

## 24.500 Topics in Philosophy of Mind

### Other Minds

Spring 2003/handout 6

Moran, *Authority and Estrangement*

#### **Preface**

A defense and articulation of a version of “privileged access”:

a person may arrive at knowledge of his own attitudes in a way that is *not* based on evidence or observation of himself. In this sense, a person may know his own mind “immediately,” yet nonetheless declare his belief with an authority that is lacking in anyone else’s observation-based description of himself . (xxix)

(a) The contrast between ‘based on evidence’ and ‘immediately’ suggests that knowledge of other minds is always based on *distinct* evidence (we never “immediately” observe that someone is in such-and-such mental state).

(b) Clearly an “inner observation” model of self-knowledge is rejected.

Privileged access is no accident:

a person does not just happen to have this remarkable capacity... it belongs to the concept of a person that he should be able to achieve knowledge of his attitudes in this way. (xxx)

Privileged access connected to “wider asymmetries of self and other” (xxxii): respect, pity, deception, etc.

#### **Chapter 1: the image of self-knowledge**

Three points to be established:

A. The “proper characterization” of privileged access

B. How “philosophical accounts of self-knowledge often fail to account for (or sometimes even to describe) a specifically first-person phenomenon... prominent accounts of self knowledge often end up describing something that could just as well be a third-person phenomenon, or transposing an essentially third-person situation to some kind of mental interior” (2).

We have yet to see why it is a mistake to give an account of self-knowledge in terms of a “third-person model”.

C. “the lingering influence of a Cartesian picture of introspection [infallible inner perception] creates unwarranted skepticism about the very possibility of self-knowledge”.

I wish to defend a view of first-person awareness that sees it as both substantial, representing a genuine cognitive achievement, but which nonetheless breaks decisively with the Cartesian and empiricist legacy...this entails...rejection of...the purely theoretical or spectator’s stance towards the self (3)

## 1.1

A characterization of privileged access:

1. “a person can know of his belief or feeling without observing his behavior, or indeed without appealing to evidence of any kind at all”
2. “judgments made in this way seem to enjoy a particular epistemic privilege not accorded corresponding third-person judgements that *do* base themselves on evidence”

The “immediacy” of our introspective judgments simply means that “they are not inferred from anything epistemically more basic”. Immediacy and infallibility are independent.

‘epistemically more basic’ could be deleted

Ordinary perceptual judgments are also taken to be immediate, but immediate judgments need not be perceptual.

## 1.2

Background: the paradox of “Content and Self Knowledge”

Assume externalism: the property of believing/desiring/... such-and-such is extrinsic

1. We know our own thoughts immediately (i.e. not by inference)
2. Either this knowledge is “based on some form of inner observation” or on “nothing—at any rate, on nothing empirical”
3. Since “you cannot tell by mere inspection of an object that it has a given *relational* or *extrinsic* property”, inner observation is not the source of knowledge

Ordinarily, in order to know some contingent proposition you need either to make some observation. or to perform some inference based on some observation.

In this sense, we may say that ordinary empirical knowledge is always a *cognitive achievement* and its epistemology always *substantial*.

Examples of knowledge that is *not* a “cognitive achievement”:

a. “the judgment *I am here now*”

Is true that “Any token of this contingent judgment would be true and justified”? On the one hand, perhaps someone else could make the same judgment at a different time and place—as it might be put, *he was there then*. On the other hand, I might make the judgment *I am here now* while spinning in deep space.

b. the judgment that the world contains substances (on the Kantian assumption that this is “a precondition” for experience)

Any such judgment would be true, but why would it be justified?

c. “the judgment that I am jealous” (on the view that this is “essentially self-verifying”)

*I am thinking* would be a better example

Back to the argument:

4. If self-knowledge is based on nothing, it is not a cognitive achievement.

5. Since self-knowledge is not like a, b, or c, it is a cognitive achievement.

So:

6. Self-knowledge is not based on observation; neither is it based on nothing

So:

(1) is false, which is absurd.

M. notes (in effect) that (4) is false (18)

### 1.3

There is...a strong suspicion of a *conceptual* requirement lying at the bottom of first-person authority...Since the following chapters will be developing an account of self-knowledge that takes them to be essential to the nature of persons generally, I want first to investigate...whether admitting some conceptual basis to first person authority undermines the assumption of first-person reports as involving genuine cognitive achievements. (21-2)

Wright’s “extension determining” account. The biconditional:

S has the intention to  $\Phi$  iff S judges that he intends to  $\Phi$

is said (by Wright) to be a priori (modulo various qualifications), and the best explanation of its a priori status is that judgments of intention do not involve “the genuine detection of some independent state of affairs” (24)

Once again, but now in a different way, self-knowledge is said to fail of a “substantial epistemology”...The “insubstantial” conclusion depends...on the case for the a priori status of the biconditional... (24)

Why is the epistemological status relevant? Or is this because if the biconditional is a priori, it’s necessary?

In any case, M. questions the step from “judgment-dependence of the sort represented by the biconditionals” to the conclusion about no cognitive achievement, citing the case of color as a counterexample.

Perhaps more fundamentally, the “insubstantial” conclusion is obscure

For our purposes, however, the chief weakness of any analysis of this sort is how little it ends up illuminating any of the familiar *asymmetries* between first- and third-person psychological discourse. (25)

Is the point that Wright’s account doesn’t go far enough?

1. response-dependence of some form or another is a feature of a great variety of concepts

Disputable, but not M.’s main point

2. it doesn’t explain why first person ascriptions often don’t rely on evidence

Suppose the immediacy was explicitly built into the allegedly a priori biconditional—would that help?

3. “we also expect and sometimes insist that he take himself to be in a position to *speak for* his feelings and convictions, and not simply offer his best opinion about them”

This point hasn’t yet been explicitly cashed out

#### 1.4

(3)—the point that the the accounts so far have shown too much “concentration on the theoretical”—is illustrated by a discussion of Mellor on conscious belief.

...the case of second-order belief is too broad to capture either the particular character of conscious belief or the specifically first-person character of conscious belief (29)

A. One can have “a first-person second-order belief [about one’s beliefs]...[that] is still not a conscious belief” (30)

B. “We apply the term ‘conscious’ to the belief itself for reasons related to why we may apply this term to certain activities of the person...it is only with respect to one’s *own* activities that ‘consciousness’ has such an adverbial function”

What *is* a conscious belief? [I never got around to looking at Mellor’s paper.] Distinguish judging—an event—from believing—a state. Is there any such thing as a conscious belief (as opposed to conscious judgment)? Arguably not. (See Crane, *Elements of Mind*, ch. 4 sect. 32.)

## Chapter 2: making up your mind: self-interpretation and self-constitution

What is left out of the Spectator’s view is the fact that I not only have a special *access* to someone’s mental life, but that it is *mine*, expressive of my relation to the world, subject to my evaluation, correction, doubts, and tensions. This will mean that it is to be expected that a person’s own awareness of his mental life will make for differences in the constitution of that mental life, differences that do not obtain with respect to one’s awareness of other things or other people. (37)

### 2.1

A discussion of Taylor’s claim that “our understanding [of certain emotions] or the interpretations we accept are constitutive of the emotion...that is why [the emotion] cannot be considered a fully independent object”

Surely it’s true that, e.g., when I come to see that my feeling of guilt was false (i.e., presumably, I didn’t do wrong and so shouldn’t feel guilty), “the emotion [itself] is different”. That is, I either don’t feel guilty any more (perhaps I feel regretful), or my feelings of guilt change in various ways. But this doesn’t support Taylor’s claim.

M. glosses Taylor’s claim as: “one’s state of mind is in some way conceptually dependent on how one interprets it”, and notes that it is explicitly restricted “to a *first-person* phenomenon”

Why is this contrasted with Wright? Don’t the points on pp. 25-6 apply equally here?

M. distinguishes Taylor’s claim from another, that various conceptual capacities are necessary for being subject to certain emotions. Moral concepts are required in order to

feel shame; envy requires the notion of others possessing things that I don't possess, etc. (at any rate something along these lines). This is because these emotions "are themselves attitudes, modes of understanding the world as well as oneself". Put a slightly different way, some emotions involve certain judgements or beliefs (e.g. the belief that one did wrong).

At the end of this section, it's not clear whether M. thinks Taylor is onto something.

## 2.2

Pursuing "self-constitution" a little deeper. Two claims:

- A. Believing oneself to be in emotional state E is necessary for being in E.
- B. Having "a certain range of thoughts about [oneself] and [one's] situation is necessary for being in E.

M. points out that (A) is not particularly plausible, and that (B) is not particularly exciting.

But, "the language of 'self-constitution', in Taylor and elsewhere, suggests that the logical relation in question is one of sufficiency and not necessity".

Two claims:

- C. Believing oneself to be in emotional state E is sufficient for being in E.

And: "adopting a new interpretation of one's emotional state suffices to constitute the state as *somehow* different"

i.e. something along these lines:

- D. Judging that one is in E, and then changing one's mind and judging that one is in F suffices for a change in emotional state.

Are there any cases that fit C? M. tentatively suggests the example of believing oneself to be "ill at ease", and discusses an analogous case of taking one's marriage to be a failure. This latter example is supposed to illustrate how C-type cases create "intriguing possibilities for unhappiness" (45).

Are these examples plausible? Might one not mistake excitement for uneasiness? And couldn't someone mistakenly judge her marriage to be a failure because she thinks any successful marriage is unalloyed bliss?

In any case, this sort of “self-constitution” is relatively rare—“for the rest of mental life, the idea of ‘self-constitution’ will have to have some other meaning, if it has any application at all” (48)

### 2.3

Some remarks about D. Two main points. First, the judgments mentioned in D need not be true. Second, that there is a “logical reason” why “a difference in a person’s own interpretation of his attitude makes a difference to what his attitude actually is”, namely that “the condition [an outsider] seeks to describe is a condition of the whole person”

This isn’t quite right, surely. The last sentence of the section (“...the total state of the person we want to characterize includes the reconception itself”) puts it better. But why is this a “logical claim”?

### 2.4

Causal vs. logical versions of “self-constitution” distinguished, and “voluntaristic implications” disavowed. ‘Interpret’, ‘describe’, etc., as they occur in the statement of some self-constitution thesis, need to be understood as expressing “the same kind of commitment as belief” (53). The point at the end of 2.1 (emotions involve judgements) is used to show why one’s emotions might change in response to one’s judgements about them. It is a special case of first-order judgments/beliefs changing in response to second-order judgments/beliefs.

### 2.5

Theoretical and deliberative questions distinguished. The former “answered by a discovery of the fact of which one was ignorant”; the latter “answered by a decision or commitment of some sort” (58). E.g., “What will I wear?” is typically resolved by a decision, not by evidence. “I don’t know how to feel about that”—I don’t know what *to* feel (deliberative)/I don’t know what I *do* feel (theoretical).

This distinction...introduces a new dimension...We are now in a position to see how there is indeed a dynamic or self-transforming aspect to a person’s reflections on his own state, and this is a function of the fact that the person himself plays a role in formulating how he thinks and feels. (59)

The distinction doesn't just apply to emotions/mental states (as M.'s "what will I wear?" shows). And there are theoretical/deliberative questions in the third person case too.

## 2.6

### Transparency

"Do I believe that p?" "is answered by reference to (or consideration of) the same reasons that would justify an answer to the corresponding question about the world" (i.e. "Is it true that p?")

...what the "logical" claim of transparency requires is the deferral of the theoretical question "What do I believe?" to the deliberative question "What am I to believe?" And in the case of the attitude of belief, answering a deliberative question is a matter of determining what is true.

...we see that the vehicle of transparency in each case lies in the requirement that I address myself to the question of my state of mind in a *deliberative* spirit...

the term 'deliberative' is best seen at this point in contrast to 'theoretical', the primary point being to mark the difference between that enquiry which terminates in a true description of my state, and one which terminates in the formation or endorsement of an attitude (63)

But isn't the "theoretical question" one that we often answer without appeal to evidence? Contrast "Do you think there will be a third world war?" (addressed, as it might well be, to someone who has never considered the matter before) and "Do you believe that you live in Cambridge/that there is a God/that O.J. Simpson is guilty?" (addressed, as they might well be, to someone who has firm opinions on these topics). The first question is answered in a "deliberative spirit"—*should* I believe that there will be a third world war?—and is answered by consideration of the relevant geopolitical facts. The enquiry terminates when I *make up my mind* that, say, there will be a third world war, and announce that conclusion prefixed with 'I believe that'. (This *is* a "true description of my state"; cf. the last quotation, which doesn't express M.'s considered view.) But the second questions are answered by *reporting the psychological facts*, not by reconsidering the evidence for O.J.'s guilt, etc. Yet

transparency (or something very close to it) holds in these cases: if I remember that O.J. is guilty, then I don't need any further evidence in order to assert with justification that I believe that O.J. is guilty. So the worry is that transparency *doesn't* "require the deferral of the theoretical question "What do I believe?" to the deliberative question "What am I to believe?".

[Tom's observation: consider "I don't believe that p and I don't believe that not-p", which might be an instance of immediate and "privileged" self-knowledge. Here there is no corresponding claim "about the world"—*the world* is not undecided on the matter.]

[Small point: sometimes one asserts that one believes p, not because one has remembered p, or investigated the issue, but simply because one has remembered that one believes p. "I believe god exists" might be a case of this sort]

### **Chapter 3: self-knowledge as discovery and as resolution**

A picture will be "sketched out...of self-knowledge as involving the ability to avow one's state of mind and not merely to attribute it to oneself" (101).

It will be argued "that it is this feature...that accounts in part both

[A] for the way in which first-person reports are made without appeal to evidence, and [B] why the ability to make reports of just this sort should be bound up with the rationality of the person" (100-1).

#### The Transparency Condition

"A statement of one's belief about X [obeys] the transparency condition when [and only when] the statement is made by consideration of the facts about X itself, and not by an "inward glance" or by observation of one's behavior" (101)

#### Avowals

"An avowal is a statement of one's belief which obeys the Transparency Condition" (101)

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

nonevidential...What is meant by calling this a rational requirement will be made clearer later..." (68)

Doubts about this strategy. First, it's not just knowledge of one's beliefs that is "immediate" (knowledge of one's desires, hopes, intentions, etc. is too). So the stuff about avowal won't give a complete account of [A]. M. implicitly acknowledges this himself. He notes that aligning one's desires with the facts about what is desirable is "not something guaranteed by the logic of the first-person, but looks more like a kind of normative ideal" (62). That is, "I desire that p, but it's not desirable that p" is, alas, often assertable. And in any case, statements of one's belief that obey the transparency condition seem to be *nonimmediate*—aren't they inferred from "the facts about X itself"? It might be objected that no inference or rational transition is involved here, on the grounds that the rule "from 'p', infer 'I believe that p'" is not even *reliably* (let alone *necessarily*) truth-preserving. Hence (the objection continues) either belief statements that obey transparency are always *irrational*, being the result of applying an unreliable rule, or (better) that they are not the result of any inference at all. But this is not a good objection, because in circumstances in which one investigates whether p, and concludes that p, one *does* believe that p—and it is only these circumstances that the rule "from 'p', infer 'I believe that p'" applies. Cf. the (invalid) rule of necessitation in modal logic: from 'p', infer ' $\Box$ p'.

Second, analogues of the TC hold for perception (see, for discussion, Stoljar, "The Argument from Diaphanousness"), so one might wonder whether it could account for [B].

Does my statement that I believe that O.J. is guilty (not made by reconsidering the evidence—see end of previous handout) obey the Transparency Condition? If it *doesn't*, then many cases of self-knowledge concerning beliefs are left untreated.

### 3.1 Wittgenstein and Moore's paradox

Two forms:

(1) P, and I don't believe that P

(2) I believe that P, but it's not true that P

Since the paradox occurs in thought (not just in assertion), the "pragmatic" diagnosis is incorrect. (See Searle, *Speech Acts*, 65.)

### The Presentational View

'I believe that P' "does not in fact have any psychological reference, but is instead a mode of presenting" the proposition that P (71).

On the Presentational View, (2) is a contradiction.

What about (1)? 'I don't believe that P', on the Presentational View, is presumably a way of expressing my doubt about whether P (contrast 'I believe that not-P'), in which case the PV doesn't offer a complete diagnosis.

A full accounting of the Presentational View...would have to explain how verbs like 'believe' that serve to describe a person's state of mind in their third-person and past-tense uses lose that function...(72)

Perhaps more seriously, the PV has to account for contexts like 'If I believe that aliens are among us, I should be locked up' (not equivalent to 'If aliens are among us, I should be locked up'), 'Smith hopes that I believe he's an idiot' (not equivalent to 'Smith hopes that he's an idiot'), etc. See Searle, *Speech Acts*, on the "speech act fallacy"

...we can agree that the normal function of the first-person present tense of 'believe' is to declare one's view of how things are...and this follows from the fact that to believe some proposition *just is* to believe that it is true. (74)

Suppose that believing that p "just is" believing that it's true that p—in the formal mode, that 'x believes that p' and 'x believes it's true that p' are synonymous (or at any rate necessarily equivalent). It doesn't follow from this that "the normal function of the first-person present tense of 'believe' is to declare one's view of how things are". To see this, note that the parallel (and equally plausible) premise about hoping (hoping that p just is hoping that it's true that p) clearly doesn't imply that "the normal function of the first-person present tense of 'hope' is to declare one's view of how things are". Further, the premise is doubtful in any case. Don't the

conceptually challenged have beliefs without having beliefs about propositions and truth?

What is believed is believed as true, and hence the possibility of answering the question about one's belief in the "outward-looking" way described in terms of "transparency". (74)

Again, I don't see how this follows (what is hoped is hoped as true, etc.). So clearly M. means something different by 'what is believed is believed as true'. Perhaps: belief, unlike desire and hope, involves a commitment to the facts—but this doesn't seem very helpful. Alternatively: beliefs "aim at truth" (cf. footnote 5, with a reference to Velleman). I suppose this is right. (Velleman himself makes the point about hoping that p/hoping that it's true that p at pp. 247-8.) Still, the argument from the claim that beliefs "aim at truth" to the possibility of transparency isn't spelt out. Velleman's account, by the way, is that "to believe a proposition is to accept it with the aim of thereby accepting a truth" (p. 251), the import of which is not immediately obvious. (On the various things meant by 'acceptance', see van Fraassen, *The Scientific Image*; Cohen, *An Essay on Belief and Acceptance*; Horwich "The Nature and Norms of Theoretical Commitment", *Philosophy of Science* 1991; "Practical Reasoning and Acceptance in a Context", reprinted in *Faces of Intention*.)

For conceptual reasons, the degree of doubt or 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. What is unavoidable from the first person perspective, then, is the connection between the question about some psychological matter of fact and a commitment to something that goes beyond the psychological facts (77)

### **3.2 Sartre, self-consciousness, and the limits of the empirical**

Sartrean jargon roughly explicated. Being-in-itself: contribution non-mental world makes to phenomena (trees, chairs, etc.). Being-for-itself: contribution consciousness makes to phenomena.

Being-in-itself: has no cause, is "positive" (no negative facts, etc.), has no explanation.

Being-for-itself: has a cause, is negative ("Man is the being through whom nothingness comes to the world" (59)), has no explanation.

“the human being..is at once a *facticity* and a *transcendence*” (i.e. is a contingent individual in such-and-such circumstances, who is free)

M. uses Sartre’s gambler to illustrate the “contrasting roles of commitment (of oneself) and theoretical knowledge about oneself”.

### 3.3 Avowal and attribution

The limits of 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. (84)

A discussion of some “familiar therapeutic contexts”.

This seems to be an example where the Moore-sentence ‘I believe I have been betrayed, but I have not been betrayed’ is assertible. (For an example concerning the other type, see Crimmins, “I falsely believe that p”.)

Rey: “avowed” beliefs are different in kind from “central” beliefs (beliefs that one can only report on the basis of third-person evidence).

[On this account] it is completely unclear how we may see the two as *clashing* at all...But avowing and reporting cannot be thus isolated from each other, if for no other reason than that any avowal is itself behavior, and thus at the very least a central part of the evidence for the explanatory attitude of belief. Hence, in principle the “two attitudes” could never be of utterly distinct types. (87)

Preserving the sense of conflict within the self...requires that we see the meaning of a psychological term like ‘belief’ as univocal across the two contexts. (87)

Is Rey arguing for an ambiguity? (There are two importantly different kinds of humans, male and female, but ‘human’ is not ambiguous.) In any case, why can’t there be the appropriate sort of clash between belief<sub>1</sub> and belief<sub>2</sub>?

Why should...the ordinary nonevidential way of avowal, be privileged in any way in our daily lives? (89)

we should consider what an idealized but still purely theoretical relation of expertise toward oneself would be missing. This will put us in a position to understand how it is not just permissible but essential that ordinary first-person knowledge proceed independently of evidence, and hence why a nonempirical or transcendental relation to the self is ineliminable. (90)

Two distinct aspects of the authority of the first-person:

A: immediacy

B: “the Transparency Condition...is a feature of ordinary first-person discourse”

Any treatment that took upon itself the task of restoring self-knowledge would have to restore a capacity that conformed to...the condition of transparency. The person might be told of her feeling of betrayal, and she may not doubt this. But without her capacity to endorse or withhold endorsement from that attitude [i.e. the feeling of betrayal], and without the exercise of that capacity making a difference to what she feels, this information may as well be about some other person...It is because her awareness of her sense of betrayal is detached from her sense of the reasons, if any, supporting it that she cannot become aware of it [the feeling of betrayal] by reflecting on that very person...The rationality of her response requires that she be in a position to avow her attitude towards him... (93)

An avowal is a statement of belief that obeys the TC. What is the avowal in question? It can't be “I believe that I feel betrayal” (or “I believe that I feel that I have been betrayed”, because that does obey the TC. Rather, it is “I believe I have been betrayed” (see 85). So the problem is that she can't become aware of her *belief* that she has been betrayed (perhaps not quite the same as a “feeling of betrayal”). (Cf. 85 on anger: the TC concerns one's belief that one is angry, not one's anger.)

### **3.4 Binding and unbinding**

The theoretical and deliberative stances illustrated by the “recognisable quandary of binding and unbinding oneself”: the example of the coin toss.

## **Chapter 4: the authority of self-consciousness**

### **4.1 Expressing, reporting, and avowing**

The claims of ch. 3 “require further development”

“in the broad sense of ‘express’ [means] simply to manifest or reveal”, “expressing and reporting are not...mutually exclusive categories”. Thus a denier of the PV need not “deny that [first-person] utterances are expressions” (103).

Conforming to transparency is compatible with the statement being a report (105).

What conforming to transparency comes to is the commitment that beliefs I call my own are beliefs I can endorse as true. (105)

#### **4.2 Rationality, awareness, and control: a look inside**

“The remainder of this chapter concerns why avowability should have any special place in our thinking about first-person knowledge”

M. wishes to defend something along these lines:

When there is an attitude of mind that I cannot become aware of through reflection on its *object*, it suggests that the attitude is impervious to ordinary rational considerations relevant to the maintaining or revising of the attitude. .. if she cannot become aware of [her sense of betrayal] by reflection on the very person she feels betrayed by, that suggests that the persistence of the sense of betrayal has nothing to do with her sense of the facts about *him*...A belief that cannot be avowed is thus cognitively isolated, unavailable to the normal processes of review and revision...Thus, we could explain why it is that the capacity not just for awareness of one’s beliefs, but specifically awareness through avowal, is both the normal condition and part of the rational well-being of the person. (108)

She believes that she has been betrayed, but she is not prepared to assert “I have been betrayed”. Still, couldn’t her belief that she has been betrayed be responsive to the facts about him? Indeed, couldn’t we imagine a case where it is *only* the *repressed* beliefs that are the rational/fact-responsive ones?

An objection to McGinn’s argument that “rationality of attitudes [requires] self-conscious awareness of them” (109-10). “[M]y beliefs, like the flow of perceptions, interact and undergo revisions all the time without any intervention on my part” (111)

I will be arguing that, whereas awareness of one’s beliefs is too strong a requirement on “rational adjustment” when this is interpreted in the first way (roughly, the sense consistent with “functional regulation”), it is too weak a requirement when [it] is understood in this second way, as a description of the person’s deliberative relation to his own beliefs. (112)

#### **4.3 From supervision to authority: agency and the attitudes**

“to be explored...the connection between the ideas of knowledge of oneself (including the authority attaching to one’s declarations) and responsibility for oneself (including one’s thoughts)” (114)

“Inflicting” attitudes on oneself (by training, drugs, etc.) vs. the production of one’s attitudes by practical/theoretical reasoning (rational control of one’s attitudes).

What is the connection with self-knowledge? One may know one’s inflicted attitudes with immediacy and authority.

#### 4.4 The retreat to evidence

4.3 “suggests a further dimension to the relevant idea of “authority”.

“If we try to imagine the abrogation of first-person authority as quite generalized, covering all questions about a person’s current beliefs, desires, and other attitudes, there is difficulty in imagining the simplest forms of explicit reasoning and concluding” (122).

What is the example on p. 125? The person “announces confidently” that he believes that  $p$ , but “can’t take the answer he delivers to necessarily report the actual belief he has arrived at”. He’s confident but wrong?

Surely someone could make his mind up, etc., without having any beliefs about his own mental states (a fortiori no authority with respect to them). So that can’t be sort of case that is supposed to be ruled out.

Doesn’t M.’s own response to McGinn (see 4.2) cast some doubt on the quoted claim from p. 122? A person’s first-order beliefs etc. can look after themselves without being monitored and guided by his second-order beliefs. So why can’t the person engage in (first-order) reasoning, despite being wrong/unreliable etc. about what he believes? A pathological case, no doubt, but it’s not clear why the lack of FPA impugns the person’s *first-order* rationality.

#### 4.5 First-person immediacy and authority

“we have seen that we do have reason to relate the capacity for self-consciousness to the rationality of the person...we need better characterizations of both the notion of “self-consciousness” and of “rational control”.”

In particular, what’s the relevance of “immediacy” to rational control of one’s attitudes?

*Anscombe’s Condition*: “the relevant question “why [are you  $\Phi$ ing]” is refused application if the agent’s awareness of what he is doing (under the...description [ $\Phi$ ing’]) is not “immediate” in our sense”

The basic argument for AC: “the agent takes the question of what he is doing to be answered by his decision as to what is worth pursuing, and that question is not a predictive or explanatory one to be answered by observation of himself”

But presumably the agent has *already* decided (to pump water, e.g.)—so no new decision needs to be made.

Sometimes a person does find out that he is, e.g., intentionally walking to the dept., by observing his eastward progression along Mass ave. (imagine he can’t remember where he’s going).

The “immediacy” of self-consciousness and first-person authority, the fact that I can be aware of my belief without inference or evidence, is a function of the fact that information about myself that I would gain through inference or evidence about myself is ruled out as irrelevant to the question of what I am to believe. (134)

Isn’t the explanation just of immediacy, not authority?

#### **4.6 Introspection and the deliberative point of view**

Features of self-knowledge:

1. Immediacy
2. Authority
3. “the capacity for first-person awareness has a special relevance to the psychic health of the person”

We need an account of (3).

#### **4.7 Reflection and the demands of authority: apprehension, arrest, and conviction**

(3) answered: reflective consciousness related to rational freedom.

Two senses of “suspending” a belief or desire.

1. Suspending its legitimacy—the “appraising sense”
2. Suspending its “actual psychological force”

Only the first person case does (1) involve (2). Answering “deliberative questions”—“a perfectly homely assertion of one’s freedom”.

#### 4.8 The reflective agent

More on the shortcomings of self-knowledge as mere awareness. “The problem with generalizing the theoretical stance towards mental phenomena is that a person cannot treat his mental goings-on as just so much data or evidence about his state of mind all the way down, and still be credited with a mental life (including beliefs, judgments, etc.) to treat as data in the first place.”

“When our analysand learns attributively that she feels betrayed [that she believes she has been betrayed?]. . . It remains open to her whether to accept those feelings at face value. . . the issue [whether she has been betrayed] remains hers to decide”.

But she *already* believes that she has been betrayed—and might not this belief “serve as a basis for further thought” (even though it is not avowed)?

### Chapter 5: impersonality, expression, and the undoing of self-knowledge

“we need a better understanding of why [the “theoretical perspective”] should contribute, in a certain range of cases, to the undoing of the very state subjected to this objectifying gaze”

#### 5.1 Self-other asymmetries and their skeptical interpretation

We need to “separate the basic fact of the [self-other] asymmetries from a skeptical interpretation of them”

#### 5.2 The partiality of the impersonal stance

Nagel’s principle: “To regard oneself in every respect as merely a person among others, one must be able to regard oneself in every respect impersonally” (160)

One interpretation: semantic principle of univocality

Another interpretation: “any thought I have about myself must be acknowledged to be a type of thought it would *make sense* to have about another person. . . *as well as* that any thought I have about myself must be one that I can translate, without alteration of meaning, into a thought expressed third-personally, as a description of the person who happens to be myself.”

“this last normative demand” is not equivalent “either to the semantic principle or to the acknowledgment [the first part of the second interpretation], and M. thinks it’s wrong, because of the deliberative/theoretical contrast.

On the one hand there’s Perry’s “essential indexical” claim. The thought that I am making a mess seems to be (somehow) different from the thought that AB is making a mess (etc.). On the other hand there’s M.’s distinction between the Theoretical and Deliberative “perspectives”, which of course does not line up with the distinction between indexical/nonindexical thoughts. What is this “thought I have about myself” that’s not translatable..., not because of Perry’s distinction, but because of M.’s distinction? (See p. 161, bottom.) It’s not clear (to me) why M.’s distinction requires rejecting the “last normative demand”.

### **5.3 Self-effacement and third-person privilege**

### **5.4 Paradoxes of self-censure**

### **5.5 Incorporation and the expressive reading**

### **5.6 Not first-personal enough?**