

14

First Person Authority

When a speaker avers that he has a belief, hope, desire or intention, there is a presumption that he is not mistaken, a presumption that does not attach to his ascriptions of similar mental states to others. Why should there be this asymmetry between attributions of attitudes to our present selves and attributions of the same attitudes to other selves? What accounts for the authority accorded first person present tense claims of this sort, and denied second or third person claims?

The point may be made, and the question asked, in the modality either of language or of epistemology. For if one can speak with special authority, the status of one's knowledge must somehow accord; while if one's knowledge shows some systematic difference, claims to know must reflect the difference. I assume therefore that if first person authority in speech can be explained, we will have done much, if not all, of what needs to be done to characterize and account for the epistemological facts.

The connection between the problem of first person authority and the traditional problem of other minds is obvious, but as I pose the former problem, there are two important differences. First person authority is the narrower problem, since I shall consider it only as it applies to propositional attitudes like belief, desire, intention; being pleased, astonished, afraid, or proud that something is the case; or knowing, remembering, noticing, or perceiving that something is the case. But I shall not discuss what are often taken to be central to the problem of other minds: pains and other sensations, and knowledge, memory, attention, and perception as directed to objects like people, streets, cities, comets, and other non-propositional entities. What holds for the propositional attitudes ought, it seems, to be relevant to sensations and the rest, but I do not explore the connections here.

All propositional attitudes exhibit first person authority, but in various degrees and kinds. Belief and desire are relatively clear and simple examples, while intention, perception, memory, and knowledge are in one way or another more complex. Thus in evaluating someone's claim to have noticed that the house is on fire, there are at least three things to consider: whether the house is on fire, whether the speaker believes the house is on fire, and how the fire caused the belief. With respect to the first, the speaker has no special authority; with respect to the second,

he does; and with respect to the third, responsibility is mixed and complex. The question whether someone intends to lock the door by turning the key depends in part on whether he wants to lock the door and believes that turning the key will lock the door; and whether this belief and desire have caused, in the right way, a desire to turn the key. Special authority attaches directly to claims about the desire and belief, less directly to claims about the necessary causal connection. These differences among the ways in which first person authority may apply to propositional attitudes are important and worth exploring. But in every case, first person authority is relevant, and it is the general case I wish to consider here. Since in almost every instance, if not in all, first person authority rests at least partially on a belief component, I shall concentrate on the case of belief.

Though there is first person authority with respect to beliefs and other propositional attitudes, error is possible; this follows from the fact that the attitudes are dispositions that manifest themselves in various ways, and over a span of time. Error is possible; so is doubt. So we do not always have indubitable or certain knowledge of our own attitudes. Nor are our claims about our own attitudes incorrigible. It is possible for the evidence available to others to overthrow self-judgements.

It comes closer to characterizing first person authority to note that the self-attributer does not normally base his claims on evidence or observation, nor does it normally make sense to ask the self-attributer why he believes he has the beliefs, desires, or intentions he claims to have. This feature of self-attributions was remarked by Wittgenstein: 'What is the criterion for the redness of an image? For me, when it is someone else's image: what he says or does. For myself, when it is my image: nothing.'¹ Most philosophers have followed Wittgenstein in this, and have extended the criterion to the propositional attitudes, as we shall see.

This feature of first person authority, suggestive as it may be, does not help explain the authority. This is so partly because of the caveats—'normally' we do not make self-attributions on the basis of evidence, but sometimes we do; 'usually' it doesn't make sense to ask someone why he believes he has a certain belief or desire or intention, but sometimes it does. Even in the exceptional cases, however, first person authority persists; even when a self-attribution is in doubt, or a challenge is proper, the person with the attitude speaks about it with special weight.

But the existence of exceptions is not the chief reason first person authority isn't explained by the fact that self-attributions are not based on evidence; the chief reason is simply that claims that are not based on evidence do not in general carry more authority than claims that are based on evidence, nor are they more apt to be correct.

Contemporary philosophers who have discussed first person authority have made little attempt to answer the question why self-ascriptions are privileged. It

¹ Ludwig Wittgenstein, *Philosophical Investigations*, §377.

is long out of fashion to explain self-knowledge on the basis of introspection. And it is easy to see why, since this explanation leads only to the question why we should see any better when we inspect our own minds than when we inspect the minds of others.

A few philosophers have denied that the asymmetry exists; Ryle is a sturdy example. In *The Concept of Mind* Ryle suggests that what we take for 'privileged access' is due to nothing more than the fact that we are generally better placed to observe ourselves than others are. Ryle writes, 'in principle, as distinct from practice, John Doe's ways of finding out about John Doe are the same as John Doe's ways of finding out about Richard Roe'. He continues,

the differences are differences of degree, not of kind. The superiority of the speaker's knowledge of what he's doing over that of the listener does not indicate that he has Privileged Access to facts of a type inevitably inaccessible to the listener, but only that he is in a very good position to know what the listener is often in a very poor position to know. The turns taken by a man's conversation do not startle or perplex his wife as much as they had surprised and puzzled his fiancée, nor do close colleagues have to explain themselves to each other as much as they have to explain themselves to their new pupils.²

I agree with Ryle that any attempt to explain the asymmetry between first person present tense claims about attitudes, and other person or other tense claims, by reference to a special way of knowing or a special kind of knowledge must lead to a skeptical result. Any such account must accept the asymmetry, but cannot explain it. But Ryle neither accepts nor explains the asymmetry; he simply denies that it exists. Since I think it is obvious that the asymmetry exists, I believe it is a mistake to argue from the absence of a special way of knowing or a special mode or kind of knowledge to the absence of special authority; instead, we should look for another source of the asymmetry.

Ayer at one time took a line similar to Ryle's. In *The Concept of a Person* he emphasizes that first person ascriptions can be in error; and he allows that such ascriptions are privileged.³ But when he comes to describe the authority of self-ascriptions, he compares it to the authority we sometimes allow an eye-witness when compared with secondhand reports. This analogy seems to me unsatisfactory for two reasons. First, it fails to tell us why a person is like an eyewitness with respect to his own mental states and events while others are not. And second, it does not suggest an accurate description of what first person authority is like. For first person attributions are not based on better evidence but often on no evidence at all. The authority of the eyewitness is at best based on inductive probabilities easily overridden in particular cases: an eyewitness is discredited and his evidence discounted if he is a notoriously unreliable observer, prejudiced, or myopic. But a person never loses his special claim to be right about his own attitudes, even when his claim is challenged or overturned.

² Gilbert Ryle, *The Concept of Mind*, 156, 179.

³ A. J. Ayer, 'Privacy'.

Joseph Agassi has actually maintained that we know the mental states and events in other minds better than those in our own mind. He distinguishes privileged access from the commonsense truth that 'every person has access to some information available to that person alone, and it involves one's self, at least as an eye-witness'. He goes on:

The doctrine of privileged access is that I am the authority on all my own experiences . . . The thesis was refuted by Freud (I know your dreams better than you), Duhem (I know your methods of scientific discovery better than you), Malinowski (I know your customs and habits better than you), and perception theorists (I can make you see things which are not there and describe your perceptions better than you can).⁴

Aside from Freud's case, there is little here to threaten first person authority. Freud's views, by extending the concepts of intention, belief, desire, and the rest to include the unconscious, do mean that with respect to some propositional attitudes a person loses direct authority. Indeed, loss of authority is the main distinguishing feature of unconscious mental states. Of course, the pre-Freudian attitudes remain as subject as ever to first person authority. But more interesting is the fact that in psychoanalytic practice, recovery of authority over an attitude is often considered the only solid evidence that the attitude was there before being noninferentially appreciated by its holder. Thus those cases of unconscious mental states that were unsystematically recognized to exist before Freud are indirectly included in the scope of first person authority by psychoanalysis. So I do not think the existence of unconscious attitudes threatens the importance of first person authority.

I turn now to philosophers who have assumed that there is such a thing as first person authority, and have accepted Wittgenstein's description of the difference between first and third person attributions.

Strawson discusses first person authority in the context of trying to answer skepticism about other minds. According to Strawson, if the skeptic understands his own question ('How does anyone know what is going on in someone else's mind?'), he knows the answer. For if the skeptic knows what a mind is, he knows it must be in a body, and that it has thoughts. He also knows that we attribute thoughts to others on the basis of observed behaviour, but to ourselves without such a basis. Strawson writes:

In order to *have* this type of concept [of a mental property], one must be both a self-ascriber and an other-ascriber of such predicates [which ascribe mental properties], and must see every other as a self-ascriber. In order to *understand* this type of concept, one must acknowledge that there is a kind of predicate which is unambiguously and adequately ascribable *both* on the basis of observation of the subject of the predicate *and* not on that basis, i.e. independently of observation of the subject.⁵

⁴ Joseph Agassi, *Science in Flux*, 120.

⁵ Peter Strawson, *Individuals*, 108. (Anita Avramides has pointed out that Strawson has contributed more to this issue than I allow. See her 'Davidson and the New Sceptical Problem'.)

This cannot be deemed a satisfactory answer to the skeptic. For the skeptic will reply that though Strawson may have correctly described the asymmetry between first and other person ascriptions of mental predicates, he has done nothing to explain it. In the absence of an explanation, the skeptic is surely justified in asking how we know that the description is correct. In particular, why should we think that a predicate that is sometimes applied on the basis of observation, and sometimes not, is unambiguous? This question, to which Strawson has not addressed himself, is a major source of skepticism about knowledge of other minds.

Richard Rorty has attempted an explanation. We are asked to imagine that originally self-ascriptions were made on the basis of the same sort of observation or behavioral evidence as other-ascriptions. It was then noticed that people could ascribe mental properties to themselves without making observations or using behavioral evidence, and that self-ascriptions turned out in the long run to provide better explanations of behavior than third person ascriptions. So it became a linguistic convention to treat self-ascriptions as privileged: 'it became a constraint on explanations of behavior that they should fit all reported thoughts or sensations into the overall account being offered'.⁶

This account is not meant to be taken seriously as a piece of folk anthropology, but it is meant to make it seem reasonable that we should treat self-ascriptions as having special authority. But the question remains: what reason has Rorty given to show that self-ascriptions not based on evidence concern the same states and events as ascriptions of the same mental predicates based on observation or evidence? There is a difference in kind in the ways the two sorts of ascription are made, and how they explain behavior is different. What Rorty describes as the discovery that self-ascriptions not based on evidence explain behavior better will be described by the skeptic as the fact that what is being ascribed is on every count apparently different.

It may come as a surprise to realize that the philosophers I am discussing have not really dealt with the ancient problem of skepticism concerning knowledge of other minds. But I think it is easy to explain. Historically the problem has been seen from either a Cartesian or an empiricist point of view, and both venues assume that each person knows what is in his own mind. The problem has therefore seemed to be that of supplying a basis for knowledge of other minds (and, of course, the external world). Philosophers now realize that part of understanding mental concepts (or predicates) consists in knowing what kind of observable behavior justifies the ascription of these concepts to others. But this answer to the skeptic does nothing to explain first person authority, or the asymmetry between self-ascriptions and other-ascriptions. We can still ask why we believe these two sorts of ascription pertain to the same subject matter. And

⁶ Richard Rorty, 'Incorrigibility as the Mark of the Mental', 416. Rorty's account is derived from Wilfred Sellars, 'Empiricism and the Philosophy of Mind'. With respect to the point at issue, Sellars's account does not differ from Rorty's.

this question is a good one, whether or not we recognize its traditional skeptical ancestry.

Perhaps it should be pointed out that no concepts aside from those applying to sensations, propositional attitudes, and the positions of our limbs show the sort of asymmetry we are discussing. Many concepts can be applied on the basis of multiple criteria, but no others are such that ascribers *must*, on particular occasions, use different criteria. If we are to explain this anomaly and avoid an invitation to skepticism, the explanation should point to a natural asymmetry between other observers and ourselves, an asymmetry not simply invented to solve the problem.

The first step towards a solution depends on becoming clear about the entities to which first person authority applies. William Alston proposes this principle to characterize the special status of self-attributions: 'Each person is so related to propositions ascribing current mental states to himself that it is logically impossible both for him to believe that such a proposition is true and not to be justified in holding this belief while no one else is so related to such propositions.'⁷

For this suggestion to be plausible, we must suppose that the proposition Jones expresses by the sentence 'I believe Wagner died happy' is the same proposition as the proposition Smith expresses by the sentence 'Jones believes Wagner died happy.' This is, of course, a highly questionable supposition. Once more, the epistemic contrast goes unexplained; and in the absence of an explanation, the question arises what *reason* one has in any particular case to believe that the proposition entertained by Jones and Smith *is* the same. Given only a description of an epistemic difference, the natural conclusion is that the propositions differ.

I turn, then, to a formulation of Sidney Shoemaker's, which makes explicit mention of language: 'Among the incorrigible statements are statements about . . . mental events, e.g. . . . reports of thoughts . . . These are incorrigible in the sense that if a person sincerely asserts such a statement it does not make sense to suppose, and nothing could be accepted as showing, that he is mistaken; i.e. that what he says is false.'⁸

I shall ignore the incorrigibility condition and substitute something less strong—something that amounts to first person authority. (This is perhaps reasonable, since Shoemaker is mainly concerned with sensations such as pain, while I am exclusively concerned with propositional attitudes.) What is important here is that Shoemaker assigns the presumption of correctness not to a kind of knowledge, but to a class of utterances. This idea might lead to an explanation of first person authority if the class of utterances could be specified in syntactic terms. Unfortunately it cannot. If Shoemaker is right, a speaker who sincerely uses a certain sort of sentence must be presumed to be right in what he

⁷ William Alston, 'Varieties of Privileged Access', 235.

⁸ Sidney Shoemaker, *Self-Knowledge and Self-Identity*, 215–16.

says. But of course this holds only if the speaker knows he is using the privileged sort of sentence; if he is not, he is misusing language. What would constitute a misuse here? Above all, one wants to say, sincerely asserting a sentence one has no special authority to assert. Perhaps so; but this is just to reiterate the uninformative and unexplained claim that it is a convention of language to treat self-ascriptions with special respect. Seen from the point of view of the interpreter, this implies that he should interpret self-ascriptions in such a way as to make them true—or to assign a special priority to their truth. The point of view of the interpreter is the only one we can take, given Shoemaker's principle, and this deprives the principle of independent application: our only reason for saying the speaker has special authority on occasion is that we are prepared to treat his utterance as a self-ascription. In other words, self-ascriptions have special authority: true; and that is where we began.

No satisfactory explanation of the asymmetry between first and other person attributions of attitudes has yet emerged. Still, focusing on sentences and utterances rather than propositions or meanings is a step in a promising direction. The reason for this is relatively simple. As long as we pose the problem in terms of the kind of warrant or authority someone has with respect to claims about an agent's attitude to a proposition (or a sentence with a given interpretation), we seem constrained to account for differences by simply postulating different kinds or sources of information. Alternatively, we may postulate different criteria of application for the key concepts or words ('believes that', 'intends to', 'wishes that', etc.). But these moves do no more than restate the problem, as we have seen, and thereby invite skepticism about knowledge of the minds of others (or of our own mind). But if we pose the problem in terms of relations between agents and utterances, we can avoid the impasse.

We now need to distinguish two related but different asymmetries. On the one hand, there is the familiar difference between self- and other-attributions of the same attitude to the same person: my claim that I believe Wagner died happy and your claim that I believe Wagner died happy. If these claims are put into words, we have the difficulty of deciding what pairs of utterances are suitably related in order to guarantee that the claims have the 'same content'. On the other hand, we may consider my utterance of the sentence 'I believe Wagner died happy', and then contrast my warrant for thinking I have said something true, and your warrant for thinking I have said something true. These two asymmetries are of course connected since we are inclined to say your warrant for thinking I speak the truth when I say 'I believe Wagner died happy' must be closely related to your warrant for thinking you would be speaking the truth if you said 'Davidson believes Wagner died happy'. But for reasons that will soon be evident, I shall deal with the second version of the asymmetry.

The question then comes to this: what explains the difference in the sort of assurance you have that I am right when I say 'I believe Wagner died happy' and the sort of assurance I have? We know by now that it is no help to say I have

access to a way of knowing about my own beliefs that you do not have; nor that we use different criteria in applying the concept of belief (or the word 'believes'). So let us simply consider a shorter utterance of mine: I utter the sentence 'Wagner died happy'. Clearly, if you or I or anyone knows that I hold this sentence true on this occasion of utterance, and she knows what I meant by this sentence on this occasion of utterance, then she knows what I believe—what belief I expressed.

It would once more make the account circular to explain the basic asymmetry by assuming an asymmetry in the assurance you and I have that I hold the sentence I have just uttered to be a true sentence. There must *be* such an asymmetry, of course, but it cannot be allowed to contribute to the desired explanation. But we can assume without prejudice that we both know, whatever the source or nature of our knowledge, that on this occasion I do hold the sentence I uttered to be true. Similarly, it would beg the question to explain the basic asymmetry by appeal to some asymmetry in our knowledge of the fact that I know what my sentence, as uttered on this occasion, meant. So again, let us simply assume we both know this, whatever the source or character of our knowledge.

So far, then, we have not postulated or assumed any asymmetry at all. The assumptions are just these: you and I both know that I held the sentence 'Wagner died happy' to be a true sentence when I uttered it; and that I knew what that sentence meant on the occasion of its utterance. And now there is this difference between us, which is what was to be explained: on these assumptions, I know what I believe, while you may not.

The difference follows, of course, from the fact that the assumption that I know what I mean necessarily gives me, but not you, knowledge of what belief I expressed by my utterance. It remains to show why there must be a presumption that speakers, but not their interpreters, are not wrong about what their words mean. The presumption is essential to the nature of interpretation—the process by which we understand the utterances of a speaker. This process cannot be the same for the utterer and for his hearers.

To put the matter in its simplest form: there can be no general guarantee that a hearer is correctly interpreting a speaker; however easily, automatically, unreflectively, and successfully a hearer understands a speaker, he is liable to serious error. In this special sense, he may always be regarded as interpreting a speaker. The speaker cannot, in the same way, interpret his own words. A hearer interprets (normally without thought or pause) on the basis of many clues: the actions and other words of the speaker, what he assumes about the education, birthplace, wit, and profession of the speaker, the relation of the speaker to objects near and far, and so forth. The speaker, though he must bear many of these things in mind when he speaks, since it is up to him to try to be understood, cannot wonder whether he generally means what he says.

The contrast between the grounds a self-ascriber has for his self-ascription, and the grounds an interpreter has for accepting that same ascription, would be stark

if we were to assume that no question can arise concerning a speaker's interpretation of his own words. But of course it can, since what his words mean depends in part on the clues to interpretation he has given the interpreter, or other evidence that he justifiably believes the interpreter has. The speaker can be wrong about what his own words mean. This is one of the reasons first person authority is not infallible. But the possibility of error does not eliminate the asymmetry. The asymmetry rests on the fact that the interpreter must, while the speaker doesn't, rely on what, if it were made explicit, would be a difficult inference in interpreting the speaker.

Neither speaker nor hearer knows in a special or mysterious way what the speaker's words mean; and both can be wrong. But there is a difference. The speaker, after bending whatever knowledge and craft he can to the task of saying what his words mean, cannot improve on the following sort of statement: 'My utterance of "Wagner died happy" is true if and only if Wagner died happy'. An interpreter has no reason to assume this will be *his* best way of stating the truth conditions of the speaker's utterance.

The best way to appreciate this difference is by imagining a situation in which two people who speak unrelated languages, and are ignorant of each other's languages, are left alone to learn to communicate. Deciphering a new language is not like learning a first language, for a true beginner has neither the reasoning power nor the stock of concepts the participants in the imagined situation have to draw on. This does not, however, affect the point I wish to stress, since what my imagined interpreter can treat consciously as evidence is exactly what conditions the first learner to be a language user. Let one of the imagined pair speak and the other try to understand. It will not matter whether the speaker speaks his 'native' tongue, since his past social situation is irrelevant. (I assume the speaker has no interest in training the hearer to cope with the speaker's original speech community.) The best the speaker can do is to be *interpretable*, that is, to use a finite supply of distinguishable sounds applied consistently to objects and situations he believes are apparent to his hearer. Obviously the speaker may fail in this project from time to time; in that case we can say if we please that he does not know what his words mean. But it is equally obvious that the interpreter has nothing to go on but the pattern of sounds the speaker exhibits in conjunction with further events (including, of course, further actions on the part of both speaker and interpreter). It makes no sense in this situation to wonder whether the speaker is generally getting things wrong. His behaviour may simply not be interpretable. But if it is, then what his words mean is (generally) what he intends them to mean. Since the 'language' he is speaking has no other hearers, the idea of the speaker misusing his language has no application. There is a presumption—an unavoidable presumption built into the nature of interpretation—that the speaker usually knows what he means. So there is a presumption that if he knows that he holds a sentence true, he knows what he believes.

15

A Nice Derangement of Epitaphs

Goodman Ace wrote radio sitcoms. According to Mark Singer, Ace often talked the way he wrote:

Rather than take for granite that Ace talks straight, a listener must be on guard for an occasional *entre nous* and me... or a long face no see. In a roustabout way, he will maneuver until he selects the ideal phrase for the situation, hitting the nail right on the thumb. The careful conversationalist might try to mix it up with him in a baffle of wits. In quest of this pinocle of success, I have often wrecked my brain for a clowning achievement, but Ace's chickens always come home to roast. From time to time, Ace will, in a jersksome way, monotonise the conversation with witticisms too humorous to mention. It's high noon someone beat him at his own game, but I have never done it; cross my eyes and hope to die, he always wins thumbs down.¹

I quote at length because philosophers have tended to neglect or play down the sort of language-use this passage illustrates. For example, Jonathan Bennett writes,

I doubt if I have ever been present when a speaker did something like shouting 'Water!' as a warning of fire, knowing what 'Water!' means and knowing that his hearers also knew, but thinking that they would expect him to give to 'Water!' the normal meaning of 'Fire!'²

Bennett adds that, 'Although such things could happen, they seldom do.' I think such things happen all the time; in fact, if the conditions are generalized in a natural way, the phenomenon is ubiquitous.

Singer's examples are special in several ways. A malapropism does not have to be amusing or surprising. It does not have to be based on a cliché, and of course it does not have to be intentional. There need be no play on words, no hint of deliberate pun. We may smile at someone who says 'Lead the way and we'll precede', or, with Archie Bunker, 'We need a few laughs to break up the monogamy', because he has said something that, given the usual meanings of the words, is ridiculous or fun. But the humour is adventitious.

¹ *The New Yorker*, 4 April 1977, p. 56. Reprinted by permission, 1977, The New Yorker Magazine, Inc.

² Jonathan Bennett, *Linguistic Behaviour*, Cambridge, 1976, p. 186. Donald Davidson, 1985.