Meaning, Reference and the Answering Machine

I. Intensions--Frege and Montague
II. Extensions--Russell and Kripke
III. Character and content--Kaplan
IV. The answering machine paradox

What I want to do today is to trace the development of at least one strand of philosophical semantics, the strand that begins roughly with Gottlob Frege, and continues through David Kaplan. Kaplan is a philosopher at UCLA, not that much older than me, so I mean to be bringing you up to the present. In fact I want to end with a little puzzle that might suggest that the current philosophical semantics is inadequate, or at least incomplete, so perhaps you might say that I bringing you more than up to date.

Gottlob Frege, was a mathematician and philosopher who taught at Jena, in what is now East Germany, in the late nineteenth and early twentieth century. Frege's thought is enormously influential among philosophers today, thanks largely to Russell, Husserl and Wittgenstein who were among the few who took him seriously initially. German philosophers are notorious for building grand and obscure systems in which everything of importance is supposed to fit, but Frege spent his whole life writing with great clarity and precision about a couple of seemingly tiny, very technical questions. The question that led to the material that I want to talk about had to do with identity statements. In particular, Frege was vexed by the question of how an identity statement, like "The Morning Star is identical to the Evening Star" could be both true and cognitively significant. There are three possible ways one might try to interpret of identity statements like this. First, one might say that they express an identity between two expressions: "The Morning Star"="The Evening Star". But in that case the identity would just be false. These expressions are not the same. The one on the left has an 'm' but no v; the one on the right has a v but no m. Second, one might say that they express an identity between objects, in this case the object (Venus) denoted by the left expression is the same as the object (Venus) denoted by the right expression (Venus). But if THAT were the correct interpretation, the sentence would be rather trivial. It would have, as Frege says, the same cognitive significance as the claim that The Morning Star is the the Morning Star. The third possible interpretation is slightly more sophisticated. The sentence does not express an identity between the names or the objects named, but rather it says that the two expressions are CO-DENOTING, i.e., that they both name the same object. But if this were correct then the statement the that Morning Star is identical to the evening star would just be a statement about the way we use words. In fact,
however, it was a great DISCOVERY of astronomy that the morning star and the evening star were, in fact, one and the same planet.

Frege's solution to this puzzle was to postulate the existence of something that mediates between the linguistic expression and the object it names. He labels this extra something the "Sinn" or "sense" of the expression. The sense, as Frege, says contains "the means of presentation" of the object. The morning star is Venus, as presented in a certain part of the sky in the morning and the evening star is Venus, as presented in some other part of the sky in the evening.

So what Frege did, in response to this seemingly tiny puzzle about identity, was to say that the "meaning" of an expression like "the morning star" is not some simple thing, but rather that it has at least two components, the reference--which is a piece of rock somewhere in our solar system, and the sense--which is something that somehow presents the reference to us or at least allows us to get at the reference. But, Frege went on to say, there is nothing special about expressions like "the morning star". Any expression in natural language can have a sense and a reference. In particular, a declarative SENTENCE, can have both a sense and a reference. Now we all learned in elementary school that a declarative sentence is an expression that "expresses" a complete thought. So we might have expected the referent of a sentence to be the thought that it expresses. But Frege has several persuasive arguments to show that, in fact, the reference of a sentence is its TRUTH VALUE. The thought expressed corresponds more closely to the sense. Just as the sense of the expression "the morning star" enables us to get to a certain piece of rock and gas, so the sense of "It's raining" enables us to get to a certain truth value. In addition to postulating the existence of these two components of meaning, Frege also postulated a principle that, although it might seem obvious, has been very important in semantics. We might call this principle "compositionality". The sense of an expression is a function of the senses of its parts and the reference of an expression is a function of the references of its parts.

Now let's leave Frege and talk for a minute about modal logic. Modal logic, narrowly conceived, is the study of principles of reasoning about necessity. Aristotle and the medieval philosophers were quite interested in this topic and discussed a number of such principles. What is necessary is not possibly not, what is possible is not necessarily false, If it is necessary that both A and B then it is necessary that A and that B. And so on. This area of logic was neglected until the twenties or so, when the philosopher CI Lewis noticed a connection between necessity and certain defects he perceived in the standard treatment of conditionals. Lewis formulated a number of different axiomatic systems for necessity, but he didn't really know which one was the correct system. Since he didn't have any semantics for these he really didn't have any good idea what could
and could not be proved in each of the systems. In fact, he didn't even know for sure that his systems were really distinct. So during the forties and fifties people had a great deal of fun proving that this formula could or could not be proved from this system or that system. All this came to an end in 1959 or so, when Kripke and others provided a semantics for these systems which made the question of whether a formula could be proved completely trivial. In the simplest version, a Kripke model for modal logic is just a set W of possible worlds and a valuation V that tells you whether the atomic sentences are true or false. Necessary A is true if A is true in all possible worlds.

Now in the years since the introduction of Kripke models, people have come to realize that the notion of a "possible world" is an extremely useful notion in all kinds of contexts. For our purposes today, it is useful because it gives us something more precise to play the role of Frege's rather mysterious "senses". In particular, we can define the intension of a sentence to be a function from possible worlds to truth values. For example the intension of "the cat is on the mat" is the function that assigns the value "true" to each world in which the cat is on the mat and that assigns the value "false" to each world in which the cat is not on the mat. The intension of a definite description is a function from possible worlds to individuals. So the "intension" of "the tallest mountain on earth" is the function which to this world assigns Everest (or, my son tells me, K2. I am not sure about this.) But to some other worlds it might assign the Matterhorn, and to still more remote worlds, it might assign the hill on which our Observatory sits. Now the intension is something like Frege's sense. It provides a means of picking out the reference or extension.

You might have noticed that have talked about expressions that refer to individuals and those that refer to truth values, but not about any other expressions. One of Montague's insights, I think, is that we don't really need to talk about any other kinds of expressions if we take Frege's principle of compositionality seriously. Take some English sentence. Say, "John loves Mary". We know "John" refers to an individual and we know the whole sentence refers to a truth value. What then does "loves Mary" refer to? Well, by the principle of compositionality, it refers to something that can be combined with an individual to produce a truth value. The simplest object that can do this is a function from individuals to truth values. So why not just say that this is the reference of "loves Mary"? In particular "loves Mary" denotes the function that assigns "true" to an object x if x loves Mary, and that assigns "false" to that object otherwise. OK, so what is the denotation of "loves". Well, we know "mary" denotes and individual, so by compositionality "loves" must denote something that combines with individuals to produce a function from individuals to truth values. So we can say that "loves" denotes a function from individuals to functions from individuals to truth values. The details here sound complicated, but the principle is as simple as it can be. Names denote individuals, sentences denote truth values, and if an expression that denotes A's
with some other expression to form an expression that denotes B's then that other expression just
denotes a function from A's to B's. Furthermore, if an expression has A's as its DENOTATIONS
it has functions from possible worlds to A's as its SENSES.

It turns out that the kind of interpretation Montague is suggesting is really equivalent to the
kind of interpretation that had been given to the artificial languages of logic all along. So Montague
was suggesting that natural languages were really just very elaborate varieties of formal languages.
He wrote a series of papers through the sixties in which he treated increasingly large "fragments" of
English in exactly this way. "Montague grammar", as we call it, is just a working out of the details
of the Fregean thesis that the sense and reference of an expression are determined by the senses and
references of their constituents.

Frege's view that every meaningful expression of a language has a sense, or Montague's
version that every such expression has an intension, runs counter to an older view about some special
expressions--namely proper names. John Stuart Mill had said that proper names have denotations,
but not connotations. This sounds an awful lot like the view that proper names have referents but
no senses. And initially Bertrand Russell said exactly that--proper names have no Fregean sense.
There is a difficulty with this position, however. If proper names have no sense, how do we figure
out what is being referred to by somebody who uses a proper name? Russell ended up by saying that
most ordinary proper names are really just abbreviations for descriptions--'Aristotle', for example
is just shorthand for 'the Greek philosopher who studied with Plato and taught Alexander the Great'.
That is because we are not, in Russell's terminology, "directly acquainted" with Aristotle. We know
him only by description. In fact because of his more general philosophical views, Russell eventually
came to believe that we really aren't directly acquainted with anything but our sense data.
Consequently, he came to believe that the only words that are genuine proper names are words like
"this", referring to a patch of color that I am currently seeing.

So now we have two views about proper names. Frege says that, like every expression, a proper
name has both a sense and a reference. Mill and Russell say that a description has a sense and
reference, but that a proper name, or at least a genuine proper name, has only a reference. How do
we decide who is right? Well Kripke, in a now famous set of lectures called Naming and Necessity
has given us a lot of evidence for thinking that Mill was right and Frege was wrong. Consider the
name "Aristotle". If Frege is right, this name picks out Aristotle in our world, but it might pick out
somebody else in another world. Suppose, as Russell might have said, it picks out in each world the
guy who studied with Plato and taught Alexander. In that case the fact that Aristotle studied with
Plato would be a necessary truth. Yet it seems pretty obvious that it was just a historical accident
that Aristotle studied with Plato. Had things worked out differently, he MIGHT have studied with the sophist Thrasy machus, or perhaps with one of Plato's brothers, or he might just have joined the army and been killed in battle. And this is just one part of the evidence. Aristotle is a guy that most of us in this culture know something about. Consider some more slightly obscure people -- Cicero, Schleiermacher, Al Kaline, Hulk Hogan, Artur Rubenstein. I know for example, that Cicero was a famous Roman orator, I know that Al Kaline played outfield for the Detroit Tigers, that Hulk Hogan is or was a professional wrestler. But I don't know enough about any of these people to identify them uniquely. Now if, as Frege thought, the sense of a name is supposed to be something that enables me to fix the referent, it would seem that I don't have the required information to do that.

Considerations like these led Kripke to believe that proper names were, as he said, "rigid designators". "Socrates" picks out the same guy in every possible world. Which guy does it pick out? Well, it's probably the guy who hung out at the Agora in Athens and inspired Plato and was put to death by Athenian democracy. But it MIGHT not be. Suppose the guy who claimed to be Socrates in Athens was an imposter, and the real Socrates, i.e., the guy whose mother named him "Socrates" spent his life as a shopkeeper in downtown Athens. Then it would not be the case that Socrates was put to death; rather the guy that we falsely think is Socrates would have been so put to death.

It might seem that proper names are not a very important part of language. In fact there may be some who would say that they aren't a part of the language at all. After all, they are omitted, for the most part at least, from dictionaries. But in recent years the phenomenon that Kripke pointed to in connection with proper names has been seen to infect other parts of language as well. Kripke himself suggested that natural kind terms, like "gold" or "water" or "human being" are, like proper names, rigid designators. And David Kaplan has argued convincingly that the so called "indexicals"--I, you, here, now, today, tomorrow, yesterday, and the "demonstraves"--this such-and-such, that so-and-so also lack a Fregean sense.

This last claim might seem rather fantastical, if you believe, as Frege seemed to, that the descriptive meaning of an expression is something like its sense. The word "I" clearly has a descriptive meaning. It means "the person who is speaking" or "the person who is writing" or in general, what Kaplan calls "the agent". Similarly, "now" means "the present time" and "here" means the location of the agent. But notice that the intension of "I" can not be a function that assigns to world w, the person who is speaking in w. For one thing there is no such unique person. In this world, for example, I am speaking, but so are some of my colleagues in the philosophy department, and so, I suspect are some people in Virginia, and Maryland, and some people on twenty or thirty
television stations and on a few thousand radio stations. I suspect that there are many millions of people speaking right now. Even if we COULD somehow identify the agent, the agent's location and the present time of a possible world, it would not do to identify the sense of the indexicals with the function that picks these features out. For if we did that the sentence "I am here now" would be true in all possible worlds, i.e., it would be necessary. Yet clearly this sentence does NOT express a necessary truth. I could be in at home, sleeping in bed. I could be working in my library carrell. I could be vacationing in Hawaii. It is not necessary that I be here now. On the other hand, there is SOMETHING obviously trivial about the claim "I am here now". When you hear it, you know its true apriori--You don't have to open your eyes and look at me and say--"yes, he's not lying. He really is there now." And it would seem to be important for a theory of language to explain both that "I am here now" is not necessary, and that it is, in some other sense, obviously true.

Kaplan does this by postulating that meaning has even more than the two components identified by Frege. Frege say that every expression has a sense and a referent. We can identify the sense with a function from possible worlds to referents, or as Kaplan prefers to say, a function between possible circumstances of evaluation and referents. Kaplan identifies this component of meaning with what he calls the "content" of an expression. It tells us "what has been said" when the expression is uttered. For example, if I say "I am here", the content is a function that assigns true to every world where Steve Kuhn is in the ICC and false to every world where he isn't. Notice that this same content could be expressed by Bill's saying "You are here", or maybe in some other context by somebody else's saying "He is there" with an appropriate kind of demonstration. So the referent of "I am here" is "true" and the content is a function from possible worlds to truth values. But there is more to meaning than reference and content. There is what Kaplan calls the CHARACTER. The character of an expression is a function from possible contexts of use to contents.

We have a picture like the following

<table>
<thead>
<tr>
<th>expression</th>
<th>character</th>
<th>content</th>
<th>referent</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;I&quot;</td>
<td>variable</td>
<td>rigid</td>
<td>me</td>
</tr>
</tbody>
</table>

| "the guy speaking in ICC 666 on 3/20/91" | rigid | variable | me |
Since I have a linguistics audience here, I should mention that I think one nice fallout of Kaplan's distinction between the context of utterance and the possible circumstances of evaluation and his concomitant distinction between character and content is that we finally have a precise way to demarcate the field of pragmatics from that of semantics. I think that people had tended to use pragmatics in kind of a loose way to mean the sort of semantics that one had trouble formalizing properly. For example, discussions of presupposition were considered part of pragmatics, as were discussions of "discourse analysis", where one wanted to look at units bigger than the sentence. But of course this way of dividing things isn't very useful, because as we get more sophisticated we will presumably learn to be just as precise about the rules of pragmatics as we are about those of semantics. And the natural way to draw the line, it seems to me, is to say that pragmatics is the study of character and semantics is the study of content. Both of these subjects, I think, have something to do with what we think of as "meaning".

I had thought for several years that Kaplan's account was basically right, and that what was left to do was to isolate more features of the context that are relevant to determining the contents of utterances and to spell out more carefully what the rules for determining the content in a given context are. For example, to figure out the TIME that I am referring to when I say "Georgetown will win the NCAA", I may need to know something about the sentences that immediately preceded my utterance. If I had just said "The year 2000 will be a banner year", for example, the time referred to will not be the same as it would if I had just said "John Thompson will return to Georgetown next year." Somebody needs to explain how prior discourse fits into the context. Similarly, somebody needs to explain more carefully than Kaplan how the content of the word "you" is determined from the context. But all this is just detail.

This fall, however, I was asked to referee a paper that made me think that maybe things are a little more complicated than I had thought. The paper turns out to be written by a guy named "Allan Sidelle". (I asked the editor to give me his name last week, so I could give him proper credit in this talk.) Sidelle's paper makes the very simple point that utterances of "I am here now" do NOT always seem to be true in the trivial way that Kaplan had assumed. In particular, when you call somebody with an answering machine you often hear his voice saying "I am not here right now." And, if he really isn't home (and isn't just screening his calls so he doesn't have to talk to you) then we seem to have the strong intuition that he is telling the truth. Now when I first thought about this, I must admit I thought it was rather silly. Surely, there must be some way to explain this rather odd example by pointing to some rather odd features of the answering-machine context. But after reading Sidelle's paper and thinking about it a little more I am no longer so sure. So I thought I
would let you think about it and help me out.

The puzzle seems to arise from the following four facts.

1. "I" refers to the speaker of u.
2. "here" refers to location of u.
3. "now" refers to time of u.
4. speaker of u is at place of u at time of u
1-4 imply utterances of "I am not here now" must be false.
5. "I am not here now" is sometimes true.

2 can be contested. But then we are talking about another use "here" (one in which "here" and "there" can be used interchangeably).

5 can be contested. One way to do this would be to say that the speaker here is the machine. But this is implausible. Would the "I" in a note left on the door for UPS be the piece of paper? A more plausible way to deny 5 is to point out that it was false at the time the recording was made. So if the the time of u is the time at which the speaker was really speaking, the sentence is false. But there is still something fishy here. We really DON'T interpret the sentence that way when we hear it. I guess we COULD say that the utterance we hear is false, but that we know the true proposition the speaker wanted to communicate--When you hear this I will not be at home.

4 can be contested. This is the option favored by Sidelle. The speaker is the guy who made the recording. But the utterance occurs at the time you hear it and the speaker may be in Tahiti at that time. But there are some problems with this view. One kind of objection comes from a general consideration about the nature of utterances and any other actions. Can a guy in Tahiti really be uttering something on his answering machine in Washington? This is a funny kind of action at a distance. Another kind of objection has to do with the fact that this solution seems to require that time be treated very differently than location. The time of the utterance is supposed to be the time that the recording his HEARD. Why, then, isn't the location of the utterance the place where the recording is heard, i.e., at the listenter's end of the line. Now you might think that the time is NOT the time the utterance is heard, but the time that that it is PLAYED. So by the correct analogy the place is the place where it is played. But it seems to me that it is just a technological fluke that we have to play the messages at a machine near the senders telephone. Suppose we could have an answering machine that somehow played its messages at the listener's end. Should my messages then be changed to say "I am not there"?
Finally, let me throw out a kind of Gricean solution that occurred to me, but that I don't really think will work. Perhaps words like "here" and "now" don't always refer to the actual location and time of the utterance, but to some location and time determined by a more complicated convention. Perhaps if the speaker has reason to think that the listener believes the speaker is at p, and if the listener knows the speaker has such reason, then "here" refers to p. When you call my answering machine and you hear me answer, I KNOW you are going to think I am home, and that you know I am going to think that. So I use "here" to refer to my home. If this view is right, then "I am not here" would not only be correct in answering machine situation, but it would also be correct in the call-forwarding situation. You call my office, my office phone sends the call to my home. Can I then truthfully say "I am not here"? Personally, I am not so sure about this. But there is another case where I am pretty sure the Gricean solution gives the wrong answer. Suppose you think I am in town and I know you think that. I want to tell you that it is snowing town, so you don't have to go to school. I am really in Florida where it is raining. I call you long-distance and say "It is snowing here". I really do communicate a thought in a very efficient way, but it seems to me that I am lying. If I am, then adding a Gricean twist will not solve the answering machine puzzle. So I guess the moral of my talk is that we have made a lot of progress in understanding meaning, but there are still some things we don't quite understand.