kripke’s argument establishes something important and correct. It shows that Kaplan’s semantics for ‘I’ doesn’t answer one very important philosophical question about the self, namely, the question of what it takes to have a self-concept—to be capable of first-person thought. For as Kripke argues, a being that lacked a self-concept wouldn’t gain one by learning rule **K**.

But although the argument does establish this conclusion in the philosophy of mind, it’s less clear that the argument establishes the conclusion in the philosophy of language that Kaplan’s rule is incomplete as a semantics for ‘I’. For one might hold that selfless beings who master rule **K** are semantically competent with respect to ‘I’. Their deficiency is psychological, not semantic. They understand ‘I’ just fine; it’s just that, given their odd psychologies, they can’t use ‘I’ in quite the same way as ordinary English speakers.

If you cannot see or hear, then in some circumstances you won’t be able to use words like ‘car’ and ‘tree’ effectively. When cars and trees can only be identified visually or auditorally, you sometimes won’t know when to use ‘car’ and ‘tree’. If you are very bad at telling whether people are angry, then your use of ‘angry’ will similarly be compromised. But your *semantic* proficiency with ‘car’,

³Compare Perry’s (1986) discussion of the shark.

⁴It may be argued that the latter beings would be *capable* of introducing a self-concept, by making use of their perspectival experiences. One might say to one of them “there is a person that is always nearby (so to speak), who is always visible in mirrors held at such-and-such angles, etc.; that person is *you!*”. Whether this really would introduce a self-concept might depend on certain aspects of the example that I haven’t specified: for instance, on what it is like for these creatures when they intentionally move their bodies. But in any case, it isn’t important to the argument that the creatures are incapable of gaining a self-concept; all that we need is that they in fact don’t have a self-concept. (Perhaps someone once taught their predecessors a self-concept, but it never caught on because they were uninterested—after all, they don’t particularly care about the people at the “centers” of their perspectives.) For even if they’re capable of gaining a self-concept in such a roundabout way, learning Kaplan’s rule for ‘I’ presumably won’t give the concept to them.

‘tree’ and ‘angry’ would not be compromised. More extremely, compare beings who use proper names the way we do, but have radically different sensory systems, so that the way they make use of the apparatus of naming—the way they initially introduce names, the way they pass them along to others, and the way they identify their bearers—is very different from ours. Their language has the same semantic apparatus as ours; they just make use of the apparatus differently.

So: the fact that a person, or a group of people, exercise different perceptual or psychological capacities when using a term, even with the result that their use of the term is in some ways limited, is consistent with the hypothesis that they mean exactly what we mean by it, and are fully semantically competent with it. And so, we might say that selfless beings who master rule **K** mean what we mean by ‘I’, and are semantically competent with that word; it’s just that their lack of a self-concept prevents them from using this semantic apparatus in exactly the way that we do.

It’s worth stressing here that the selfless beings’ use of ‘I’ would not be completely compromised. Kripke’s argument stresses the circumstances in which it would indeed be compromised: the selfless beings wouldn’t know when to utter ‘I’ sentences.⁵ But the selfless beings would be able to *understand* ‘I’ sentences just fine. They would understand, for instance, that when I say ‘I am a philosopher’, I mean that Ted Sider is a philosopher. Thus the selfless beings’ difficulties with using ‘I’ are localized (albeit significant and robust). They have a psychological shortcoming which renders them unable in certain situations to recognize when the word should be used.

One might worry that the question of what to classify as semantic is merely verbal. I don’t think so (though there are difficult issues here). For one thing, the commonalities between us and the selfless beings when it comes to understanding others’ uses of ‘I’ must be recognized at some level of theorizing, regardless of what we call that level. For another, counting the selfless beings as being semantically competent with respect to ‘I’ fits with a certain tradition of thinking of the semantic as a relatively “objective” level of analysis, which generally abstracts away from the psychological idiosyncracies of individual speakers.

⁵Well, Sally, one of these beings, would know that Sally ought to say ‘I am happy’ to express the proposition that Sally is happy. But she wouldn’t know to say ‘I am happy’ to express the proposition that Sally is happy, since she doesn’t know that she herself is Sally.

3. Achieving self-reference

I have been saying that a semantic account of ‘I’ needn’t shed light on the question of what it is to have a self-concept. Let me say just a couple words about Kripke’s views about that latter question—or rather, on the closely related question of what it takes to refer to oneself.⁶

Suppose I want to refer to myself. How do I do it? Kaplan and John Perry (1979) argued against the idea that we do it by description. Ordinary descriptions might turn out not to be unique, or might even fail to apply to me; and it seems unlikely that I have some “primitive aspect” of which I am aware by introspection. Kaplan then supplied his rule **K**. But as Kripke’s argument shows, possession of that rule doesn’t put one in a position to think about oneself first-personally, and so does not put one in a position to use ‘I’ to refer to oneself.

Kripke stresses the importance of a special sort of acquaintance we have with ourselves, and says that we use it to achieve self-reference.⁷ He’s right, I think, to insist that such a relation needn’t be conceived descriptively, but rather could simply be a direct psychological relation each person bears to him or herself. Alternatively—though Kripke doesn’t say this—one might say that one bears this relation to one’s mental states.

It’s perhaps worth thinking a bit about how exactly we could use self-acquaintance to achieve self-reference. Recall Kripke’s point that if someone lacked a self-concept, we couldn’t give one to her by teaching her Kaplan’s rule **K**. One might make a parallel point that such a person couldn’t achieve self-reference by employing the description “the being to which *I* bear Kripke’s relation of acquaintance”—the description presupposes ‘I’.

But when Kripke says that we achieve self-reference by self-acquaintance, I don’t take him to mean that I pick out myself as “the being with which I am self-acquainted”. Indeed, I don’t think he thinks that there is any way to define ‘I’ using only nonindexical concepts plus his relation of acquaintance. (Compare his rejection of the need to use the “scientific language” when giving meanings.) Rather, I take it, the idea is that the fact that I am self-acquainted explains my ability to use the word ‘I’.

If Kripke agreed with my earlier claims about the separability of the seman-

⁶I mean this in the “de se” sense.

⁷He says things like this in a few places, though most of them are in his discussion of Frege. So I’m a bit fearful that I am misinterpreting when I attribute this view to Kripke himself. If I am misinterpreting, I’m sorry!

tics of 'I' from the psychological abilities needed to use it, he might put the point as follows. The selfless beings who master rule **K**, and thus understand 'I', nevertheless can't use the term 'I' properly to talk about themselves. What they lack is acquaintance with themselves. If they came to bear that relation to themselves, then they would come to be able to use 'I' sentences in new ways, for they would then be capable of recognizing when they themselves had certain features, and then could use the sentence 'I have such and such features' appropriately.

I'd like also to briefly mention something else Kripke says:

...the first person use of 'I' of course does not have a Fregean sense, at least if this means that it has a definition. But it might be a paradigmatic case, one that I did not mention in *Naming and Necessity* (1980), of fixing a reference by means of a description: it is a rule of the common language that each of us fixes the reference of 'I' by the description 'the subject'. However, since each of us speaks a natural language, and not an imaginary 'scientific language' spoken by no one, for each of us the referent can be different. (p. 304)

Now, 'the subject' doesn't pick out one thing uniquely, not by its conventional meaning anyway. So what does Kripke have in mind here?

A bit later on the page we read this:

For it follows from the way I fix the reference, as the subject of my own thought, that I must exist.

So perhaps the suggestion is that the description is not just 'the subject', but rather a fuller description that brings in the speaker's thoughts. But which description, exactly? Surely not 'the subject of *my* thoughts'.

One possible answer runs as follows. The description we use to fix the reference of 'I' is: 'the subject of *these* thoughts'. 'These' is a demonstrative. Now, a demonstrative can have its reference secured by a relation of acquaintance. That is not to say that the demonstrative is replaceable by some description containing 'acquaintance'; it is rather to say that the speaker's reference to something when using the demonstrative is explained by the fact that she bears a relation of acquaintance to that thing. In the case of 'these thoughts', the relation of acquaintance explaining the reference is Kripke's relation of self-acquaintance, only now that relation is understood as holding between a person and her thoughts rather than between a person and herself.

Is this proposal what Kripke had in mind? I'd be interested in hearing his thoughts here.

4. Frege's hierarchy

In "Frege's Theory of Sense and Reference" Kripke makes a fascinating suggestion about Frege's hierarchy of indirect senses. Frege accepts the following:

- (β) When words appear in indirect contexts, that is, "says that", "believes that", and so on, they refer to their senses in the clause following the "that".

Thus the referent of a word in an indirect context is not its customary referent, but rather its customary sense. But then, what is the sense of a word in an indirect context? Not its customary sense: each sense determines a unique referent for Frege, and the referent determined by the customary sense is the customary referent. So it would seem that a competent speaker must now memorize two senses for each word, the customary and the indirect sense.

Moreover, consider doubly indirect contexts—indirect contexts inside indirect contexts. An argument like the one before can be given for the conclusion that each word has a distinct third sense. Triply indirect contexts then lead to a further sense, quadruply indirect contexts to yet another sense, and so on. This is the problem of Frege's hierarchy.

Kripke's first point about the hierarchy problem is that there is a parallel situation with direct quotation that no one has thought to be problematic. Frege (arguably) accepts a principle parallel to (β) for direct quotation:

- (α) When words appear in direct quotes, they refer autonomously, that is, to themselves

Thus in direct quotation—inside quotation marks—words don't refer to their customary referents. But that means that in such contexts they cannot have their customary senses. Moreover, direct quotation can occur inside direct quotation. So is there also a problem of a hierarchy of senses for direct quotation?

Here is the core of Kripke's solution to both problems. (α) and (β) provide general rules for determining the reference of expressions in direct quotation and indirect contexts. But, as Kripke puts it, "whatever determines a reference is a Fregean sense" (p. 269). So (α) and (β) themselves give the sense of expressions in direct quotation and indirect contexts. Moreover, (α) and (β) can be applied recursively, and so give the entire hierarchies of senses.

This strikes me as a powerful and plausible idea about Frege, but there are a few things I'd like to look at a little more closely. Suppose I use rule

(β) to determine that the reference of ‘grass is green’ in ‘Jones believes that grass is green’ is G , the customary sense of ‘grass is green’. Reference has been determined, and so Kripke says, we have a Fregean sense. But there is a question of exactly which sense this is. Specifically, there is a question of how much of the procedure I followed gets built into the sense.

After all, part of the procedure that (β) specifies for determining reference is linguistic in nature, and Kripke is clearly not saying that these elements are to be built into indirect senses.⁸ When following (β), I observed that ‘grass is green’ occurred in an indirect context in the sentence of interest, and inferred that its referent in this case is its customary sense. But this part of the procedure involves linguistic notions—the concept of a word, the concept of an indirect context, the concept of the clause following ‘that’, and so forth—and when I say that Jones believes that grass is green, I am not expressing a thought that involves the concepts of words, clauses, indirect contexts, and so forth.

So: not every part of rule (β)—and, more generally, not every part of what we do when we compute the reference of a linguistic expression in an indirect context—gets built into the indirect sense. So which parts get built in?

Here is one possibility. The full procedure associated with (β) (and rejected above as the indirect sense of ‘grass is green’), delivered as its “output” G , the customary sense of ‘grass is green’; it did so in part via linguistic considerations. We could consider a more austere procedure that also delivers G as output, but not via linguistic considerations; and indeed, not via any considerations at all. Such a procedure is the “constant” procedure whose output is G . When I follow rule (β) and determine that G is the referent of ‘grass is green’ in ‘Jones believes that grass is green’, I can be regarded as also following this more austere procedure as well, in addition to the full, partly linguistic one. This more austere procedure—this way of determining a referent—is another candidate for the indirect sense of ‘grass is green’. That is, of all we do when we follow (β), we might build into the sense only something very minimal: the outputting of G .

It may help to be a little more concrete about what exactly these senses, these procedures or ways for determining referents, are. Let’s think of them as properties that can be instantiated by at most one thing. (On this conception, the referent that is presented by a sense is simply the object, if any, that instantiates that sense.) Thus construed, the linguistic indirect sense discussed (and

⁸Though it may be that the corresponding linguistic elements *are* intended to be built into the senses in contexts of direct quotation.

rejected) a bit ago might be regarded as the property of *being the sense of the words following 'that' in the indirect context 'Jones believes that grass is green'*, and the austere indirect sense might be regarded as the property of *being identical to G*.

We can bring out another conception of the nature of indirect senses by considering the following quotation from Kripke's paper:

However, there is the question of how an individual speaker will apply (β). My suggestion, once again is that Frege, like Russell, has a doctrine of direct acquaintance. Every time we determine a referent, we are introspectively acquainted with how the referent is determined, and that is the corresponding sense. And our introspective acquaintance with this sense gives us a way of determining it, and of referring to it, and this is the indirect sense. Thus the Fregean hierarchy of indirect senses, doubly indirect senses, and the like is given this way. Each level of the hierarchy is the acquaintance-sense of the previous level. (2011, pp. 271–2)

Now, one thing this passage is saying is that when we use, or apply, rule (β), we exploit the fact that we bear a certain relation of acquaintance to senses. (β) specifies, for instance, that G is the referent of 'Grass is green' in an indirect context; but if I am to *use* this rule to know what the referent of 'Grass is green' is, I need to be acquainted with G . So acquaintance is part of the total procedure we use, when we follow (β) to determine the reference in indirect speech. But does anything about acquaintance get built into indirect senses?

Some parts of the quoted passage could be read as suggesting that acquaintance does indeed get built in ("Each level of the hierarchy is the acquaintance-sense of the previous level", and "our introspective acquaintance with this sense gives us a way of determining it, and of referring to it, and this is the indirect sense"). On this view, in the example of 'grass is green', for instance, the indirect sense involves acquaintance with G . Acquaintance-senses do not build in linguistic information (such as the notion of an indirect context), but they do build in more than the austere senses do: they build in something about acquaintance.

Note that one could agree with the first part of what Kripke says in the quoted passage, that Frege relies on a doctrine of acquaintance, while not agreeing that anything about acquaintance is to be built into indirect senses. For one might hold that indirect senses are the austere ones mentioned above—the indirect sense of 'grass is green', for instance, is the "constant" rule delivering G , i.e. the property of being identical to G —while conceding to Kripke that

in order to be able to follow this austere rule—in order to grasp the indirect sense—one must be acquainted with G .

So one question of clarification I'd like to ask is whether Kripke intended acquaintance to be built into indirect senses, or whether instead he had in mind the more austere conception of indirect senses.

If the former is the intended interpretation of Frege, I have a followup question about the nature of acquaintance-senses. Let's continue to think of senses as properties—properties concerning the way in which the referent is determined. How exactly is acquaintance built into these properties? There seem to be two possibilities here.

One is that for any speaker s , the indirect sense of 'grass is green' is the property *being an object that is identical to G and with which s is acquainted*. On this conception, different people associate different acquaintance-senses with words in indirect discourse; the acquaintance-senses differ because each one builds in a particular person's acquaintance-relations, so to speak. Given this conception, different people express different thoughts by "Jones believes that snow is white" and other sentences involving indirect discourse. (Given Kripke's interpretation of Frege's view of 'I' and 'now', this needn't be regarded as immediately disqualifying this interpretation of Fregean indirect senses.)

A second possibility is that the indirect sense doesn't build in anything specific to the speaker; it just—intuitively—involves the *general* method of determining the customary sense by acquaintance. A natural way to model this would be to think of an indirect sense as no longer being a property of the customary sense, but rather as a relation between speakers and customary senses. For then we could think of the indirect sense of 'grass is green' as being the relation born by an arbitrary speaker s to x iff x is identical to G and s is acquainted with x . *Each* person who grasps G by acquaintance bears this relation to G , and thus we can think of this relation as being the way in which each such person determines G for herself.

On this second possibility, acquaintance-senses would be admittedly unlike familiar Fregean senses, since they are relations to speakers. Nevertheless, this second possibility meshes with the idea that indirect senses involve a speaker's determining reference to a sense by bearing a direct cognitive relation to that sense. The more usual Fregean model of the determination of reference, after all, is that a speaker determines the reference by employing a property that determines the referent. Given this usual model, the sense can be the property. But if reference to a sense is determined instead by a direct relation of acquaintance to the sense, we've departed from this model. Since there is no

employed property to identify with the sense, the only property available to be the sense is the first possibility just mentioned: a property that builds in the speaker herself. But that has its own disadvantages (different indirect senses for different speakers), and so we might prefer the second possibility, in which each speaker uses the same “centered” property—i.e. relation—to determine the referent.

This second possibility is a move in the direction of “two-dimensionalism”, and so it inherits some of the issues of that tradition. If indirect senses are relations rather than properties, they do not present their referents absolutely, but only relative to speakers. Thus they have truth values only relative to speakers, and so the view will need to say something a little subtle for sentences like “Everything Jones believes is true”. But notice something about the proposed indirect sense for ‘grass is green’, namely, the relation of being an s and x such that $x = G$ and s is acquainted with x . Granted, this indirect sense does not present its referent G absolutely, only relative to a speaker s who is acquainted with G . But it never presents any customary sense *other* than G with respect to any other speaker, and so G can be associated with it as a sort of absolute referent, which may be enough.

To sum up. If indirect senses have acquaintance built in, and are conceived in the second way (“two-dimensionally”), the hierarchy for ‘grass is green’ looks like this:

- Customary sense: G
- First indirect sense: the relation, r_1 , of being an s and x such that $x = G$ and s is acquainted with x
- Second indirect sense: the relation, r_2 , of being an s and x such that $x = r_1$ and s is acquainted with x
- etc.

If indirect senses have acquaintance built in, and are conceived in the first way, the hierarchy for ‘grass is green’, for a speaker A , looks like this:

- Customary sense: G
- First indirect sense: the property, p_1 , of being an x such that $x = G$ and A is acquainted with x
- Second indirect sense: the property, p_2 , of being an x such that $x = p_1$ and A is acquainted with x
- etc.

And if indirect senses are austere, the hierarchy looks like this:

Customary sense: G
First indirect sense: the property, p_1 , of being an x such that $x = G$
Second indirect sense: the property, p_2 , of being an x such that $x = p_1$
etc.

My question here is simply: which of these hierarchies is best, and which is best to attribute to Frege?

References

- Kaplan, David (1989). "Demonstratives." In Joseph Almog, John Perry and Howard Wettstein (eds.), *Themes from Kaplan*, 481–563. New York: Oxford University Press.
- Kripke, Saul A. (2011). *Philosophical Troubles. Collected Papers Vol I*. Oxford University Press.
- Lewis, David (1979). "Attitudes *De Dicto* and *De Se*." *Philosophical Review* 88: 513–43. Reprinted in Lewis 1983: 133–56.
- (1983). *Philosophical Papers, Volume 1*. Oxford: Oxford University Press.
- Perry, John (1979). "The Problem of the Essential Indexical." *Noûs* 13: 26–49. Reprinted in Perry 1993: 3–25.
- (1986). "Perception, Action, and the Structure of Believing." In Richard E. Grandy and Richard Warner (eds.), *Philosophical Grounds of Rationality: Intentions, Categories, Ends*, 333–61. Oxford: Clarendon Press. Reprinted in Perry 1993: 121–49.
- (1993). *The Problem of the Essential Indexical and Other Essays*. New York: Oxford University Press.