

# Self-Knowledge: Discovery, Resolution, and Undoing

Richard Moran

This paper is concerned with ways in which some of the special features of self-knowledge relate to somewhat less familiar problems in the moral psychology of the first-person. The discussion is framed, in part, by an examination of Wittgenstein's remarks on Moore's Paradox, and I want to draw from these remarks some lessons about self-knowledge (and some other self-relations) as well as use them to throw some light on what might seem to be a fairly distant area of philosophy, namely, Sartre's view of the person as of a divided nature, divided between what he calls the self-as-facticity and the self-as-transcendence. I hope it will become clear that there is not just perversity on my part in bringing together Wittgenstein and the last great Cartesian. One specific connection that will occupy me here is their shared hostility to the idea of theoretical certainty as our model for the authority of ordinary self-knowledge, and their relating of such a theoretical model to specific forms of self-alienation. This, in turn, is related to another concern they share, a concern with the difficulties, philosophical and otherwise, in conceiving of oneself as but one person in the world among others. They share the sense, I believe, that while I recognize that I am a finite empirical being like anyone else, I must also recognize that the inescapable peculiarities of the first-person point of view oblige me to think of myself as both something more and something less than another empirical human being. The aim of this paper, however, is not to draw parallels between these two philosophers, but to develop the outlines of an argument concerning self-knowledge, one which re-locates some of its special features nearer to moral psychology than epistemology. But I find I need to draw on both writers to do so.

## I

Wittgenstein was not the first writer to note that, in normal circumstances, a person does not learn of his own attitudes by consideration of behavioural evidence. Typically, we do not need to interpret our own speech or other behaviour to determine what we think about something. To say this much is not yet to say anything about either the extent or the reliability of what we take ourselves to know about our own mental lives. Certainly nothing approaching Cartesian infallibility follows immediately from this independence from 'external evidence'. The idea of the independence of self-knowledge from such evidence is

often associated with the name of Descartes. However, this minimal idea does not originate with him, and the theses which are much more distinctively Cartesian ones are the substantive epistemic claims he made for the deliverances of introspection. The minimal claim for our purposes simply states that the ordinary reliability of first-person psychological discourse typically proceeds independently of what we may call 'outer evidence'; and *this* idea does not itself include either the claim of the self-intimating character of the mental, or the infallibility of the deliverances of introspection.

Even if this independence of self-knowledge from 'outer' evidence does not commit one to the full Cartesian claim of the transparency of consciousness to itself, it is still troubling enough for epistemologically-minded philosophers, and this disturbance is surely part of what motivates the tendencies to construe the faculty of introspection itself as providing the missing evidence; evidence now of a special quasi-perceptual kind which is better, more certain than any possible evidence available to any other person. From this vantage, then, we may see epistemic disquiet turn to despair in the face of Wittgenstein's persistent arguments in his later writings that first-person awareness cannot be a matter of any 'inner' evidence either. As I interpret these arguments, the claim is not that there is no such thing as a conscious mental event, nor is it denied that these are the sorts of things we may have unmediated awareness of (if all 'unmediated' means here is: not grounded in anything epistemically more basic). Rather, the point is that nothing in the way of an occurrent episode of consciousness could in principle provide one with the knowledge which one does indeed have of one's own thoughts. I'll have more to say about this later, but one fairly self-contained place to see this idea at work is in Wittgenstein's discussion from the *Blue Book* of imagining or thinking that King's College is on fire.<sup>1</sup> If we imagine the interior event as involving something like a good image (and how could it be *more* than that?) we may be brought to realize that even the best genuine picture could not reveal to us, could not be our evidence for, just *which* college it is that we are imagining on fire (that is, King's or another just like it), or indeed that what we are doing is *imagining* it on fire and not *remembering* it being on fire, or thinking of it in some other way. Yet, when a person imagines or remembers something these are the sorts of things we take for granted that he does know, and to deny such knowledge would amount to denying that people ever do succeed in imagining or remembering anything. And yet, so the argument proceeds, none of this knowledge could even in principle be based on 'inner' evidence of the sort provided by mental images or any other introspectible items of consciousness.

In this way then, bereft now of appeal to either 'outer' evidence or 'inner' evidence, various philosophers have felt driven to conclude (from these as well as other considerations) that what we think of as self-knowledge cannot sustain a 'substantial epistemology' at all; that is, that the very idea of self-knowledge as involving the apprehension of independent facts about one's mental life must be rejected. Thus, for example, Crispin Wright concludes that 'The roots of first-person authority for the self-ascription of these states [he mentions 'desire, belief, decision and intention'] reside not in cognitive achievement, based on cognitive

privilege, but in the success of the practices informed by this cooperative interpretational scheme.<sup>2</sup> On this view, it is just a defining feature of our concepts of certain psychological states that the person's own reports, however based, enjoy what he calls a 'positive presumption', barring evidence of self-deception or other distorting factors.

I will want to argue instead that we can see it as a rational requirement on belief, on being a believer, that one should have access to what one believes in a way that is radically non-evidential, that does not rely on inferences from anything inner or anything outer. What is meant by calling it a rational requirement will be made clearer later, but one impression I do mean to dispel is the sense of this independence of evidence as involving a kind of licence or courtesy we extend to people, a sense which encourages the sort of analysis of first-person authority in terms of social permissions that is found in Wright and elsewhere. The thought seems to be that if what we say about ourselves from the first-person perspective is not construed as based on evidence of *some* kind, then we cannot think of the thought thereby expressed as representing a cognitive achievement of any sort. And thus the expressions 'authority' and 'privilege' take on a kind of fully juridical meaning, in the sense of a decree that is answerable to nothing and no one. What I would like to suggest, on the contrary, is that it is the very answerability of first-person expressions, both to others and to certain requirements of rationality, that is responsible for their special status as contrasted with third-person reports. Such statements are not simply *allowed* to go by without the support of evidence; rather the kind of answerability they do have (to impersonal rational considerations) *obliges* them to be statements which do not rest on evidence. It will only be in situations of compromised rationality that the person will be in any need of, or be able to make any use of, evidence for what she believes or intends.

## II

Moore's Paradox can be seen as an emblem for peculiarities in the first-person point of view, specifically how the possibilities for thinking and talking about oneself are systematically different from the possibilities for thinking and talking about other people. An examination of it shows how the familiar appearance of first-person privilege is based on differences in the commitments and obligations of the first-person point of view. The paradox concerns statements of either of two forms: 1) 'P and I don't believe it', and 2) 'I believe that P, but not-P'. The initial problem, of course, is to say just *what* is supposed to be wrong with such statements, for the first thing pointed out in discussions of them is that they do not express formal contradictions, although they seem in some way absurd or incoherent. It is perfectly possible for both parts of such a statement to be simultaneously true, for they have different subject matters. The two versions given above describe possible situations a person might very well be in: the first describes the case in which there is some fact P of which one is unaware, and the

second represents one as believing something false. In fact, it should be clear that to say that these statements describe possible situations one might actually be in is to speak much too guardedly. For it's obvious from these paraphrases that they describe situations which one is certainly *always* in, for as long as there are any facts of which one is ignorant or any beliefs one has which are mistaken. This is not a condition one can reasonably expect to outgrow.

So the question is, if these sentences describe either possible or actually obtaining situations, what could be wrong with asserting them? Or indeed, what would be wrong with the unspoken *thought* that I believe it's raining outside and it isn't? Something surely *would* be wrong, both with this thought about falsity and the corresponding one concerning one's ignorance. If so, then this suggests that the prevalent diagnosis of Moore's Paradox as the paradigm of a 'pragmatic paradox' is at best inadequate, for the puzzling quality persists outside the context of utterance and speech act.<sup>3</sup> Were someone to think to himself, as he looks out the window, that it's raining outside, and conjoin this with the thought that he doesn't *believe* that it's raining, his thought would risk incoherence in just the same way as it would if he were to assert the whole thought to someone else. Since 'pragmatic' analyses like these are often associated with Wittgenstein, it is worth pointing out that nowhere in his various remarks on Moore's Paradox does he offer an analysis of it in terms of what we would call the pragmatics of assertion.<sup>4</sup>

A different view of the paradox, which he certainly does consider, but which I believe he rejects as an adequate account, in effect claims that such sentences are covert contradictions after all. The basis for this idea is the denial that first-person statements of the form 'I believe that P' are about the speaker as a believer at all, but are instead to be understood simply as ways of presenting the embedded proposition P. That is, when someone says, 'I think it's raining outside, his statement does not refer to his (or anyone else's) state of mind, but is instead simply a more guarded way of making the assertion about the rain. In this context, the word 'believe' is not operating as a psychological verb at all. On this view, then, the two parts of the Moore-type statement do *not* in fact have different subject-matters (one part about the rain, the other about someone's belief), and hence their conjunction really does form a contradiction. And so what was puzzling about the original statement reduces to the fact that it is a contradiction after all, but in a disguised form. We may call this the Presentational view, since its central idea is that in the first-person present-tense the verb-phrase 'I believe' is not in fact psychological, but rather represents a mode of presenting the relevant proposition which follows it. It is this type of view that Wittgenstein is alluding to when, for instance, he concludes one line of thought with 'Don't regard a hesitant assertion as an assertion of hesitancy' (1956, p. 192). That is, we are to see the hesitancy, expressed by the apparent reference to one's belief, as qualifying the assertion about the rain, and not as describing anyone's state of mind.

However, to ascribe the Presentational view to Wittgenstein one would have to understand this passage and related ones as not just warning against a confusion we may be prone to, but as claiming that, for instance, hesitancy can apply

only to assertions and not to persons and their states of mind. This is not what he says; and had it been what he meant, it would have made less sense to warn against confusing one thing with another than simply to declare that the very idea of an assertion of one's own hesitancy (or doubt, or conviction) is an illusion. But there certainly are situations in which one does intend to make an assertion about one's own hesitancy, or one's conviction, or one's belief as a fact about oneself. To insist on this much does not depend on any metaphysics of 'inner states', or any ideas about what, if anything, may constitute a person's belief. And to deny it would amount to saying that while you can talk and think about the psychological life of other people, you are peculiarly barred from doing so in your own case. There is no need to underestimate the profound differences of such thought and talk in the two cases, but the task of understanding these differences can only be short-circuited by an analysis which implies that you cannot think or talk about the very same matters that another person does when he talks about what you believe or intend.<sup>5</sup> Were this the case, you would not simply lack first-person authority; rather you would be unable so much as to entertain the thought that there is something you believe or something you want, nor would it be possible for you even to have *mistaken* beliefs about such matters. There would be one person in the world whose psychological life would not be so much as darkness to you, for you could not even turn your attention to it.

Now I take it that it *will* be part of Wittgenstein's view that what we call first-person privilege is systematically connected with privation or disadvantage of a certain sort, but this does not consist in being unable to think of oneself as a psychological subject. There are instead special conditions on conceiving of oneself as a psychological subject, which do not apply to one's relations to other people. Wittgenstein expresses one such condition in one of his more well-known remarks on Moore's Paradox, when he says, 'If there were a verb meaning "to believe falsely", it would not have any significant first-person present indicative' (1956, p. 190). As before, whatever problem there would be for the significance of such a first-person statement of false belief would apply equally to the corresponding first-person *thought*. If so, then *certain* attitudes toward the person as a psychological subject will have their home in our relations with others, and will have at best only some (philosophically and psychologically) problematic application to oneself. The special status of the first-person position is not exclusively one of favour or authority. But it is not the case that the very idea of conceiving oneself psychologically (e.g., as a subject of belief) is ruled out for grammatical or other reasons. Some have seen such a view as a convenient way to dissolve the philosophical problems of self-knowledge, but it is a heavy price to pay, and Wittgenstein himself explicitly repudiates the conclusion that there is no properly psychological use of verbs like 'believe' in the first-person present-tense.<sup>6</sup>

For all that, we can still agree that the normal function of the first-person present-tense of 'believe' is to declare one's view of how things are, out there in the world beyond oneself, and this stems from the fact that to believe some proposition *just is* to believe that it is true. This connection between belief and truth is responsible for another related feature of the avowal of belief that was

suggested by Wittgenstein, and later taken up by Gareth Evans and others.<sup>7</sup> If my beliefs just are what I take to be true, then when I am asked what I believe about something I will answer this by directing my attention to the world independent of my mind. When asked whether one believes Oswald acted alone, one normally responds by attending to facts about *Oswald* etc., and does not scan the interior of one's own consciousness. This feature is sometimes called the 'transparency' of one's own thinking, in that, for me, a question about my belief is 'transparent' to a question which is not about me but about the world, and is answered in the same way. Thus I can answer a question about my belief by directing my attention to what is independently the case, and not by considering evidence, behavioural or otherwise, about anyone's state of mind.<sup>8</sup>

However, whatever the implications of this relation of transparency, Wittgenstein does not deny that the two questions may relate to what are different matters of fact. Nor does it even mean that a person cannot admit the difference between the two from within the first-person point of view. It is, after all, one thing for it to be raining outside, and quite another for me (or anyone else) to believe that it is. There are in this way two quite different types of commitment involved in my avowing a belief of mine. On the one hand, in saying 'I believe it's raining outside' I commit myself to the state of the weather's being a certain way. My avowal of this belief expresses the fact that it is not an open question for me whether it is raining or not. At the same time, however, I must acknowledge myself as a finite empirical being, one fallible person in the world among others, and hence that my believing something is hardly *equivalent* to its being true. And even when a person's fallibility is not the issue, anyone must recognize that his believing P is nonetheless an additional fact, distinct from the fact of P itself.

Neither commitment is avoidable, although it is clear that they can pull one in different directions. Insofar as I recognize myself as a finite human being, I must acknowledge that the question of what my belief is concerns an empirical matter of fact distinct and independent from the question about the object of belief. For me to deny this, either implicitly or explicitly, would be for me to deny that what my beliefs are about, what my attitudes are directed toward, is an independently existing world. And yet here I am, assuming transparency, answering the question about my belief as if these were *not* distinct matters; *without* directing my attention either to the behavioural evidence or the inner state of the person whose beliefs I'm reporting. It may thus seem like there is some evasion, or at least some questionable indirection involved here, if I answer a question about one subject matter by means of reflection on another logically independent one. Or: it may seem that my stance toward the question of what my beliefs are does implicitly deny the fact that, on one level at least, this concerns a matter of empirical psychological fact.

Whether or not there is any implicit denial here of one's status as an empirical human being, it does seem appropriate to distinguish between different levels at which one conceives oneself as a psychological subject. To believe some proposition is to take it to be true. And of course, one also takes other people to have true beliefs sometimes. But the beliefs of other people represent facts (psychological

facts, to be sure) on the basis of which one may make up one's mind about some matter, whereas one's own beliefs just *are* the extent to which one's mind *is* (already) made up. That is, the beliefs of another person may represent indicators of the truth, evidence I may infer from to some conclusion about the matter. I may trust them or mistrust them. With respect to my own beliefs, on the other hand, there is no distance between them and how the facts present themselves to me, and hence no going from one to the other. It is for reasons of this kind that Wittgenstein says, 'One can mistrust one's own senses, but not one's own belief' (1956, p. 190). What this must mean is not that I take my beliefs to be so much more trustworthy than my senses, but that neither trust nor mistrust has any application here.<sup>9</sup> One way to express this might be to say that, in any particular case, it is a fully empirical question for me whether my own senses or another person's beliefs reveal the facts as they are. And even when my confidence in either one of them is complete, this itself will be an empirical matter, based on various things I may know about either my own senses or about the other person, and including considerations of trust, evidence and reliability.<sup>10</sup> Whereas, from the first-person point of view, the relation between one's belief and the fact believed is not evidential or empirical, but rather categorical. That is, to speak of one's belief just is to speak of one's conviction about the facts, and not some additional thing one might be convinced by. Hence it is quite a different matter to take one's own belief about something to be true and to take someone else's belief to be true, even when these beliefs concern the very same proposition.

Some such distinction is necessary if we are to understand Wittgenstein's remark as something other than a declaration of our greater complacency in the relation to our beliefs than in the relation to the deliverances of our senses; or, say, as describing the matchless confidence one has in one's own judgement which cannot be approximated by the confidence one has in anyone else's.<sup>11</sup> Instead, what is being described here is the distinction in kind or category between something one may treat as evidence on which to base one's judgement, and the judgement itself which one arrives at. As an empirical matter, the fact of anyone's believing P leaves open the question of the truth of P itself, although another person may close this opening by inferring from a psychological fact to a non-psychological one. But for the person herself, if her own belief that it is raining outside does not constitute the question's being settled for her, then nothing does. To have beliefs at all is for various questions to be settled in this way. Referring to a *categorical* rather than an empirical relation here is a way of saying that to be a believer at all is to be committed to the truth of various propositions. The possibility of scepticism does not belie this categorical relation between belief and truth, for the sceptic renounces belief itself and the commitment it entails. (Doubting, then, will bring with it *different* categorical commitments.) Nor is this relation absent from those situations where I have reason to think that my judgement in some matter is skewed and I therefore have less than total confidence in my own opinion. For in that case the relation between belief and truth expresses itself in the fact that insofar as I lack confidence in my judgement about X, I have no settled belief about it. For conceptual reasons, the degree of mistrust will entail

a corresponding qualification in the original attribution of the belief to me. This is not an empirical matter, and it doesn't apply to one's relation to the beliefs of others.

From the first-person perspective, then, what is unavoidable is the connection between the question about some psychological matter of fact and a commitment to something that goes beyond the psychological facts.

### III

This distinction of levels, and the possibilities of tension between them in one's relations with oneself, are thematic in the early work of Sartre. In Sartre's language, consciousness of anything at all involves a kind of negation, a distancing or distinguishing of the subject of consciousness from its object. Indeed, he will sometimes say that consciousness just is this 'negating' of the in-itself (the world of facticity) by the for-itself (the conscious subject which 'transcends' that world). Consciousness of *oneself* will involve this same distancing and separation of the 'transcendental' subject from that aspect of the person which such self-consciousness is directed upon. Hence he normally identifies the object of self-consciousness with the aspect of the person as a 'facticity' rather than as 'transcendence', that is, with one's status as an empirical psychological subject, a particular human being. And so, for him, self-consciousness of one's own belief involves a distancing of oneself from the perspective of the declaration or endorsement of one's belief.

If I *believe* that my friend Pierre likes me, this means that his friendship appears to me as the meaning of all his acts. [. . .] But if I know that I believe, the belief appears to me as pure subjective determination without external correlative. This is what makes the very word 'to believe' a term utilized indifferently to indicate the unwavering firmness of belief ('My God, I believe in you') and its character as disarmed and strictly subjective ('Is Pierre my friend? I do not know; I believe so').<sup>12</sup>

From a purely empirical point of view, the fact of one's belief is just a fact about one's psychological life, like anyone else's, a 'pure subjective determination'. As such it bears no special relation to the truth, or only one of aspiration, and fallibility is built in to the attribution of any such attitude. Metaphors of interiority are characteristic of this perspective, of course (one state going on inside me, which may or may not correspond to what is going on outside) but they are not essential to it. The tension in question can be described without recourse to the idiom of the inner and the outer. What is crucial is the distinction between what is true of the person, and what truth independent of the person she is thereby committed to.

Sartre's case of the akratic gambler who resolves to stop gambling is in some ways a more helpful example for considering the two stances and the contrasting

roles of commitment (*of oneself*) and theoretical knowledge about oneself. For the gambler to have made such a decision to quit is for him to be committed to not going to the gaming tables. He is committed to this truth categorically, that is, insofar as he actually has made such a decision, *this* is what it commits him to. For him his decision is not just (empirical) *evidence* about what he will do, but a resolution which he is responsible for. But now at the same time he does know himself empirically too, and from this point of view his 'resolution' is a psychological fact about him with a certain degree of strength. And it is the psychological strength of this resolution that will justify any theoretical expectation that he actually will avoid the gaming tables. From this theoretical point of view on his (past) resolution (as facticity now, rather than as transcendence) it seems then an ungrounded, inconstant thing on which to base any confidence about what he will in fact do. In Sartre's view, his relation to his decision is transformed when it becomes for him an empirical object of consciousness, and he relates to it as facticity rather than identifies with it as a transcendence.

The resolution is still *me* to the extent that I realize constantly my identity with myself across the temporal flux, but it is no longer *me* – due to the fact that it has become an object *for* my consciousness. [ . . . ]

It seemed to me that I had established a *real barrier* between gambling and myself, and now I suddenly perceive that my former understanding of the situation is no more than a memory of an idea, a memory of a feeling. In order for it to come to my aid once more, I must remake it *ex nihilo* and freely.<sup>13</sup>

Contrary to what the wording of this passage may suggest, it is not simply the fact that his resolution has become an object for his consciousness that makes it suddenly seem something ephemeral. After all, the awareness of another person's resolution (as empirical fact about him) need not diminish its endurance or reality in my eyes.<sup>14</sup> Indeed, as empirical realities the acts and intentions of other people may present themselves with the status of something more secure and law-like than one's own perpetually modifiable resolutions could ever hope to achieve. So it should not be assumed that it is apprehension of the empirical aspect alone that makes the gambler's resolution appear to him as less than reliable. Rather, his anxiety is provoked by a disengagement from his resolution, a hedging of his endorsement of it, combined with the simultaneous desire to rely on it like a natural fact. He relates to his resolution as something independent of him, like a machine he has set in motion and which now should carry him along without any further contribution from him. He seeks confidence about his own future behaviour at the empirical level, but then realizes that any such theoretical confidence is utterly inadequate on its own to settle his mind, because it can only be totally parasitic on his practical-transcendental resolution. His problem lies in the fact that if he *needs* empirical support, that can only be because he feels that his resolution is not strong enough on its own. And he now realizes that the empirical perspective cannot provide any additional strength of its own, for all of

it is borrowed from the strength of the resolution itself. Hence the resolution itself appears as inadequate ('no more than the memory of an idea, a memory of a feeling') and at the same time he feels helpless to provide it with any additional strength, because he is seeking it in the wrong place. With his attention diverted from the practical reasons which issued in the resolution, and considering it now purely psychologically, there seems nothing especially compelling in it.

What, then, does being psychologically realistic about oneself mean in such a situation? One of Sartre's themes is the thought that I cannot simply *accept* the theoretical conclusion, however empirically well-grounded, without this opening me to the charge of indulging in *acquiescence* in my weakness under cover of being hard-headed and without any illusions about myself. What I am aware of concerning myself empirically, as a facticity, cannot substitute for what I am committed to categorically; commitments I have simply in virtue of having any beliefs about the world or any decisions about my action. Being 'without illusions' about oneself is precisely a theoretical relation, and someone for whom this is the supreme virtue, the one unchallengeable imperative, will be to that extent in a relation of bad faith and evasion with respect to his beliefs and his actions. (This has not prevented the interpretation of the early Sartre as a champion of just such a 'heroism of disillusionment'.)

These considerations may also provide us with the beginnings of an explanation for why *akrasia* can be such a peculiarly corrosive condition. For it often begins with the tactical substitution of the theoretical point of view for the practical one (e.g., telling myself that I'm bound to backslide at some point, so it's better that I do so on this occasion in a controlled and self-aware manner). But soon my accumulated history of backsliding provides more and more good theoretical evidence for predictions of my future conduct which conflict with what I decide to do. What gets obscured is the provisional status of the original 'tactical substitution' which was, after all, presented in the guise of a practical concern, to minimize the harm of anticipated backsliding. My repeated so-called decisions now are poor things both epistemically, as indicators of the future, and practically, as promoters of my well-being. While at the same time, the theoretical perspective is successful on its own terms (predictively), as well as presenting itself as looking out for the efficacy of my decisions. Hence my theoretical understanding of myself is shown to be much more reliable than my practical self-understanding, and my akratic behaviour itself provides more and more evidence of the superiority of the theoretical over the practical point of view on my decisions and my future.

But even when I am confident about myself, and not worried about backsliding, this does not mean that I can be complacent about my resolution as constituting a real empirical barrier between gambling and myself, for I must recognize that the resolution remains mine to keep or to break at any time. For me the fact of my resolution cannot be something on which I may base a confident prediction, because I must recognize that the resolution exists as a fact for anyone to base a prediction on only insofar as I endorse it. Any my endorsement is in principle answerable to how I relate myself to the reasons in favour of some

course of action. My resolution is only as strong as my hold on those practical reasons. If they seem insufficient to me such that I seek to avail myself of empirical, predictive reasons for my confidence that I will abstain, then to that same degree I lose the empirical basis in fact for making that prediction, for I thereby reveal myself as *unresolved* about the question.

And naturally these limitations in the empirical relation to oneself do not apply to another person's relation to one's resolution. For the other person there is no demand that predictive reasons defer to practical ones. With respect to beliefs the parallel asymmetry would be the instability in the idea of trust or mistrust applied to one's own belief, in the sense of treating the empirical fact of one's *having* the belief as evidence for its truth. If a generally reliable person believes that it's raining outside, that fact *can* certainly be taken as evidence for rain. But in my own case, as with the resolution not to gamble, I must recognize that the belief is mine to retain or to abandon. (In Sartre's language, 'I posit my freedom with respect to it.') That is, my belief exists as an empirical psychological fact only insofar as I *am* persuaded by the evidence for rain, evidence which (prior to my belief) does not include the fact of my being persuaded. If I am unpersuaded enough to need additional evidence, then by virtue of that psychological fact itself I lose the empirical basis for any inference from a person's belief to the truth about the rain. For someone's *unconfident* belief about the rain provides that much less reason for anyone to take it to be good evidence for rain itself.

#### IV

Moore's Paradox provides us with a kind of paradigm formula for difficulties arising from the introduction of an empirical or theoretical point of view whose deliverances may clash with what we want to say from the point of view of a rational deliberator, a point of view 'transcendent' with respect to the psychological facts as currently constituted. For empirically, I can well imagine the accumulated evidence suggesting both that I believe that it's raining outside, and that it is not in fact raining outside. Theoretically these are perfectly independent matters of fact, and I can in principle recognize the possibility of their co-occurrence, just as I can imagine my future conduct clashing with what I now decide to do. But, as I conceive of myself as a rational agent, my awareness of my belief is awareness of my commitment to its truth, a commitment to something that transcends any description of my psychological state. And the expression of this commitment lies in the fact that my reports on my belief conform to the condition of 'transparency' mentioned earlier: that I can report on my *belief* about X by considering (nothing but) X itself. Hence a situation that one can conceive of as a theoretical, empirical possibility clashes with the conception of oneself as a rational agent. And this clash is not avoidable by opting out of this conception of oneself; for were one *not* a rational agent, there would be no psychological life to have empirical views about in the first place.

When a person's relation to her belief conforms to the transparency condition, then the belief is expressed by reflection on its subject matter, and not by consideration of the psychological evidence for a particular belief attribution. That is to say, one is not treating the belief-attribution to oneself as a purely empirical or theoretical matter. We have seen that, were it to make sense for one to take such a purely theoretical view of oneself, then the thought expressed in a Moore-type sentence would describe a perfectly coherent empirical possibility which one could sensibly report on. Hence we will gain a better understanding of the limitations of the theoretical or empirical point of view on the self by examining the situation in which transparency fails, and hence one's relation to oneself approaches a fully empirical one. The limits on the empirical point of view on oneself will in turn shed light on how it can be, among other things, a *rational* requirement that one have a kind of access to one's beliefs that is not based on evidence of any kind.

In certain psychoanalytic contexts, for instance, the manner in which the analysand becomes aware of various of her beliefs and other attitudes does not necessarily conform to the transparency condition. Some such attitudes (e.g., the belief that one has been betrayed by one's brother), may be available *only* through the eliciting and interpreting of evidence of various kinds. The person might become thoroughly convinced, both from the constructions of the analyst, as well as from her appreciation of the evidence, that this attitude must indeed be attributed to her. And yet, at the same time, when she reflects on the world-directed question itself, whether she has indeed been betrayed by this person, she may find that the answer is no, or can't be settled one way or the other.<sup>15</sup> So transparency fails because she cannot learn of this attitude of hers by reflection on the object of that attitude. She can learn of it only in a fully theoretical manner, taking an empirical stance toward herself as a particular psychological subject.

In such a case we might say that the analysand can *report* on such a belief, but that she does not *express* it, since although she will describe herself as feeling betrayed she will not in her present state affirm the judgement that this person has in fact betrayed her. When the belief is described, it is kept within the scope of the psychological operator, 'believe'; that is, she will affirm the psychological judgement 'I believe that P', but will not avow the embedded proposition P itself.

In various guises, some such general distinction between reporting and expressing will be more or less familiar. In an interesting recent paper on *akrasia* and self-deception, Georges Rey (1988) has developed similar terms for describing the difference between beliefs and other attitudes which I become aware of by virtue of their *explanatory* role, and beliefs which I am aware of because I *avow* them, that is, explicitly endorse them. On this view, a *report* on an attitude of mine has an explanatory basis, and need not imply a commitment to its truth or justification, any more than its third-person equivalent would. An *avowal* of one's belief, by contrast, is not made on any psychologically explanatory basis, and is rather the expression of one's present commitment to the truth of the proposition in question. Rey describes both self-deception and *akrasia* not as involving conflict between first- and second-order attitudes, but as constituted by conflict between

attitudes which one avows and those he calls one's 'central' explanatory attitudes. Much as I think there is to gain from considering Rey's general approach, a central problem with it is that he understands the distinction drawn here to be a distinction between two different *kinds* of attitude, the 'central' and the 'avowed' ones. And in formulating the distinction he makes the comparison between our psychological terms like 'belief', and natural kind terms like 'jade' which has a divided reference between two distinct natural kinds of stone. Just as we can greet this discovery by simply saying that there are two kinds of jade, so we may say that each psychological attitude may come in either a central or an avowed variety.

But if the beliefs which I express when I avow them (either by saying 'P' or 'I believe that P') are simply of a different *kind* from the beliefs and other attitudes which are the central explanatory ones, then it is completely unclear how we may see the two as *clashing* at all. If they are anything at all, conditions like *akrasia* and self-deception are some kind of *conflict* within the person, expressive of conflicted relations to the same thing, and this sense is lost if we see the avowed belief that P and the central explanatory belief that not-P as distinct attitude types, and each all right in its own way. This would simply reinstall an unanalysed instance of Moore's Paradox, and leave us no way to say what's wrong. That is, I may *report* on the 'central' explanatory attitude that I feel betrayed, while I *avow* the belief that I have *not* been betrayed. (I look inside and see one thing, and I look outside and see another.) But avowing and reporting cannot be thus isolated from each other, if for no other reason than that any avowal is itself behaviour, and thus evidence for the *explanatory* attitude of belief. Hence the 'two attitudes' could never in principle be of utterly distinct types.

Preserving the sense of conflict within the self, whether in a psychoanalytic context or not, requires that we see the meaning of a psychological term like 'belief' as univocal across the two contexts. Only thus can we remain open to the idea that, although when I avow my belief I do not avail myself of psychological evidence of any kind, I nonetheless take what I say in that context to be *answerable* to the whatever psychological evidence there may be. In this light, compare Anscombe (1976) on two possible stances toward an expression of intention. A well-known remark of hers says that 'If a person says "I am going to bed at midnight" the contradiction of this is not: "You won't, for you never keep such resolutions" but "You won't, for I am going to stop you" ' (p. 55). Since an expression of intention is not a prediction, it is not contradicted by a contrary prediction. This is not, however, to deny that in declaring the intention the person is committed *both* to the practical endorsement of the action and the expectation of a future event. And the hearer is thereby told something about the future which he may doubt, or count on like a prediction. This is recognized by Anscombe later, although not explicitly related to the earlier remark, when she says: 'If I say I am going for a walk, someone else may know that this is not going to happen. It would be absurd to say that *what* he knew was not going to happen was not the very same thing that I was saying *was* going to happen' (p. 92). The person who announces that she is going for a walk does not base her statement on evidence;

and yet her statement may be true or false, and for her interlocutor deciding on its truth or falsity *will* be a matter of evidence.

We should, then, see the stance of avowal and the stance of explanation as two ways of coming to know the same thing, and the Sartrean and Kantian language employed earlier is meant to convey this idea of different stances toward what is in some sense the same state of affairs, the same person. When I *avow* a belief I am not treating it as just an empirical psychological fact about me; and to speak of a *transcendental* stance toward it is meant to register the fact that it is explicit in the avowal that it commits me to the facts *beyond* my psychological state; and as a commitment it is not something I am assailed by, but rather is mine to maintain or revoke.

If, then, we are looking at two routes to knowledge of the same facts, we must ask: what would be missing from a restoration of self-knowledge that remained theoretical or descriptive in this sense? Why should one route, the ordinary non-evidential way of avowal, be privileged over some possible theoretical route, so long as we arrive at knowledge of the same state of affairs? It is virtually definitive of psychoanalytic treatment that it does not begin by taking first-person declarations as necessarily describing the truth about the analysand's actual attitudes, and this might be taken to mean that the knowledge of oneself it seeks to culminate in will dispense with avowal as something unsophisticated and unreliable, and substitute for it something more interpretive and theoretically grounded. This picture, however, neglects the crucial therapeutic difference between the merely 'intellectual' acceptance of an interpretation, which will itself normally be seen as a form of resistance, and the process of working-through leading to a fully internalized acknowledgement of some attitude which makes a felt difference to the rest of the analysand's mental life.<sup>16</sup> This goal of treatment, however, requires that the attitude in question be knowable by the person, not through a process of theoretical self-interpretation but by avowal of how one thinks and feels. We will get a better sense of what's inadequate about 'merely intellectual' acceptance of an interpretation, by considering what an idealized, but still purely theoretical, relation of expertise toward oneself would be missing. We would then understand better why it is not just permissible but essential that ordinary first-person knowledge proceed independently of any evidence, and hence why a non-empirical or transcendental relation to the self is ineliminable.

The deliverances of the 'ideal symptomatic stance', as we are imagining it, could be as spontaneous as the most basic judgements we make about the world. Nor need there be any loss of special reliability: what is learned about one's attitudes in this way might well be so certain and so complete as to be unchallengeable by anyone else. Further, we need not think of the evidential basis here as restricted to anything like overt behaviour; but may also include one's dreams, passing thoughts, associations and feelings. That is, theoretical expertise as we are imagining it here could extend to the private 'inner' realm as well, such that only I could report on it and my reports were invariably accurate. And yet, the claim is, such an epistemic capacity would still not provide what is known from within ordinary first-person knowledge. What the consideration of such a case

provides us with is a full-blown realization of the theoretical, or perceptual, picture of self-consciousness that both Wittgenstein and Sartre are concerned to combat. And one irony of this critique is that part of the *criticism* of this picture – that is, the picture of an inner eye focused upon essentially inner events which no one else in principle could witness and which I am unable to reveal or describe to another person – is that the full-blown metaphysical picture of privacy here essentially *underestimates* the radicalness of the difference between my relation to myself and my possible relations to others.

For in essence what we have here is a picture of self-knowledge as a kind of mind-reading as applied to oneself. Hence in imagining the manner of the person's self-knowledge we may even dispense with any work of interpretation on her part altogether. That is, she may know these psychological facts with immediacy, in a way that does not depend on any external 'medium', and which involves no inference from anything else, and yet she would still not enjoy ordinary first-person knowledge. What the distinction between the (empirical) explanatory stance and the stance of avowal enables us to see is that the ordinary first-person point of view combines two different features which are not commonly distinguished. It is part of the ordinary first-person point of view on one's mental life that behavioural evidence is not consulted, *and* in addition that the expression of one's belief here carries with it a commitment to its truth. Thus, from the first-person point of view a claim concerning one's attitudes also counts as a claim about the world they are directed upon. But we can now see that these two aspects of the first-person point of view can in principle come apart. Insofar as it is possible for one to adopt an empirical or explanatory stance towards one's beliefs, and thus to bracket the issue of what their possession commits one to, it will be possible for one to adopt this stance to anything theoretically knowable, including private events or attitudes that one may be somehow aware of immediately, without inference. The kind of alienation we have been picturing here would remain possible on even the most generous epistemology for self-knowledge, so long as it was construed purely theoretically. We may allow any manner of inner events of consciousness, any exclusivity and privacy, any degree of privilege and special reliability, and their combination would not add up to the ordinary capacity for self-knowledge. For the connection with the *avowal* of one's attitudes would not be established by the addition of any degree of such epistemic ingredients.

The picture of gaining knowledge of one's attitudes from an explanatory perspective, and the possible distance of the person from the attitudes reported on in this way, is more graphic and is easier to latch on to when we imagine the case as involving a second person. This might be an expert of some sort whom I trust, interpreting the evidence in such a way as to ascribe a belief to me, which I am not prepared to endorse as true. What the foregoing considerations are meant to show is that these external features of the empirical or explanatory stance are inessential to it. Such theoretical knowledge of oneself would still not be properly first-personal even if this situation were fully 'internalized' and involved no one other than the person herself, reflecting on nothing other than the contents of her

consciousness. Beyond simply being *aware* of the attitude in question (by whatever mode of awareness) there would remain outstanding the question of her endorsement of it. And no further amount of psychological knowledge about the attitude entails such endorsement. But without it, her awareness of this attitude still doesn't express her actual point of view on the world, doesn't provide her with *reasons* for action or belief. From this stance one may as well be talking about *anyone's* attitude. And as far as the first-person phenomenology of the case is concerned, the person may as well be reporting on the voices in her head. However loud they may be, they do not count as providing the person with the self-knowledge she seeks, concerning what she believes or what she wants. And this is so precisely because of their independence from considerations in favour of, or against, this or that attitude.

Any treatment that took upon itself the task of restoring self-knowledge would have to restore a capacity that conformed to what we've been calling the condition of transparency. For whatever expert knowledge a person may gain of herself, there would still be something seriously wrong, for instance, with her relation to *her brother*, so long as she could not become aware of her attitude toward him by reflection on that very person, and was instead restricted to reflecting on either her own internal or external states, however immediate or effortless this reflection might be in practice. The rationality of her response requires that she be in a position to *avow* her attitude toward him, and not just describe or report on it, however accurately. For it is only from the position of avowal of her attitude toward her brother that she is necessarily acknowledging facts about *him* as internally relevant to that attitude (say, as justifying or undermining it); and thereby (also) as relevant to the fully empirical question of whether it remains true that she indeed has this attitude of resentment. From the purely theoretical or explanatory point of view she might respond to the information that her sense of betrayal is in fact groundless by saying 'Quite so, but how is that relevant to the psychological question of what my attitude [Rey would say her "central" attitude] actually is? Surely I *do* feel betrayed by him. We both agree that this is the unavoidable conclusion to draw from the rest of my thoughts, feelings, and behaviour. And I'm as certain of that as ever.'

The point of referring to such things as 'commitment' and 'endorsement' here has been to show the limitations for self-knowledge of this kind of 'certainty'. We may compare the case of belief with the case of decision and the two different possible ways of knowing what one will do. We may imagine Sartre's gambler as in a state such that he can only learn of the strength of his resolution to avoid the gaming tables by empirical psychological reflection on himself, and *not* by practical reflection on the *reasons* he has for not gambling. To the extent that he cannot become sure of what he will do by reflection on (normative) practical considerations rather than predictive considerations, he does not take the commitment or direction of his will to be a function of what his best practical reasons are. His actual will, such as it may be, is not understood by him as a response to those reasons, and what determines its strength or weakness is seen as something independent of them.

What is wrong with 'direction of gaze' here, the shift to the theoretical or empirical perspective, is that it suggests that his reflection on his best reasons for belief or reasons for action still leaves it an open question what he will actually end up believing or doing. This is not a stable position one can occupy and continue to conceive of oneself as a practical and theoretical deliberator. One must see one's deliberation as the *expression and development* of one's belief and will, not as an activity one pursues in the *hope* that it will have some influence on one's eventual belief and will. Were it generally the case (for Sartre's gambler, say) that the conclusion of his deliberation about what to think about something left it still open for him what he does in fact now think about it, it would be quite unclear what he takes himself to be *doing* in deliberating. It would be unclear what reason was left to *call* it deliberation if its conclusion did not count as his making up his mind; or as we sometimes say, if it didn't count as his coming to *know* his mind about the matter.

Nothing further in the way of *evidence* about himself could establish the connection between his reflection on the world and his knowledge of what he will (in fact) do or what he does in fact believe. Considering himself empirically (or theoretically), he must see these as quite separate questions. Nothing but some commitment on his part could make it the case that his world-directed reflection on the reasons available to him for believing something or doing something does indeed settle the question of what he does believe or what he will (in fact) do.

## V

Another well-known sentence from the *Investigations* concerning Moore's Paradox is Wittgenstein's claim that 'If there were a verb meaning "to believe falsely", it would not have any significant first-person present indicative' (p. 190). This statement also occurs in the first volume of the *Last Writings in the Philosophy of Psychology* (§ 141), but later in that volume there also occurs the claim that 'If there were a verb "to seem to believe" then it would not have a meaningful first-person in the present indicative' (§ 423). On the face of it these represent very different ideas. The first is an expression of the internal relation between belief and truth that has been our guide throughout. And as such it is a conceptual claim about belief that, I think, even those most suspicious of conceptual claims in general would want to concede. But the second statement appears to claim something similar for the relation between belief and *awareness* of belief, and that is something that even someone well-disposed to the grammatical or the conceptual might not be willing to grant. For it seems that it could only be true that it makes no sense for me to say 'I *seem* to believe –' something, if it makes no sense for me to be in a state of empirical uncertainty about what my belief is.

Speaking for myself, I should say that this is not what I've tried to argue here, but rather that the empirical point of view on oneself cannot be a stable or dominant position. And I think we can read these two statements of Wittgenstein's, and understand the relation between the two of them, as expressions of the

priority of what I've been calling the transcendental over the empirical point of view on the self. That is, they are statements which outline limits on the applicability of the explanatory, theoretical stance toward oneself. However, as with such statements in general, the difficulty is in determining from what stance they themselves are made, from what position one can declare the priority of one totalizing stance over another. For the critique must recognize that the empirical, theoretical attitude is perfectly consistent within itself and acknowledges no limits to its domain. It is, after all, the total realm of empirical fact, and that includes all the people in the world, including oneself. From within the explanatory attitude any talk of 'limits' to its applicability can only appear as evasion and a counsel to ignorance or denial. Any epistemic considerations will simply be part of this response, and no *non*-epistemic considerations can establish their validity within this domain. On the other hand, from within the transcendental point of view of oneself, the *refusal* to acknowledge limits in principle to epistemic or evidential considerations can only appear as evasion of another kind; that is, abdication of the responsibilities of being a rational epistemic agent. From within this stance, the idea of limits to the validity of stepping back from oneself as a presently constituted psychological subject can only appear as counsel to complacency, as a limit to the applicability of rational criticism and endorsement to one's beliefs.

\* \* \*

It is not essential to whatever is missing from the theoretical perspective described above that there be any lack of certainty about the contents of one's mind. At one point Wittgenstein remarks that 'We say "I hope you'll come", but not "I believe I hope you'll come"' (1967, § 70). But what is wrong with the position of belief here? Whatever would be bizarre and less than welcoming in being told 'I believe I hope you'll come', the problem is not that you are here being offered mere belief instead of something expressing greater certainty. The sense of something wrong in this encounter would not be mitigated by the amendment, 'No, you misunderstand me. I'm quite *sure*, in fact *I have no doubt* that I hope you'll come'. The essential thing would still be lacking, that which is provided by the ordinary expression of oneself in 'I hope you'll come'. But if something would be missing, it seems fair to ask what more could reasonably be demanded of the person speaking here. After all, it may be thought, she is speaking her mind, to the very best of her knowledge of it, not holding anything back. So what is it that is wanted by the person unsatisfied by all this, the person who feels let down by this scrupulous self-description? That she should somehow speak *more* than sincerely; that is, speak somehow *beyond* her present knowledge of herself? In one sense this is precisely the demand, insofar as it means that what is wanted from the expression of hope is something that is not settled by evidence about herself or other theoretical considerations. And this will indeed involve speaking beyond one's fixed observation of oneself to the extent that even the most searching theoretical self-description will still leave

room in principle for disavowal, for a failure or refusal to identify with what one discovers.<sup>17</sup>

Richard Moran  
Harvard University  
Cambridge, Mass., U.S.A.

#### NOTES

<sup>1</sup> Wittgenstein (1958), pp. 39–40.

<sup>2</sup> Wright (1986), p. 402. This sentence follows an earlier passage which says, ‘The authority which our self-ascriptions of meaning, intention, and decision assume is not based on any kind of cognitive advantage, expertise or achievement. Rather it is, as it were, a *concession*, unofficially granted to anyone whom one takes seriously as a rational subject’. (p. 401) See also Wright (1989).

<sup>3</sup> By ‘pragmatic analysis’ I mean any account which explains the paradox in terms of the self-defeat of the purposes of assertion (by representing oneself as disbelieving what one is asserting). John Searle (1969) was perhaps the first to give an explicit development of this view.

<sup>4</sup> In addition to the sustained discussion of Moore’s Paradox in section X, Part II of Wittgenstein (1956), there are frequent remarks on the paradox in the four volumes of late writings on the philosophy of psychology. In particular, see (1980a) §§ 460–504, 700–719, 735–751, 810–823; (1980b) §§277–292, 415–420; (1982) §§ 81–88, 141–145, 416–428, 522–526; and (1992) pp. 8–12.

<sup>5</sup> Further implications of insisting on univocality across first- and third-person contexts in philosophy of mind are explored in Moran (1994).

<sup>6</sup> For example: ‘The language game of reporting can be given such a turn that a report is not meant to inform the hearer about its subject matter but about the person making the report’ (1956, p. 190).

This more explicit in § 502 of (1980a):

But now, we do nevertheless take the assertion ‘He believes p’ as a statement about his *state* [*Zustand*]; from this indeed there results his way of going on in given circumstances. Then is there no *first-person present* corresponding to such an ascription? But then, may I not ascribe a state to myself now in which such-and-such linguistic and other reactions are probable? It is like this, at any rate, when I say ‘I’m very irritable at present’. Similarly I might also say ‘I believe any bad news very readily at present’.

<sup>7</sup> Evans (1982), p. 225.

<sup>8</sup> On a related point from Wittgenstein, see (1980a) § 815: ‘Asked: “Are you going to do such and such?” I consider *grounds for* and *against*.’ That is, I may address a question concerning the expectation of a future event as transparent to a question concerning what I have most reason to do.

<sup>9</sup> The ancient contrast between the seductive, misleading Senses, and the trustworthy dictates of Reason can be seen, in part, as resting on a misrecognition of a related difference in kind between the two. The deliverances of the Senses can be compared to an unruly

mob, in conflict with each other, because they belong to the category of data on the basis of which one forms a judgement. Whereas, insofar as Reason represents the unifying judgement one forms from such evidence it is not a faculty superior to or in competition with the Senses. Rather, it just is the overall conviction one arrives at, and hence 'obeys'.

<sup>10</sup> This formulation is intended to leave open the question of the basis of our *general* entitlement to believe what we are told by others. On this see Coady (1992), and Burge (1993).

<sup>11</sup> Similarly, for several of the *Maximes* of La Rochefoucauld (e.g., 'Everyone complains of his memory, but none of his judgement' (# 89)), it would be mistaken to see them as purely psychological, and not also categorical, remarks.

<sup>12</sup> Sartre (1956), p. 114.

<sup>13</sup> Sartre (1956), p. 70.

<sup>14</sup> Consider Hume's prisoner and his guard. 'A prisoner, who has neither money nor interest, discovers the impossibility of his escape, as well when he considers the obstinacy of the gaoler, as the walls and bars, with which he is surrounded; and, in all attempts for his freedom, chooses rather to work upon the stone and iron of the one, than upon the inflexible nature of the other.' (*An Enquiry Concerning Human Understanding*, § VIII, 'Of Liberty and Necessity'.)

<sup>15</sup> At this point it might be asked whether this means that we must give up the original attribution of the belief in question. To do that, however, we would need a reason to give decisive weight to the avowal, as but one piece of evidence, over all the other evidence accumulated in and out of therapy.

<sup>16</sup> In Freud, see especially (1914).

<sup>17</sup> Earlier versions of this paper were read at Columbia University, Harvard University, The University of St. Andrews, The University of Michigan, Ann Arbor, Duke University, Cornell University, The University of Vermont, and McGill University. I am grateful to the audiences on those occasions. Special thanks to Dorit Bar-On, Garrett Deckel, Randall Havas, Christine Korsgaard, Douglas Long, Jim Pryor, Amelie Rorty, David Velleman, and George Wilson.

## REFERENCES

- Anscombe, G. E. M. (1976), *Intention*. Ithaca, New York: Cornell University Press.
- Burge, Tyler (1993), 'Content Preservation', in *Philosophical Review* 102, pp. 457–488.
- Coady, C. A. J. (1992), *Testimony*. Oxford: Oxford University Press.
- Evans, Gareth (1982), *The Varieties of Reference*. Oxford: Oxford University Press.
- Freud, Sigmund (1914), 'Remembering, Repeating, and Working-Through', in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, 12: 147, edited and translated by James Strachey. London: Hogarth Press.
- Moran, Richard, (1993), 'Impersonality, Character, and Moral Expressivism', in *The Journal of Philosophy* 90, pp. 578–595.
- Moran, Richard, (1994), 'Interpretation Theory and the First-Person', in *The Philosophical Quarterly* 44, pp. 154–173.
- Rey, Georges (1988), 'Towards a Computational Account of *Akrasia* and Self-Deception', in A. O. Rorty and B. McLaughlin (eds.), *Perspectives on Self-Deception*. Berkeley and Los Angeles: University of California Press.
- Sartre, Jean-Paul (1956), *Being and Nothingness*, translated by Hazel Barnes. New York: Philosophical Library.
- Searle, John (1969), *Speech Acts*. Cambridge: Cambridge University Press.

- Wittgenstein, Ludwig (1956), *Philosophical Investigations*, translated by G. E. M. Anscombe. Oxford: Blackwell.
- Wittgenstein, Ludwig (1958), *The Blue and Brown Books*. Oxford: Blackwell.
- Wittgenstein, Ludwig (1967), *Zettel*, translated by G. E. M. Anscombe. Oxford: Blackwell.
- Wittgenstein, Ludwig (1980a), *Remarks on the Philosophy of Psychology*, vol. I, translated by G. E. M. Anscombe. Oxford: Blackwell.
- Wittgenstein, Ludwig (1980b), *Remarks on the Philosophy of Psychology*, vol. II, translated by C. G. Luckhardt and Maximilian A. E. Aue. Oxford: Blackwell.
- Wittgenstein, Ludwig (1982), *Last Writings on the Philosophy of Psychology*, vol. I, translated by C. G. Luckhardt and Maximilian A. E. Aue. Oxford: Blackwell.
- Wittgenstein, Ludwig (1992), *Last Writings on the Philosophy of Psychology*, vol. II, translated by C. G. Luckhardt and Maximilian A. E. Aue. Oxford: Blackwell.
- Wright, Crispin (1986), 'On Making Up One's Mind: Wittgenstein on Intention', in Weingartner and Schurz (eds.), *Logic, Philosophy of Science and Epistemology*, Proceedings of the 11th International Wittgenstein Symposium. Kirchberg and Vienna: Holder, Pichler, Tensky.
- Wright, Crispin (1989), 'Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics', in Alexander George (ed.), *Reflections on Chomsky*. Oxford: Blackwell.