Replies to Heal, Reginster, Wilson, and Lear

Book Symposium on Authority and Estrangement: An Essay on Self-Knowledge, by Richard Moran

Replies to Heal, Reginster, Wilson, and Lear

Philosophy and Phenomenological Research

I'm very grateful for the attention given to my book by all the commentators, and their various and thoughtful responses have helped me in many ways. Several related issues are raised by the comments of Heal and Reginster, and to avoid repetition I will discuss them together here. Both of them raise questions about the scope and authority of rationality over a person's beliefs and other attitudes, and ask what is supposed to be wrong with adopting what I describe as a spectator's point of view on oneself, and whether this stance by itself involves the evasion of rational responsibility for one's attitudes. They also, in their different ways, provide searching discussions of the 'rakehell' case from Chapter 5, where several of these issues come together, so I'd like to respond to those parts of their comments together as well.

Rationality, etc.

First I should try to say more about how I see the ideas of rationality, responsibility, and agency entering the understanding of ordinary self-knowledge of states of mind such as belief or intention. For I do come back again and again to the relations I see between these families of ideas, and drawing connections between them is certainly an aim of the book. I realize that there is a kind of sweeping movement between these families of ideas throughout the book, and I think sometimes the sweep is understood to mean something sweeping in the exaggerated claims for the actual governance of rationality or responsibility in our lives. But my avowed strategy, in any case, is to try to secure common ground by being as minimalist as possible in commitments about how far our empirical lives are governed by, or explainable by, the normative structures of these concepts and other ones. I do think that the correct answer cannot be that our lives are not governed at all by these norms and concepts, and I think that anything like a global error theory of folk psychology would risk eliminating the distinct subject matter of philosophy of mind. So I feel that I'm trying, not always successfully I'm sure, to say that we need only agree that, among other abilities, human beings have the ability to act for reasons and ask for reasons, whatever philosophical account we may want to give of reasons at the end of the day. That is, the idea is that, if this much is agreed to, then the basic issue of self-other asymmetries is there before you in all its varieties, centrally including the relation of self-knowledge to agency and responsibility. So while agreement on this score is not exactly philosophically innocent (what is?), the idea of acting or believing for reasons should be compatible with a host of competing views about such questions as how far humans actually manage to be rational, how the natural world permitted a creature to develop who could demand and respond to reasons, how its cognitive and volitional abilities are made possible by the flesh and blood hardware of the person, or the metaphysics of reasons themselves. After all, even an Eliminativist or an Expressivist wants to be able to say the sorts of things we all say in describing someone as having his reasons and acting on them.

Both the modesty and the unavoidability of the assumption of rationality are reflected in such facts as that we could not distinguish the bodily movements that befall us from the actions we perform without appeal to the notion of something done for a reason. If some occurrence involving a human being thoroughly resists being understood in such terms, then we aren't seeing it as an action at all. This is not always a bad thing. History seems to show that for some human occurrences, abandoning that idiom of reasons can be an important advance in understanding. It is something always to-be-determined to what extent some human phenomenon can be understood in terms of reasons and responsibility at all, and some days are more depressing than others in the effort to find rationality in the goings on around us or within us. It is not an effort guaranteed to succeed, and it will never be the whole story of the shape of an action or an attitude (leaving aside just what that is supposed to mean). But I take it that the assumption of rationality is basic to identifying the phenomena of thought and action in the way that the assumption of meaningfulness is basic to understanding something as speech. Here again, we are familiar with the fact that not all speech does succeed in making sense, and that the capacity for meaningful speech rests on an empirical foundation that shapes its possibilities, and which can fail at the level of anything from neuro-chemical transmission to political trauma. And it is also true that the possibilities for meaningful speech are not given once and for all by either a theory of rationality or a semantics, but are worked out over cultural history by the activities of speakers and writers themselves. Still, as with reason and action, although speech can indeed be described and studied at the level of acoustics or of neuro-chemistry, that is, without attention to what makes for meaning in language, the assumption of meaningfulness is needed for the identification of the phenomenon in question as speech. Similarly, a philosopher needn't think that the psychological is exhausted by the rational to think that, at some level of description, it is a minimal commitment to any study that aims to describe or explain human thought or action, however tenuous our hold on reasons or meaning may be.

Most explicitly, rationality and agency enter the story I tell about first-person authority through an understanding of the idea of Transparency, which in turn I understand in relation to the idea of the 'immediacy' of first-person awareness. Both are familiar enough ideas. 'Transparency' stands for the claim that a person answers the question whether he <u>believes</u> that P in the same way he would address himself to the question whether P itself. From the first-person point of view, the one question is treated as 'transparent' to the other. And to say the knowledge that I believe that P is 'immediate' is to say that it is not based on any observations or inferences that I make. I argue that such immediacy is best explained as a consequence of what I call the 'Transparency Condition' on first-person statements of belief and other attitudes. What that means, however, is that immediacy and transparency, understood as describing the conditions for first-person authority, stand or fall together.

And the reason for that lies in the role of rationality in the argument. For insofar as my knowledge of my attitude is immediate, it is not based on the observation of myself or any other evidence. How could this be? One way it could be is expressed in the Transparency Condition, which tells us that I can answer a question concerning my belief about, e.g., what happened last week, not by considering the facts about myself as a believer but by considering what happened

last week. But how could that be legitimate? That is, how could it be even permissible, let alone some kind of normative requirement, for someone to answer a question concerning what some person's belief or other attitude is not by consideration of the fact about that person but by consideration of the facts about last week? Those are two quite distinct matters we might be asking about, the events of last week, and some person's current state of mind. But if the person were entitled to assume, or in some way even obligated to assume, that his considerations for or against believing P (the outward-directed question) actually determined in this case what his belief concerning P actually is (the inward-directed question), then he would be entitled to answer the question concerning his believing P or not by consideration of the reasons in favor of P. This sort of account would explain how the person's relation to his own belief is essentially different from anyone else's because his role in determining what he is to believe is different from anyone else's. And in this way, we would be relating transparency and the immediacy of self-knowledge both to each other and to the authority a person has both in making up his mind and in speaking his mind. But all of this, the immediacy, the transparency, and the authority, will only be in place to the extent that the person's reasons really do determine what his beliefs and other attitudes are. That's why I said that they stand or fall together. First-person authority and the immediacy of self-knowledge are aspects of what it is to be a subject of belief in the first place, but only insofar the person is entitled to the assumption that, e.g., what he believes about something on reflection is determined by what he has reason to believe. And I take it that this assumption, while indispensable to understanding someone as a believer at all, refers to a human capacity that is partial, fragile, and imperfect. And from my point of view this is all to the good, for I am interested in why the infirmities that first-person authority is subject to should have any special consequences for the rationality of the person or the attitudes in question.

In summing up a discussion of activity and passivity in belief, Jane Heal says the following:

> "I am passive with respect to many of my beliefs, not just in the weak sense that I allow myself to acquiesce in them, but in the stronger sense that I have no option but to carry on with them. To put this in epistemological terms, Moran's view has a strongly internalist flavor and he does not give much consideration to those facts about our cognitive position which an externalist would stress. Moran presents us as making up our minds under the guidance of reason. But the fact is that to a considerable extent we have our minds formed for us, by a variety of factors entirely beyond our control."

I think we would both agree in denying that we are active with respect to our beliefs either in the sense they are voluntarily adopted or in the sense that would imply that most of a person's beliefs or other attitudes are actually formed through deliberation by the agent. If voluntarism is the view that beliefs or other attitudes can be adopted at will, either for purely practical reasons or on a whim, then I take the rejection of voluntarism to be implied by, and not only consistent with, the role I give to deliberation and reason in belief-formation, for the reasons in question are going to be epistemic reasons, that is, reasons purporting to relate to the truth of the claim in question. And in the book I emphasize that the role I give to the deliberative stance is not meant

to suggest that most of our beliefs are actually formed through explicit deliberation or reasoning. We'd end up with many fewer beliefs for coping with the world than we actually have if we could only acquire them through explicit reasoning or deliberation. Here again, I'm trying to keep to a modest claim, and show that it has surprising consequences for understanding self-knowledge. The modest claim is that, while most of our beliefs and other attitudes either never arrive at consciousness at all, or only do so from we know not where, the fact remains that it is possible for a person to draw a conclusion, reach a finding, determine his belief about something on the basis of his assessment of the reasons supporting it. Put this way, I take this to mean something the denial of which would be equivalent to denying that people ever actually reason to a conclusion, or act or hold beliefs or other attitudes for reasons. And although it is not unheard of for a philosopher to make such a claim, I don't think this is a denial Heal means to make.

If my claim is understood this way, then I don't see it as in conflict with either epistemological externalism or the thought that our beliefs are formed by a variety of factors beyond our control. I understand epistemological externalism as a claim about the conditions that must obtain for a belief to be justified or for it to count as knowledge, and primarily what the externalist denies is that justification or knowledge require that the person's reasons be accessible, or that the conditions justifying the belief be known by the person to obtain. As such, externalism is a claim about the conditions for knowledge or justification, not an expression of skepticism about the possibility of being justified. It is a weaker condition than the internalist requirement that whatever justifies a belief must be known by the person to obtain. An externalist need not deny that among the factors justifying some belief of mine may well be the fact that I engaged in some explicit reasoning which resulted in my drawing the relevant conclusion, and that in this case this was both a genuinely reliable belief-producing mechanism and was known by me to be so. What I take it the externalist will insist on is that it is the reliability that matters, whether or not the conditions of reliability are known by the believer to obtain. Now, there may be those who deny that such activities as reasoning, considering reasons, weighing evidence, etc. ever play either a causal role in bringing someone to a belief or are among the conditions which actually contribute to the justification of a belief, but I don't see anything in the externalist position itself that should move one in that direction.

Perceptual belief is a favored case for eliciting externalist intuitions, and here certainly it cannot be denied that the process of belief-formation is dependent on a "variety of factors entirely beyond our control", and to a great extent beyond our knowledge or understanding as well. Here, we might well think, a person's beliefs are not "up to him" in any realistic sense at all. However, what I intend by "activity" or "rational control" here is compatible with the sense that perceptual presentations are not typically under the direct control of the person, and that they normally compel belief in an automatic and unreflective manner. Opening my eyes and looking about the kitchen, I have no choice in the matter as to acquiring various beliefs about the items lying around there. But I have no choice here because I take myself to have no reason to believe otherwise than what my senses suggest to me. My senses are in good working order, nothing seems awry, what they appear to present to me does not conflict with anything else I believe or am attending to at the moment; in short there is every reason to accede to the habit of belief and none to oppose it. And no effort of mere will could budge me from this belief. However, in a different situation, where I seem to see something unbelievable and I also have independent

reason for thinking that my visual impressions can't be trusted here, where it has been explained to me that I am the subject of an experiment and what I seem to see or hear will be unreal, then I may well not accede to forming a belief on the basis of these impressions. All this is familiar enough. Equally familiar is the fact that there are situations where the pull of belief persists in spite of my knowledge that there are clear reasons not to take appearances at face value here. I don't mean to deny this, but only insist that ordinarily what we believe, even when perceptually based, is sensitive to our grasp of how this belief fits in to the rest of our network of beliefs. The sense of "activity" I am appealing to is not the invocation of a superhuman power to either to adopt a belief or to suspend judgment when there is no epistemic reason to do so, but is rather the ordinary adjustment of belief to the total evidence. And indeed, there are some propositions so deeply implicated in the very idea of something being evidence for something else that it may be impossible to imagine conditions under which they could be subject to revision.

Failures of Transparency, Rationality, and Responsibility

I am grateful to Heal and Reginster for taking up the issues I raise in the final chapter of the book, in particular with their remarks on my diagnosis of what goes wrong in the reflections of the rakehell of the story and how that relates to the adoption of a spectator's point of view on his own attitudes and motivations. My discussion in that chapter is quite compressed and takes up issues that get less development than the questions of the previous four chapters, so I appreciate their taking that discussion further. They both suggest that it is my view that adopting a point of view on one's own attitudes that does not conform to the Transparency Condition, or that brackets the deliberator's stance in forming his attitudes, is necessarily something pathological or at the very least involves an abdication of responsibility for what one thinks or feels. For our purposes here, we can understand the third-person or spectator's point of view as a standpoint on one's own beliefs and attitudes that is not governed by the Transparency Condition, so that reflection on what my belief or other attitude about something may be is treated independently of the question of the truth of the belief or the world-related facts that would justify the attitude. Within the generous bounds of the Principle of Charity, this stance is an ordinary possibility in thinking about other people, but is problematic as applied to oneself, which is why, for instance, the various formulations of Moore's Paradox lose their paradoxical quality when shifted to the third-person. But my introduction of the Transparency Condition in Chapter Two is immediately accompanied by the insistence that there are conceivable situations where the two questions do come apart in the first-person case, the self-directed question and the world-directed question (p. 62). So when Reginster says that "according to Moran, first-person self-knowledge must obey the Transparency Condition", I want to raise a question or two concerning how this "must" is understood. For it is by way of rejecting the idea that Transparency is a logical feature of first-person discourse that I describe situations of first-person discourse that do not conform to it, situations in which the person either cannot come to know his attitude through avowal, or in which he is obliged to adopt a different stance toward his attitude. Hence I do not think that self-knowledge must obey the Transparency condition if that means that self-knowledge without immediacy or Transparency is not a genuine possibility. Rather, I see conformity to Transparency as an achievement of the person, the satisfaction of a normative

requirement, precisely because I want to understand the situations in which Transparency fails, and in this way understand why self-knowledge has the problematic importance in our lives that it does.

This matters to the interpretation of the rakehell story, since my diagnosis of what goes wrong with his reflections there does appeal to the tactical shift to a third-person perspective, and does relate this shift to an evasion of responsibility on the part of the man in the story. But I do want to distance myself from what Reginster describes as the most prominent version of the diagnosis that I offer, "according to which the sole shift to the third-person stance undermines responsibility". Like Reginster, I find such a diagnosis implausible, and for similar reasons to his. Sartre does sometimes speak as if it were solely the shift to a third-person stance itself that must per se involve an abandonment of responsibility for oneself, but this I chalk up to his inability to resist the pithy formulation of a paradox when it occurs to him. Bad faith, however, can exploit the first-person stance of commitment as easily as it exploits the third-person stance of prediction and explanation (p. 81). And even for Sartre himself, his preferred formulations of bad faith do not refer simply to the adoption of a third-person stance on oneself, but describe the person as tactically shifting between the two stances and interpreting one in terms of the other.

> "The basic concept which is thus engendered utilizes the double property of the human being who is at once a facticity and a transcendence. These two aspects of human reality are and ought to be capable of a valid coordination. But bad faith does not wish either to coordinate them nor to surmount them in a synthesis. [...] It must affirm facticity as being transcendence and transcendence as being facticity, in such a way that at the instant when a person apprehends the one, he can find himself abruptly faced with the other." (Sartre,'Patterns of Bad Faith')

Chapter Four is in part concerned with the distinction between two senses in which a person may take responsibility for some attitude of his, an 'internal' and an 'external' sense, only the former of which is connected to avowal. The latter kind of responsibility belongs to situations that may not only permit but require the adoption of third–person stance on one's attitudes (pp. 116 - 120), and involves the sort of control and intervention that brackets the question of truth, and thus an awareness of them that is not to be had through avowal. Hence I agree with Reginster here that there is no incompatibility between adopting a third-personal stance on some aspect of oneself and still assuming responsibility for it. Heal's case of the person who has been asked to perform some burdensome task, and reflects on his shame at a previous episode of failure by way of thinking about whether he can be trusted here, is a good example of this possibility.

The sorts of cases explored in Chapter Four are mainly ones in which the person fails to endorse some attitude of his or cannot make good rational sense of it, but nonetheless seeks to assume an 'external' responsibility for it, seeking to direct it one way or another. As with a phobia, for instance, he knows that the persistence or strength of the attitude is not responsive to his own sense of what is genuinely to be feared or not, and yet that needn't mean that he

abandons all responsibility for it. So I don't see that adopting a third-person perspective on some attitude of mine, which is resistant to the responsibility of ordinary rational control, is incompatible with other relations of agency and responsibility with respect to it. Concerning the rakehell case, Reginster describes my account as claiming that the rakehell's expressive interpretation of his shame ("using it as a ground for self-praise") is incompatible with continuing to endorse the original sense of shame. I would agree, however, that there is no incompatibility between fully endorsing some attitude and reflecting on it as a psychological state for its indicative or expressive value. For instance, someone could wholeheartedly endorse his own response of anger or compassion while reflecting with satisfaction on what kind of person this shows him to be. This is an unlovely moral reaction, but there is nothing impossible about it. Even here, however, there is the risk of something wrong with the person's sense of priorities and direction of attention. When a person takes pride in some world-directed attitude of his, there is the risk that the pride itself now assumes a role in sustaining the attitude, quite apart from the world-relating facts that elicited that attitude in the first place. The pride itself becomes a motive for holding fixed the attitude itself, so as to continue to serve as the basis for this pride, and thus the attitude becomes partly unhinged from, or distracted from, the world-relating reasons that made it sensible, let alone admirable, in the first place. That's a genuine risk, but I don't think that the expressive stance itself must involve any abdication of responsibility.

The question remains, assuming that moral reflection on oneself is not <u>always</u> misplaced, as to what has gone wrong with the Rakehell's reflections. I would agree with Heal that "The rakehell does not need to know what moral significance an observer would attach to his shame", but he may well need to know (or we may want him to be concerned to know) the actual moral significance of what he has done including his immediate response of shame. And, on the picture of moral reflection under consideration, the actual moral significance is what is revealed from an impersonal point of view. And it is here that things start to go wrong for him. I also agree that there is an issue in his misplaced use of his cognitive and emotional resources, but I don't think that explains his final criticism of himself: "not liking myself at all for feeling rather a good chap". This criticism is not equivalent to saying that he has misused his cognitive and emotional resources which could be better used elsewhere, and it is also more, I think, than a criticism of his priorities. It is something more like the claim that he has no <u>right</u> to this assessment of himself, even though it might have started off as a judgement that was true enough from an impersonal point of view. And he has no right to it because the Expressive interpretation of his shame relies on a perspective that holds fixed the fact of his shame itself, in a way that would make sense for another person but not for himself.

What is special about the perspective the rakehell adopts is that he engages in this bit of reflection with the aim of self-exculpation, or at least the replacement of his acute sense of shame with something more tolerable. This project requires the rakehell to <u>reopen</u> the question of how he is to feel about himself, and it is because of this that his reflections land him in paradoxes not shared by all expressive or indicative reflection on one's own attitudes. He re-opens the question of how he is to feel by seeking to avail himself of the expressive or evidential import of his original feeling of shame. Not every expressive interpretation of my emotional

state re-opens the question of how I am to feel. My prideful reflections on my anger or my compassion are not part of my answering the question for myself whether I should, after all, feel anger or compassion, or feel them to such a degree. But the rakehell's reflections are part of his re-opening the question for himself whether he should continue to feel (so) ashamed. In re-opening the question of his shame, his reflections place his original shame in suspense, for it is now placed in question by him. Being placed in question by him alters the sense in which this shame represents a standing evaluation of his. This is the aspect of the case that represents a self-other asymmetry, since another person might reflect on the rakehell's shame, as part of re-opening the question whether he should continue to feel (so) ashamed, without that reflection or that questioning making any difference to the settled quality of the rakehell's actual sense of shame. Only for the person himself is there this relation between the self-regarding attitude of shame and re-opening the question of whether this is really how he is to feel. A genuine observer could undertake this sort of reflection from a safe distance, assured that his own opening of the question did not alter the facts of the attitude he is reflecting upon.

In brief then, not all reflection on one's attitudes is part of a re-opening the question of what one is to feel. But the rakehell's reflections are part of such a deliberation. In the first-person case, re-opening the question alters the settled aspect of the attitude itself. But the settled aspect of the attitude of shame was necessary for it to have the expressive or evidential import that he is seeking to avail himself of. The interest for him of the evidential import of his shame is for what it may tell him about whether he should feel (so) ashamed. His reflection on this evidential import is undertaken to help him answer this question. But the settled attitude of shame only exists as something with this kind of evidential import insofar as it is not an open question for him how he is to feel here. When the reflection is in the service of answering a question about how he is to feel, then the original attitude is placed in question, is unsettled for him. And then if the redeeming interpretation of his shame depended on it being the interpretation of a settled attitude of shame, he cannot get there from here.

Deliberation and Self-Awareness

George Wilson raises some searching questions concerning the relation between reflective agency and self-knowledge of one's attitudes, and helpfully puts my account in the context of what he calls the "Practical Knowledge Model" of first-person knowledge of both our actions and our attitudes. Early in his comments, Wilson raises some doubts as to what kind of relation I could have in mind between a thesis about agency and the attitudes (i.e., that my beliefs are in some sense "up to me", and involve the assuming of a certain responsibility) and an account of how it is that we typically have unmediated awareness of the attitude we have. As he puts it, "In short, there is an initial question about how being a reflective agent in relation to one's attitudes can tell us anything about our privileged first-person awareness of those attitudes." This question arises because many beliefs and other attitudes are formed unconsciously, and hence are not within the conscious control of the person. Hence it is hard to see how one is an agent with respect to these attitudes at all, and hence the account offered would not apply to them. And from the other side, with respect to such things as the movements

of one's mouth that are involved in speaking, one is definitely an agent, and yet one does not have any kind of privileged awareness of one's mouth movements under that description. So in a case like this agency doesn't seem to bring with it any special link to first-person awareness. We might conclude from this that even if it were shown that one is an agent with respect to one's attitudes themselves (a controversial enough claim) this would not shed any light on the question of first-person awareness of attitudes, because various other happenings to which one is incontestably an agent do not involve any such awareness.

I see the issue of agency arising at a different stage of the argument than this description would suggest. I start from what I take to be a basic self-other asymmetry, a familiar starting point for discussions of both self-knowledge and the problem of other minds. While I can know the attitudes of other people only by watching what they do, listening to what they say, etc., I seem to be able to know my own attitudes without attending to or appealing to what I do or say. I seem to be able to know what I believe (or hope or intend, etc.) "immediately", that is without appealing to evidence at all. The claim of the basic asymmetry does not say that first-person access is the only kind of access that a person has to his own attitudes. It is certainly possible, and indeed not uncommon, for a person to come to attribute some attitude to himself from the same sort of evidential basis that he would appeal to in making the attribution to another person. It is important to the view I develop that both perspectives have a place in our thinking about ourselves. And the fact of their both being part of the person's ordinary relation to himself raises various questions. Since both perspectives make attributions of the same sorts of mental kinds, isn't it possible that they will deliver competing answers to the question of what I believe, or wish, or intend? And if they do conflict, how is such a conflict to be adjudicated? Do the deliverances of one perspective always have priority over the other, and if so, why? And even if we decide that there is no general priority such that the attributions from one perspective always stand to be corrected by the other one, there is another more general question about priority: if both perspectives constitute routes to the same kinds of attributions, would it be anything more than unusual for a person to have only one such mode of access to his attitudes? Does the rationality of beliefs and other attitudes require that the person have access to them from one such perspective, with the other perspective being dependent on it? Is there anything in the nature of being a subject of beliefs and other attitudes in the first place that would require the sort of access we associate with the first-person perspective?

Starting from this basic asymmetry, I take it for granted that a person can, under the right circumstances, answer the question of what his belief or other attitude is, in a way that does not appeal to the behavioral evidence. In seeking to explain how this could be, philosophers from Wittgenstein to Evans and beyond have claimed that the reason a person can answer this question about his belief without appealing to the behavioral evidence is that, when asked, e.g., whether I believe that my neighbors own a dog, I will not look within myself at all or at what I say or do, but rather direct my attention to my neighbors and their possible dog-ownership. And here we are, then, at the Transparency Condition, and the question arises how this could possibly be right. As Wilson says, the reasons that would justify an answer to the question 'Do my neighbors own a dog?' are quite different from the reasons that would justify an answer to a

question about what my, or anyone's, beliefs about the neighbors are. Any formulation of the Transparency Condition has to respond to the problem that it seems to assert that I can answer a question about one subject matter (my actual state of belief) by means of reflection on a wholly independent one. If we consider the issue in terms of reasons for belief then the issue is what could justify the assumption that I can answer the question of what my belief concerning P is by way of considering the reasons in favor of P itself. And my suggestion is that a person is entitled to this assumption insofar as his answering the question proceeds from the understanding that his sense of the reasons in favor of P itself does determine what his belief about P is. Otherwise, he would have to think that, even though he is considering the reasons in favor of P and coming to some conclusion about it, something other than that consideration is determining what his actual belief about P is. That can happen, of course. It can happen that processes having nothing to do with deliberation determine the beliefs I have that never become objects of assessment or critical reflection. And in certain situations, with respect to certain particularly fraught subject matters, it can happen that when I do deliberate about the question I either fail to arrive at a stable conclusion, or the conclusion I arrive at explicitly is one that, for some reason, I suspect may not be my genuine or abiding belief about the matter. In that sort of case, all the deliberating or critical reflection may be so much rationalization, a well-meaning story I tell myself that has little or nothing to do with what my actual belief is. This is a familiar enough situation of compromised rationality. But as familiar as it is, I don't think it should distract us from the pervasive extent to which we take it for granted that the conclusions we arrive at in the course of deliberation really do represent the beliefs we have about the matters in question, and I don't see how anything like ordinary argument or deliberation could have anything like the role in our lives that they do if this were not the case. The aim of deliberation is to fix one's belief or intention, and it could not do so if in general the conclusion of one's deliberation (e.g., about the neighbor's dog) left it as an open question, one needing to be answered in some other way, what one's actual belief about the matter is. And if there is no additional step to be made here, if I am entitled to assume that the conclusion of my reasoning tells me what my belief or intention is, then I think we have the form of the only kind of vindication that the Transparency Condition could have.

Agency thus enters into the account by way of the idea that in ordinary deliberation I exercise a particular authority over what my beliefs and intention actually are, since, unlike another person's reflection on my reasons and my beliefs, my own reflection goes beyond either a normative assessment or a rational recommendation of my belief and actually determines my belief. Again, this is not to say that all or even most beliefs or other attitudes are formed through anything like explicit deliberation. And even of the beliefs that are the upshot of explicit deliberation, my claim is not that this deliberation produces both the belief and the self-awareness of it. Rather, my attempt is to explain how it is that when the question arises what my belief about something is, I can answer that question in a way that conforms to the Transparency Condition. Wilson raises the point that if answering the question from the deliberative stance means that I make it true that I believe that P, this still doesn't help us understand the relation of this to self-knowledge of the belief in question. We can see this from the fact that I can equally make it true that I am producing the mouth movements that go with pronouncing the word

'dinosaur' without having knowledge of that fact, either under that description or under any description at all. As he says, "There still remains the question about how the agent comes to know that she believes that P just on the basis of considerations contained in her deliberations about P."

There are two different possible questions here. One question asks how it is that deliberation, and the considerations contained in it, produces not only the belief itself (if that much is allowed) but also self-awareness of the belief. On my account, however, deliberation does not of itself produce awareness of the belief or other attitude. Much ordinary deliberation, after all, does not concern itself at all with a question about self-knowledge, but rather simply seeks to determine whether, e.g., the neighbors next-door own a dog. In coming to an answer to that question, the issue of what my beliefs are need not arise for me explicitly at all, and hence the conclusion of that deliberation does not provide me with self-awareness of my beliefs. On the other hand, the quoted question may be read as asking how it is that, when I am seeking to answer the question of what my belief about something is, I can answer that question by way of avowal, that is, by taking a deliberative stance toward that something and the reasons in favor of it, and take my conclusion there to be the answer to what my belief actually is. If the conclusion reached could not be adopted or endorsed as my own belief about the matter, the deliberation wouldn't be doing me much good. The relation to self-awareness of the belief presumes the situation where it is a question about one's belief or other attitude that is being asked, and I take it that it is not believable that, in the ordinary cases, a person can only address such a question from the same third-person point of view he would adopt in thinking about someone else's belief. My claim is that the best explanation of this fact appeals to the capacity for avowal, hence conformity to the Transparency Condition, and that this is only possible to the extent that the person is entitled to the assumption that the rational agency he exercises in reasoning to a conclusion actually determines what his belief is. I think this is an assumption, or an entitlement, that is basic to reasoning itself, but I am also at pains to show how fragile or compromised it can be.

My reconstruction of the Transparency Condition raises the question of whether the actual first-person content of the answer is lost under such conditions, so that an utterance of the form "I believe that P" is only apparently about my own belief but really in fact reduces to a more or less hesitant claim about P itself. And that's why I spend time in Chapters Three and Four criticizing what I call the Presentational View, and defending the idea that conformity to the Transparency Condition does not annul the apparent first-person reference of judgments of the form "I believe that P". The thought is that if it is granted that a person can tell what his belief is without appeal to external evidence, and it is also granted that conformity to Transparency is consistent with the statement still being about one's belief, then Avowal must count as a way of answering the question what one's belief is, and hence a way of coming to know it. So my account does not attempt to explain how the deliberative forming of one's beliefs produces self-awareness of the beliefs formed, because I don't think that's true. What I do try to account for is how it can make sense for the first-person question about one's belief to be answered by way of reflection on the reasons relevant to the truth of the belief, rather than by reflection on the behavioral or other evidence that would be relevant to the psychological ascription to a particular

person. In brief my answer is that this is possible to the extent that the person's reflection on the reasons in favor of P does not simply result in the assessment of the belief-worthiness of P, or in a recommendation to himself to adopt that belief, but amounts to the person actually believing P. It is here that agency enters into the picture, by way of underscoring the difference between the thinking that issues in either the observation or assessment of an attitude and the thinking that is part of the formation of the attitudes themselves.

Avowal, Evasion, and Unfreedom

Authority and Estrangement makes intermittent reference to Freud and psychoanalysis, and I'm glad that Jonathan Lear has taken this aspect of the book seriously. The connections between the philosophical and the therapeutic issues concerning self-knowledge are more important to the book's aspirations than few explicit references may suggest. The book aims to give an account of why there should ever be the kinds of differences between knowledge of oneself and knowledge of others that we commonly recognize. The aim is not to assimilate all self-knowledge to immediacy and first-person authority, but to account for why central forms of it should exhibit these difference between relations to oneself and relations to others. This much Freud, for one, is committed to, and it is because he can take it for granted that he can concentrate his clinical and theoretical interest on the failures and limitations of such ordinary self-knowledge. But as to why any knowledge of one's psychological life should be able to proceed independently of external evidence, he has little to say. At the same time he shares another, more philosophical, assumption, with roots in Plato, Spinoza and others, that self-knowledge is central to the health and freedom of the person; that the restoration and expansion of self-knowledge (of certain kinds) is of deeper importance than the satisfaction of our boundless curiosity about ourselves. A special virtue of psychoanalytic theory and practice is that it focuses attention on this assumption itself, and makes it less easy to simply take it for granted. That is, it helps to show that we need an account of why knowledge of oneself (of a suitably specified kind) should matter at all to the health of the person, more than the obvious practical advantages of not being in error about one's own personal information.

There is a familiar, Hollywood-version, of the undoing of repression and its relation to consciousness and freedom that goes something like this: Young man (let's assume a man here) suffers traumatic experience in his childhood, which leaves behind an unconscious fear that marks the rest of his life in displaced, symbolic forms that are indecipherable to him and those around him. Certain actions or compulsions of his fail to make sense to him because they are covert repetitions of the original trauma, which is ruling his behavior from the unconscious. One day, however, something intervenes. Say, he gets hit on the head, or by narrative contrivance he finds himself in a situation which mimics the structure of the original trauma, and this reminds him of the whole earlier incident and the fears and compulsions that followed. And then, it hardly needs saying, bringing the original incident to consciousness, finding himself able to put those fears into words in the present tense, makes the fears and compulsions vanish. The neurosis evaporates in the light of consciousness and our hero is now free.

Alas, consciousness and reason are just as devious in their own ways as are the subterranean wiles of the unconscious, and will be employed by any reasonably skillful neurotic to preserve and protect the neurosis itself. The narrative form given above is itself the expression of a fantasy, both in its picture of what it is for something to be repressed and in the magical powers ascribed to self-consciousness. For one thing, there is no reason to think that simple awareness of a buried memory or fear is any more clear-sighted in it implications for practical reason than is the self-conscious, reflective awareness that, e.g., this person, this flattery, or this drink, represents a temptation I should resist. That is, without even venturing into psychoanalytic theory we must admit that we already know what reflective self-consciousness is and that it is often impotent and confused with respect to the most mundane matters. Why should it be any different when the object of self-consciousness is something repressed? Hence I welcome Lear's description of the process of analysis as a protracted practical deliberation that does not produce results by simply restoring some attitude to first-person awareness. Having reached (a kind of) first-person awareness is often only the beginning of a longer struggle with defense and rationalization. For me there remains the question, however, of why the coming to self-knowledge of a repressed attitude should be thought to matter <u>at all</u>, either to psychological health, or to rationality, or to freedom. That is, I want to understand why bringing something to consciousness or to the ability to enunciate should matter <u>at all</u> to the rationality of the attitude or the health of the person, especially once it is acknowledged that consciousness of any kind at all does <u>not</u> seem to matter to the good functioning of vast areas of our mental lives. Further, I want to understand not only what the point of lifting repression might be, but more importantly just <u>what</u> the lifting of repression is supposed to be once it is acknowledged that an attitude can remain fully under repression even while the person has full attributional awareness of it. And since the possibility of such 'attributional awareness' seems to show how a form of self-awareness can obtain independently of the ordinary sense of an attitude as "one's own", I want to understand what 'identifying' with an emotion or other attitude may be, in either Freud's or Frankfurt's sense, and why that should be thought to have anything to do with either freedom, consciousness, or psychological health. These questions were left open for me by the philosophical accounts of self-knowledge that I was familiar with, and I see them as questions that need answering for understanding any version of the "talking cure", any treatment that centers on the ability to bring something to speech.

The book outlines the distinction between one kind of self-awareness I call "attributional" and another kind that I characterize in terms of the ability to avow the attitude in question; that is, come to know the attitude in a manner that conforms to the Transparency Condition. And various turns of the argument emphasize how the one kind of awareness of one's attitude (the "attributional") is compatible with a lack of first-person access to it, and a lack of the sense of ownership and control we associate with self-consciousness. These are two ways of becoming aware of an attitude of one's own, and there are doubtless others, or more complex developments of these. It's true that much of the time I am emphasizing that, until we are talking about a form of self-consciousness that includes the capacity for the particular form of <u>avowal</u>, we are not talking about a form of self-consciousness that has the features that have made it of philosophical interest over the years. So I am saying that the capacity for avowal is

crucial for describing a form of self-consciousness that would make any sense, even critical sense, of its presumed relation to such things as freedom or rationality, and that shows the person's own relation to his attitudes to be different from his relation to other processes going on inside them (including psychological ones). A shift in concept takes place there, in the difference between attributional awareness to something avowable, that is a necessary condition for there to be any therapies of talking and self-consciousness, for there to be 'self-consciousness' in that sense. But that shift from attribution to avowal on the part of the analysand is not a sufficient condition for cure, truth, reason or freedom, as Lear's case story beautifully illustrates.

Part of what it means to say that Mr. A lives in a world of betrayal is that he can, and does, find betrayal everywhere, co-extensive with the world itself. And indeed, that world is the object of his reflections when he comes to know his attitude by avowing it, just as the Transparency Condition would require. That is, the 'world of betrayal' plays the role that the world is described as playing in avowal, and that's the whole problem. As Lear says, "For him, the movement from attribution to avowal of early betrayal to himself would be part of his defense." Since the world is infinitely rich, and life is infinitely ambiguous, reasons can always be found for discovering betrayal everywhere. For the right kind of neurotic, as with the virtuoso of confession, avowal itself can be so much shtick: a world-directed form of self-consciousness to be sure, but so is a compulsive, angry, meandering rant (which can also be a form of avowal). So I think it is entirely right that not only can 'stepping back' often be a sham, but even when it is perfectly real *qua* 'stepping back' it can be exploited and manipulated for neurotic ends just as can any other sophisticated, acquired skill. And for someone practiced in such rationalizations, bringing his sense of betrayal to avowability simply brings it into the space of manipulable reasons where the neurosis is most comfortable operating. Here it is not only that avowability is insufficient for cure, truth, reason or freedom, but that the very exercise of avowal itself is employed in keeping truth and the rest out of reach of the analysand. There are pathologies of avowal, and avowal can indeed play a crucial role in keeping us unfree, and Lear shows this to be a phenomenon worth exploring, both for its own sake and for the questions it poses for how we typically think about, e.g., self-deception or self-consciousness.

The story Lear tells also illustrates that what is said about the case of belief will only carry over to attitudes such as various desires, jealousies, or periods of grief with substantial modifications, and yet I would argue that these are world-relating attitudes as well, and hence subject to norms of intelligibility. For many forms of desire, for example, any application of a Transparency condition will have a different application than it does in the case of belief. For one reason, even if it is conceded that (for the broad class of 'judgement-sensitive' desires) the desire carries with it the assessment of its object as worth desiring in some way, that assessment would hardly carry the subject very far toward desire itself. I can acknowledge all sorts of things as being worthy objects of desire in some way (e.g., a stick of gum, a trip to Geneva, etc.) without having any desire for them myself. And further, any connection between desire and being found worth desiring has to be interpreted very broadly to be psychologically realistic at all. Desire is not always or perhaps even often for The Good or The Reasonable, as philosophers

have often understood these notions. I may be ashamed of my desire for something, because I take it not only to be a desire for something both trivial and disgusting, but more, to be a desire for that thing because it is both trivial and disgusting, something I desire under that very aspect; and yet I still acknowledge it as my own. But this is not so much to deny the connection between desire and some assessment of the characteristics of the object that make it desirable, as rather a reason to broaden and deepen our sense of what such an assessment can be, and how it can provide the coloring and descriptive content of a desire. For surely some activity's disgusting character can be just what appeals to me and makes the pursuit of it alluring, even rewarding, even if it conflicts with other values of mine. What would seem to me to leave the character of desire behind is if the object were pursued under no aspect at all, with nothing in its presentation to me (consciously or unconsciously) under which it is somehow appealing or beckoning or uncomfortably compelling.

There is indeed a kind of idealization in the appeal to Transparency, and in one way, of course, that makes it distant from a description of how we actually are. But it can also be seen as an idealization in the way that the avoidance of contradiction in one's beliefs is an idealization. That too, is not a description of us as we actually are, but at the same time it is not simply a false description of us or a description of how we would like to be. Instead, avoidance of contradiction is internal to the very concept of belief, given that it is internal to belief to purport to represent the world. And so, avoidance of contradiction is not simply an alien, ideal demand that we fail to live up to. Rather, it tells us something about what it is to have beliefs in the first place, what it is to be a believer. And that is why the exposure and acknowledgment of contradiction is something considerably more than the admission of imperfection, but rather carries with it the demand to do something, even if it is no more than to halt or to treat one's belief with less confidence than before. (It also doesn't follow from seeing contradiction this way that even as an ideal it is something we would always avoid.) I think that transparency can be seen as an idealization of self-knowledge (of a certain kind) in something like this way, that one's attitude toward something should be available to one via reflection on that very thing.