Mentalism and the Gricean program

Jeff Speaks
PHIL 93914
March 27, 2008

1 Mentalism

Let mentalism be the view that we can explain what it is for an expression to have a meaning in a language in terms of certain of the propositional attitudes of users of the language.

According to most mentalists, not just any propositional attitudes will do. The background motivation for mentalist is to display the priority of thought over linguistic meaning, and the class of propositional attitudes is very broad, and contains attitudes very closely linked to the use of language, such as saying, asserting, and telling. When mentalists try to give an account of language in terms of the attitudes, they try to account for facts about meaning in terms of facts about what agents intend, believe, or judge, and not in terms of what agents say or assert by uttering sentences of their language.

One way of distinguishing between the relevant classes of propositional attitudes is via distinction between action-entailing and non-action-entailing attitude verbs. In the case of action-entailing attitude verbs, we can expand an ascription \(\alpha V\)'s that \(\sigma\) to one of the form \(\text{By } \phi\text{ing, } \alpha V\)'s that \(\sigma\), where ‘\(\phi\text{ing}’ denotes some action of the referent of \(\alpha\).’ This class of propositional attitude verbs includes, for example, “says,” “means,” and “asserts”; we can expand an ascription of the form \(\alpha \text{ said that } \sigma\) to one of the form \(\text{By } \phi\text{ing, } \alpha \text{ said that } \sigma\), and an ascription of the form \(\alpha \text{ asserted that } \sigma\) to one of the form \(\text{By } \phi\text{ing, } \alpha \text{ asserted that } \sigma\). For example: “By uttering ‘Schnee ist wiess,’ John said that snow is white”; “By spreading his arms, the umpire meant that the base runner was safe.” By contrast, attitude verbs like “believes” and “desires” are not action-entailing; a
similar expansion of a sentence of the form "α believed that σ" to "By φing, α believed that σ" yields a sentence which is at best awkward, and at worst nonsensical.

One way to think of the mentalist program is as trying to reduce meaning to the non-action-entailing attitudes. Why do this? The problem with propositional attitude verbs like "asserts" and "says" which the above grammatical distinction illustrates is that they are too closely linked to the meanings of signs. For usually, in a sentence of the form "By φing, α V's that σ", the action denoted by ‘φing’ will include mention of a sign which has a meaning in a language or population. In the case of our example sentence — “By spreading his arms, the umpire meant that the base runner was safe” — the action performed has a meaning in its context, and there is a clear link between the meaning of the action or gesture and what the umpire meant by performing the gesture. In cases where the action is the assertion of a sentence, the connection is even more obvious. In such cases, the obvious way to give an account of what constitutes an agent asserting or saying something will make reference to the meaning of the sentence uttered or gesture performed. So, showing that there’s a link between meaning and, for example, assertion, wouldn’t suffice to show the priority of thought over language.

So how do we get from meaning to the non-action-entailing attitudes? There are, broadly, two going answers to this question:

1. The meanings of sentences are fixed by the typical causes and effects of utterances of the sentence, where these causes and effects are specified in terms of facts about mental content.

2. The meaning of a sentence is determined by the effects which speakers intend to bring about by uttering that sentence.

The former route naturally yields an account of meaning in terms of belief; the latter, obviously, an account of meaning in terms of intention.

The intention-based route to an account of linguistic meaning will have to use action-entailing propositional attitudes as intermediaries. To see why, note that the effects which a speaker intends to bring about by an utterance need not be closely related to the meaning of the utterance; metaphor, sarcasm, and jokes are all cases in which there seems to be no very close connection between the meaning of the sentence uttered and the intentions of the speaker. But, one might think, there is, even in these cases, a close connection between the intentions of the speaker and what the speaker means, asserts, or communicates by her utterance. This suggests that intentions are better used to account for these action-entailing propositional attitudes than, directly, for the meanings of linguistic expressions. The hope for the mentalist would then be to use these action-entailing attitudes to give an account of meaning, and so to construct, indirectly, an account of meaning in terms of the intentions of speakers. The picture of mind and language envisaged by this indirect mentalist strategy might then be represented as follows:
2 The Gricean program

Since Grice's article “Meaning,” the most popular choice of an action-entailing propositional attitude to play the middle role in an indirect mentalist account of linguistic meaning has been the propositional attitude of speaker-meaning.

Grice thought that (1) facts about what expressions mean are to be explained, or analyzed, in terms of facts about what speakers mean by them; and he thought, further, that (2) facts about what speakers mean by their utterances can be explained in terms of the intentions of speakers. These two views comprise the ‘Gricean program’ for reducing meaning to mental content. The Gricean program is the most popular version of the indirect mentalist strategy. We’ll be focusing on part (2) of the Gricean program, the attempt to explain speaker-meaning in terms of audience-directed intentions.

2.1 The analysis of speaker-meaning

The basic version of the Gricean analysis of speaker-meaning in terms of the intentions of speakers is the following:

\[ [G] \ a \ means \ p \ by \ uttering \ x \ \equiv \ a \ intends \ in \ uttering \ x \ that \ \\
(1) \ his \ audience \ come \ to \ believe \ p, \ \\
(2) \ his \ audience \ recognize \ this \ intention, \ & \ \\
(3) \ (1) \ occur \ on \ the \ basis \ of \ (2) \]

For reasons we will not go into, this is meant to be understood as requiring a single, self-referential intention, rather than three separate intentions.

\footnote{See Grice (1957, 1969).}
The motivation for clause (1) should be clear: in paradigm cases of telling someone something, we are trying to convey a belief to an audience. Why did Grice not just stop there? What are clauses (2) and (3) meant to rule out?

2.2 Some counterexamples to clause (1)

I think that none of (1)-(3) are necessary for speaker-meaning, and that (1)-(3) are not jointly sufficient. If this is right, the question then becomes whether the analysis can be fixed. We'll consider some possible modifications along the way.

Challenges to (1):

- Reminding someone of something.
- Confessing something to someone who’s caught you red-handed.
- Taking an oral exam.
- Saying a prayer while believing in the omniscience of God.

A possible fix: change (1) from ‘believes’ to ‘occurrently believes’ or ‘actively believes.’ This helps with cases of reminding, but has no obvious application to the others. For example, consider cases of examination in which a student is asked a question, and gives an answer which, he believes, the teacher already believes; since the teacher is presently asking him the question, it would be reasonable for him to think that she actively believes it. Hence he does not intend that she come to actively believe it.

The defender of the Gricean account may object that although all these cases falsify [G] as it stands, in each case I do intend that someone be in a mental state of some sort which is closely related to what I mean by my utterance. For example, perhaps in all the above cases I intend my audience to believe that I believe $p$.

A worry here is that in many cases, I might well not care if my audience comes to believe anything about my beliefs. (‘There’s a man behind you with a gun!’) A more plausible way to integrate the suggestion is by following Grice [1969] and distinguishing between protreptic, and the latter exhibitive utterances, the former which can be analyzed by [G], and the latter which can be analyzed by a formula which replaces ‘believes $p$’ in clause (1) with ‘believes that I believe $p$’.

That might look like this:

\[ a \text{ means } p \text{ by uttering } x \equiv \]

(a) $a$ intends in uttering $x$ that
(1) his audience come to actively believe $p$,
(2) his audience recognize this intention, &
(3) (1) occur on the basis of (2); or
(b) $a$ intends in uttering $x$ that
(1') his audience come to actively believe that \( a \) believes \( p \).
(2') his audience recognize this intention, \&
(3') (1') occur on the basis of (2')

Problem: this makes it impossible to mean that you believe \( p \) without also meaning \( p \).
(This is also a problem for the ‘all exhibitive’ version of the analysis.

Possible fix: build the protreptic/exhibitive distinction into the account. But this runs into the problem that it is hard to know how to state the distinction without appealing to the facts about speaker-meaning we are out to explain.

Possible further counterexample to the ‘all exhibitive’ analysis:

Bob is an air-traffic controller at Newark Airport, who relays information from the head controller to arriving pilots. Bob does not believe this information; he thinks that the head controller is involved in a conspiracy to cause air accidents. Nevertheless, Bob has an overly strong sense of duty, and feels it his duty to convey the information passed to him from the head controller to pilots. So when the head controller tells him that Runway 2 is open for landing, Bob says over his radio to the appropriate pilot, “Runway 2 is open for landing.” Bob intends by uttering this that the pilot should believe that Runway 2 is open for landing, and intends that he should come to that belief on the basis of recognition of his intention. Nevertheless, Bob certainly does not intend that the pilot come to believe that he, Bob, believes that Runway 2 is open for landing; Bob doesn’t believe this, and would say so were an accident to occur.

What have these examples shown us so far? The Gricean account relies on there being a correlation between what a speaker means by an utterance and the contents of beliefs that he intends that his audience come to have. In each case discussed so far, there is a belief ready to do the work the Gricean needs done; the problem is that which mental state this is varies from case to case.

The Gricean might reply to this sort of problem by turning from beliefs to some other sort of propositional attitude. For example, perhaps, as Neale (1992) has suggested, they are intentions to cause agents to entertain certain propositions.

Problem: utterances of disjunctions. Speaker-meaning (like assertion, telling, and other closely related attitudes) does not distribute over disjunction; but I can’t intend someone to entertain a disjunction without intending them to entertain the disjuncts. (Or, at least, I can often intend them to entertain the disjuncts while not meaning them.)

Possible fix: appeal to the notion of a ‘primary intention’, in Schiffer’s sense: “To specify one’s primary intention in doing \( X \) ... is to give one’s reason for doing \( X \),” as opposed to specifying an intention which one has because one has a given primary intention (Schiffer (1972), 62). But this seems too restrictive: Suppose that your car runs out of gas, and I say to you, “There’s a gas station around the corner.” Clearly, though I do intend that you come to believe that there is a gas station around the corner, I only have this intention
because I intend you to come to believe that you can fill your car up around the corner; so the former is a secondary intention and not, in the relevant sense, the reason for my utterance. Nevertheless, it does seem as though I meant by my utterance that there is a gas station around the corner; and this is a counterexample to any version of the analysis restricted to the primary intentions of speakers.

A more fundamental problem with the switch from believing to entertaining propositions: the conditions given make it almost impossible, in many contexts, for speakers to mean things by their utterances. Suppose I know that, usually, when you hear a certain sentence which in the context means \( p \), you immediately, almost reflexively, entertain \( p \). This is, after all, not a farfetched scenario; it is the condition that normal speakers are in with respect to sentences they understand. But if I know this, then I can hardly intend that you come to entertain \( p \) on the basis of your recognition of my intention when I utter this sentence. (This is why Neale drops clause (3); more on this below.)

### 2.3 Counterexamples based on persuasive discourse

The Gricean account requires for a speaker to mean \( p \) not only that she intend that her audience come to believe \( p \) and that her audience recognize this intention, but also that her audience come to this belief on the basis of that recognition. The problem is that, when giving an argument of some kind, a speaker will typically intend that her audience come to believe the conclusion on the basis of belief in the premises, and \( p \) on the basis of a recognition of the speaker’s intention.

(Another class of counterexamples to clause (3) of the account, though a less significant one, comes from a sub-class of the class of cases of reminding, discussed above. Suppose that you cannot remember your friend’s name, though you feel as though it is on the tip of your tongue. In such a case, I can’t intend that you come to form a belief about the name on the basis of recognition of my intention; I know that hearing the name alone will cause you to form the belief.)

Although Grice (1969) was the first to notice this problem for his account of speaker-meaning, he never offered a solution. It is tempting to try to solve these cases by treating them as exhibitive utterances; one might think that, even if a speaker does not intend that the audience believe \( p \) on the basis of recognition of this intention, the speaker might intend that the audience believe that the speaker believes \( p \) on the basis of this intention. But in neither case will the amendment solve the problem; we can imagine cases in which the speaker knows that the audience already knows that the speaker believes the proposition of which he is being reminded or persuaded, or cases in which the claim that a speaker has a certain belief is in fact the conclusion of the argument; these will raise the problems with exhibitive analyses discussed above.

The only response on offer to these problems which counts them as cases of speaker-meaning is a radical one, which has been proposed by Stephen Neale. Neale proposes that condition (3) simply be dropped from the account, to yield the following:

---

\(^{2}\)This is suggested in Rumfitt (1995).
a means \( p \) by uttering \( x \equiv a \) intends in uttering \( x \) that
(1) his audience come to believe \( p \), &
(2) his audience recognize this intention.

Problem: Consider a case in which we have a surgeon capable of manipulating a patient’s brain so as to make him disposed to assent to certain sentences. The surgeon, in this case, intentionally manipulates the patient in such a way as to make the patient disposed to accept both “\( S \)” and “My surgeon intends me to believe that \( S \)”. Plausibly, the patient’s being so disposed is sufficient to guarantee that the patient believes that \( S \), and believes that the surgeon intends that he believe that \( S \). By Neale’s account, then, it follows that by manipulating the brain of his patient, the surgeon meant that \( S \) by manipulating the brain of the patient, which seems clearly to be incorrect.

2.4 Speaker-meaning without intended effects

The second broad class of cases which are incorrectly classified by the Gricean account are cases in which the speaker means something by an utterance but does not intend to bring about any beliefs in his audience at all. Indeed, this is a special case of the more general fact that it is not a necessary condition on meaning something by an utterance that the speaker intend to bring about any effects in an audience:

- I might mean that trespassing is not allowed by erecting a sign to that effect on my property even though there is no audience present when I put the sign up.
- Writing in a diary.
- Uses of language in thought or calculation.

Why the first sort of cases are the easiest: the idea of an intended audience. This does not help with the second two.

Possible fix: the audience is yourself. Problem: Are we to say that by writing something in my diary I intend to bring about a certain belief in myself by recognition of this (my own) intention? I already know that I have the relevant belief; this is presumably part of the explanation of my writing as I did in my diary, and not the intended effect of my writing.

These audienceless cases point to a fundamental problem with the Gricean account: intending to bring about effects by one’s utterance is not an essential part of meaning something by an utterance. Audienceless cases are one way of making this manifest, but there are others which do the same. Consider an innocent person arrested on charges of espionage by his own government; under torture, he continues to claim that he is not a spy, simply because he feels that it is his duty to do so. He knows that his torturers will not believe him; at this point he doesn’t even care if they do. Nonetheless, the torturers are his audience, and he does mean by his utterance that he is not a spy.

\footnote{The case is a variant of one presented in Harman (1974).}
2.5 Meaning, speaker-meaning, & Moore’s paradox

I mentioned at the outset of this chapter that there is a persistent thought to the effect that, even if we have not yet arrived at the right form of the Gricean account, some true version is out there to be discovered. The discussion of the problems raised by revising Grice’s original account has been designed to discourage this thought. I suggest that one source of these problems is a connection between the meaning of the sentence uttered by the speaker and what the speaker meant by uttering the sentence; this connection may be expressed by the following principle:

\[[M/S-M]\] If an agent utters a sentence which means \( p \) sincerely and seriously (without sarcasm, irony, etc.), then the agent means \( p \) by her utterance.

This principle connecting meaning and speaker-meaning has been in the background of many of the arguments of the preceding sections, for many of the cases for which the Gricean has been unable to account have been cases in which among the things meant by the speaker was the meaning of the sentence in the context of utterance. Inasmuch as he seeks to account for speaker-meaning wholly in terms of the intentions of speakers, it is unsurprising that this principle should pose a problem for the Gricean.

As it turns out, though, this principle can also be employed in a general argument involving Moore’s Paradox which shows that no intention-based account of speaker-meaning can be correct. G. E. Moore drew attention to the oddness of uttering sentences like ‘It is raining, but I do not believe that it is raining.’ However odd it is to utter such a sentence — and it is difficult to imagine a scenario in which a speaker who understands this sentence could utter it sincerely and seriously — there are contexts of utterance in which a sentence of this form is clearly true. To construct such a context, after all, we need only select a speaker who does not believe some truth, find some sentence ‘\( S \)’ which expresses this truth, and assign that speaker as the speaker of \( \langle S, \text{ but I do not believe that } S \rangle \) in the context.

Other quasi-paradoxical sentences may be generated by focusing, not on belief, but on intention; the one relevant to evaluation of the Gricean account of speaker-meaning is

\[\text{ It is raining, but I do not intend you to believe that it is raining.}\]

As with Moore’s original paradoxical sentence, it is clear that there are some contexts of utterance in which the sentence is true. The key question here is: are there contexts in which a speaker could utter this sentence seriously and sincerely?

It seems so. For consider the following three examples:

An unfaithful husband is found out by his wife; he might deeply regret his act and love his wife, but know that there is no way to convince her of this. Still,
he might well say to her, “I love you, though I don’t expect (intend) you to believe that.” It seems clear that what he says might well be true.

A harbors ill feelings for B. Moreover, he knows that B knows this, and that B will disbelieve anything that he says. A calculates to himself the best way to cause B the most discomfort, and hits on the following strategy: he will tell B that it is raining outside, knowing that B will upon hearing this immediately come to believe that it is not raining outside. Then, he hopes, B will fail to bring an umbrella on his next trip outside, get wet, and feel all the worse for knowing that his enemy A had warned him truthfully in advance. For added confusion, A also decides to tell B, truthfully, that he does not intend him to believe that it is raining. So A says to B: “It is raining, but I don’t intend you to believe that it is raining.”

Suppose that A has just taken a truth serum, knows that B hates being wet, and, as above, harbors ill feelings for B. While gazing at B malevolently, he says, “It is raining, but I don’t intend you to believe that it is raining.”

In each of these cases, an agent utters a sentence which has a meaning; and in each of these cases, it seems that the agent utters the sentence seriously (they are not being sarcastic, ironic, joking, etc.). But if this is right, then it follows from the above principle connecting linguistic meaning with speaker-meaning that each of the agents in these examples means by their utterance what the sentence means in the context of utterance.

With these cases in mind, we can then argue as follows:

[1] The speaker (A) utters a sentence to his audience (B) which means
   \((p & A \text{ does not intend } B \text{ to believe } p)\)

[2] The sentence uttered by the speaker is true.

Since the sentence uttered was in each case true, each of its conjuncts must be true; so we get as a further premise the second conjunct:

[3] \(A \text{ does not intend } B \text{ to believe } p\) ([1],[2])

But since the utterances are serious (nonsarcastic, etc.), we can derive a further claim about what the speaker means by her utterance:

[4] By her utterance, A means \((p & A \text{ does not intend } B \text{ to believe } p)\) ([1],[M/S-M])

\footnote{Thanks to Jonathan Beere for this example.}
So we have reached the conclusion that some conjunctive proposition is meant by the speaker in these contexts. To get the desired conclusion, we must employ a further premise:

[5] The propositional attitude relation expressed by “means” distributes over conjunction; that is, indexicality aside, a sentence of the form $\langle \alpha \rangle$ means that $\sigma$ and $\sigma' \rightarrow$ entails $\langle \alpha \rangle$ means that $\sigma \rightarrow$ and $\langle \alpha \rangle$ means that $\sigma' \rightarrow$

Premise [5] seems to me intuitively quite plausible; how could a speaker mean that $p$ and $q$ without also meaning that $p$? The case seems even clearer when speaker-meaning is laid alongside other action-entailing propositional attitudes, such as assertion. It seems clear that assertion distributes over conjunction; if John asserts $p \& q$, then John asserts $p$. Indeed, virtually all of the action-entailing propositional attitudes which come to mind — telling, informing, saying, communicating, commanding — clearly do distribute over conjunction. It would be surprising if, lone among the members of this class, the propositional attitude of speaker-meaning did not distribute over conjunction.

The problem for the Gricean is that [4] and [5] jointly entail


This is problematic because the conjunction of [6] and [3] is inconsistent with the Gricean claim that what it is for a speaker to mean $p$ by an utterance is for that speaker to intend to bring about the belief $p$ in her audience.

Of course, we have by now discussed a number of other cases in which having this sort of audience-directed intention fails to be a necessary condition for meaning something by an utterance. The interest of the present argument is that it does not depend on the details of specific formulations of the Gricean account. The examples discussed above began with an utterance of the form $\langle \neg S \rangle$, but I do not intend that you believe that $S$; as such it counts against the version of Griceanism which takes intentions to bring about believing $p$ to be constitutive of meaning $p$; but we just as easily could have begun with utterances of the form $\langle \neg S \rangle$, but I do not intend that you believe that $S$ or $\neg S$, but I do not intend that you actively believe that $S$ or $\neg S$, but I do not intend that you entertain the proposition $S$. So this is a class of counterexamples with substantial generality; it seems likely that, whichever audience-directed intention is seized upon by a particular version of the Gricean analysis, we could construct quasi-paradoxical sentences of a form appropriate to refute that analysis.

2.6 Assessment of Grice’s account

It is false.

Might the Gricean account fare better as an account of some other propositional attitude relation? Nothing in the indirect mentalist strategy, after all, requires that the action-
entailing propositional attitude which bridges the gap between linguistic meaning and non-action-entailing propositional attitudes be speaker-meaning.

Some evidence for this comes from a grammatical distinction between two classes of action-entailing propositional attitude verbs. For some such verbs, an ascription of the form "By φing, α V’s that σ to β" can be expanded to an ascription of the form "By φing, α V’s that σ to β", which includes mention of an audience as the indirect object of the verb. The Gricean analysis makes essential reference to an audience; it thus seems plausible that, if it is a successful analysis of any propositional attitude relation employed in ordinary speech, it will be one expressed by a verb in this class. Given this, it is striking that the propositional attitude of speaker-meaning is not in this class; there is no natural sentence "By uttering x, α meant that σ to β". Verbs which naturally take a singular term referring to an audience as indirect object are, unsurprisingly, verbs very closely related to communication, such as “communicated,” “told,” “informed.”

Might it be the case that the Gricean analysis is an adequate analysis of communication? If so, then perhaps the following analysis would be true:

\[
a \text{communicates } p \text{ to his audience by uttering } x \equiv \\
a \text{intends in uttering } x \text{ that} \quad \\
(1) \text{his audience come to actively believe } p, \\
(2) \text{his audience recognize this intention, } \\
(3) \text{(1) occur on the basis of (2)}
\]

But meeting these conditions is neither necessary nor sufficient for communication. The conditions are not sufficient because “communicates” is a success verb; if I say that I communicated something to an audience, that entails that my audience understood what I was trying to get across; otherwise, it would be natural to describe the situation as one in which I tried to communicate something, but failed. The conditions are not necessary because, in many cases, we would take the statement of the conclusion of an argument as a case of communication; but, as we saw before, persuasive discourse does not meet condition (3) of the Gricean analysis.

Possible reply: the Gricean account is a stipulative definition of a technical notion of S-meaning:

\[
a \text{S-means } p \text{ by uttering } x \equiv a \text{intends in uttering } x \text{ that} \quad \\
(1) \text{his audience come to actively believe } p, \\
(2) \text{his audience recognize this intention, } \\
(3) \text{(1) occur on the basis of (2)}
\]

Perhaps we can then define expression meaning in terms of S-meaning; after all, that was the ultimate target.

But this poses a problem once we see the shape of the intended account of expression-meaning.

Since the publication of David Lewis’s *Convention* in 1969, it has become standard practice
for mentalist proponents of the indirect strategy to try to account for facts about the meanings of linguistic expressions in terms of conventions governing the use of those expressions by speakers to mean things. (See, e.g., Schiffer (1972, 1982); Bennett (1973, 1976); Loar (1976, 1981).)

The simplest way to present the thesis of the convention-theorist is to define convention in terms of mutual knowledge, where \(a\) and \(b\) mutually know \(p\) just in case \(a\) knows \(p\), \(b\) knows that \(a\) knows \(p\), \(a\) knows that \(b\) knows that \(a\) knows \(p\), and so on, and vice versa.\

Using this notion, we can state the view of the convention-theorist as follows:

\[
[C] \ x \ \text{means} \ p \ \text{in a population} \ G \equiv \\
(1) \ \text{almost all members of} \ G \ \text{utter} \ x \ \text{only when they mean} \ p \ \text{by uttering} \ x, \\
(2) \ \text{almost all members of} \ G \ \text{mutually know (1), &} \\
(3) \ (1) \ \text{obtains because of (2)}
\]

But the argument above forces us to state the convention-theoretic analysis of language in terms of \(S\)-meaning, rather than speaker-meaning. But the relation of \(S\)-meaning is a technical notion which was just defined a few pages back; what grounds are there for saying that speakers must have beliefs about \(S\)-meaning?

Does this, once noticed, raise doubts about \([C]\) itself? Could agents speak a meaningful language without having any beliefs about speaker-meaning (as opposed to other propositional attitudes related to communication)?

References


---

\(^6\)This is the route of Schiffer (1972), ch. 5.
