

Richard Moran

AUTHORITY
AND
ESTRANGEMENT

AN ESSAY ON SELF-KNOWLEDGE

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

COPYRIGHT © 2001 BY PRINCETON UNIVERSITY PRESS
PUBLISHED BY PRINCETON UNIVERSITY PRESS, 41 WILLIAM STREET,
PRINCETON, NEW JERSEY 08540
IN THE UNITED KINGDOM: PRINCETON UNIVERSITY PRESS, 3 MARKET PLACE,
WOODSTOCK, OXFORDSHIRE OX20 1SY

ALL RIGHTS RESERVED

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA

MORAN, RICHARD
AUTHORITY AND ESTRANGEMENT : AN ESSAY ON SELF-KNOWLEDGE /
RICHARD MORAN
P. CM

INCLUDES BIBLIOGRAPHICAL REFERENCES AND INDEX.
ISBN 0-691-08944-2 (CLOTH : ALK. PAPER)—(ISBN 0-691-08945-0 (PBK. : ALK. PAPER)

1. SELF-KNOWLEDGE, THEORY OF. I. TITLE

BD438.5.M67 2001

126—DC21 2001021157

THIS BOOK HAS BEEN COMPOSED IN ITC GARAMOND

PRINTED ON ACID-FREE PAPER. ∞

WWW.PUP.PRINCETON.EDU

PRINTED IN THE UNITED STATES OF AMERICA

1 3 5 7 9 10 8 6 4 2

1 3 5 7 9 10 8 6 4 2

(PBK.)

To my parents

CHAPTER TWO Making Up Your Mind: Self-Interpretation and Self-Constitution

2.1 SELF-INTERPRETATION, OBJECTIVITY, AND INDEPENDENCE

Ordinary "realism" about the mental suggests a relation of logical independence between the description of some feature of mental life (e.g., a thought or emotional response) and the feature or state itself. But in the case of the *self*-interpretation of various aspects of psychological and social life, some philosophers have argued that this independence does not obtain. The hermeneutic tradition and Taylor's Constitutive Claim.

We want to understand why this idea of a constitutive relation between interpretation and object is restricted to "intentionally characterized" phenomena, and why within these it is restricted to their *first-person* interpretation.

Distinguishing the Constitutive Claim from the idea that certain conceptual capacities are necessary for the possibility of certain emotional responses.

2.2 SELF-FULFILLMENT AND ITS DISCONTENTS

To say that a person's self-interpretation "constitutes its object," even partially, suggests that, in those cases, the new interpretation *suffices* for a new description to be true of it (perhaps a description conforming to the new interpretation itself). Sometimes taking oneself to be a certain way (e.g., uncomfortable, ambivalent) is sufficient for being truly characterized in those terms. The negative character of such "compromising" self-descriptions, and the self-fulfilling logic of contamination. The appeal of the Constitutive Claim may draw strength from such cases, but they cannot be said to characterize psychological life generally.

2.3 THE WHOLE PERSON'S DISCRETE STATES

The privilege accorded to a person's own interpretation of his state need not be restricted to those cases where the constitutive rela-

tion involves the state's *conforming* to the person's interpretation of it. Even an interpretation which we, on the outside, can see as importantly *mistaken* may nonetheless have a claim to defining his state overall, in a way that is not shared by an outsider's interpretation of it. Seeing one's own pride as sinful constitutes it as importantly different from what it would otherwise be, since conditions like pride are orientations of the person and not atomistic particulars.

2.4 BELIEF AND THE ACTIVITY OF INTERPRETING

Verbs such as "interpreting" and "describing" can denote ordinary activities that can be performed at will, but in that sense the activity of "merely describing" one's state a certain way will not be expected to constitute it as different, any more than mere describing makes a constitutive difference to other things in the world. Rather, the examples that make the best sense of the Constitutive Claim concern the self-interpretation of one's emotional state where "interpreting" it a certain way means actually taking it to *be* that way. This involves cognitive commitment which, like belief, is not something that can be undertaken at will. And this begins to clarify why reflection on one's own intentional states should be linked to the transformation of their character.

2.5 THE PROCESS OF SELF-CREATION:

THEORETICAL AND DELIBERATIVE QUESTIONS

Sometimes reflection on one's state of mind is a purely theoretical matter, where the question is how some feature of one's mental life is to be correctly identified. But often one's reflection is more "deliberative" in spirit and seeks to bring one's state of mind to some kind of resolution. This sort of uncertainty is answered by something more like a decision than a discovery. The difference between "I don't know what to feel here" and "I don't know what it is that I do feel here."

The situation of deliberative, rather than theoretical, reflection on one's state helps to account for why it should be *self*-interpretation

alone that is said to "help shape the emotion itself," and why, even on a commonsense "realism" about mental life, we should *expect* self-interpretation to play this special role.

2.6 RELATIONS OF TRANSPARENCY

It is sometimes claimed that, from the first-person point of view, the question "Do I believe that P?" is *transparent* to a corresponding question "Is P true?," a question which involves no essential reference to oneself at all. "Transparency" here means not reduction, but that the former question is answered in the same way as the latter. But what is the basis for such transparency, when it obtains?

There do seem to be situations where the person can or must answer the psychological question of what his attitude *is* in a way that is *not* "transparent" in the relevant sense. Rather than being guaranteed by logic, the claim of transparency is grounded in the deferral of theoretical reflection on one's state to deliberative reflection about it. Conforming to transparency as a normative demand. Both the "transforming" character of self-interpretation and the "transparency" of one's present thinking are grounded in the interaction between theoretical and deliberative reflection on one's state of mind and the primacy of the deliberative stance within the first-person.

CHAPTER THREE Self-Knowledge as Discovery and as Resolution

3.1 WITTGENSTEIN AND MOORE'S PARADOX

Why ordinary self-knowledge should be "nonevidential" rather than a matter of theoretical attribution to oneself, and why this difference matters. Moore's Paradox as a way of describing situations where one's attribution of an attitude to oneself does not match the attitude one would explicitly express or endorse (hence, where the Transparency Condition is violated).

What is paradoxical in Moore's Paradox is not restricted to situations of speech or the pragmatics of assertion.

The Presentational View, whereby the first-person present-tense of 'believe' does not have any psychological reference, but *only* serves to "present" the embedded statement (e.g., saying "I *believe* it's going to rain" as expressing uncertainty about *the rain*). Rejection of this view, and its attribution to Wittgenstein.

First-person authority and first-person subjection; blindspots.

How is transparency consistent with the fact of the different subject matters of the two questions (a state of mind, a state of the weather)? Belief as empirical psychological fact about a person, and belief as commitment to a state of affairs beyond the self. Inferring from someone's having some belief to the truth of that belief, versus the categorical relation between belief and "taking to be true" in the first-person.

3.2 SARTRE, SELF-CONSCIOUSNESS, AND THE LIMITS OF THE EMPIRICAL

The "self-as-facticity" (seeing belief as a psychological fact about oneself) and the "self-as-transcendence" (seeing belief as a commitment of oneself). The case of the gambler as illustrating conflict between these two perspectives, neither of whose claims is avoidable. One type of "bad faith" as the exploitation of the purely theoretical perspective on oneself; another type as the empty transcendental assertion of one's freedom and clean slate.

For the first-person, the sense of the primacy of the practical, deliberative question over the predictive, theoretical question.

3.3 AVOWAL AND ATTRIBUTION

"Transparency" as a normative requirement on rational agency. Reporting one's state of mind, where this is an evidence-based attribution, in the service of psychological explanation, versus expressing or avowing one's attitude, where this is a matter of one's declared commitment. The description of forms of psychic conflict

CHAPTER TWO

Making Up Your Mind: Self-Interpretation and Self-Constitution

It is often said that the act of reflection alters the fact of
consciousness on which it is directed.

—Sartre, *Being and Nothingness*, p. 121

We have seen how natural it is to think that if self-knowledge is something substantial at all, it must be something that corresponds at least roughly to the Perceptual Model of it. We speak of the "realm" of the mental, and one of the more innocent things suggested by this way of speaking is that if we are to speak of awareness here, it must be like any other awareness of any other realm of empirical phenomena. At this point in the argument it hardly matters whether we think of this realm as containing states, processes, objects, or whatever. We may be willing to countenance certain apparent peculiarities of the first-person

access (e.g., privacy, authority, immediacy); but if the appearance of knowledge or awareness is not to be a sham, it seems we must somehow accommodate such peculiarities within a more familiar picture of observational knowledge. Hence, the recurrent attractiveness of the picture of something like a sealed room, to which only I have access (but which I am also unable to leave).

Some version of this picture may seem to be required by even the most modest sort of realism about mental life. In the case of any given person, the thoughts, the mental phenomena, are there; and while most people have only a kind of external access to them, one person has an immediate internal access to this realm. When this picture is found extravagant, it is usually by way of rejecting either the presumed substantiality of self-knowledge or the supposedly special features of the first-person position. By contrast, rather than criticizing it as a metaphysical extravagance, I hope to show how this picture radically *underdescribes* the differences between self-knowledge and the knowledge of others. 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. For this reason, introspection is not to be thought of as a kind of light cast on a realm of inner objects, leaving them unaltered. We looked briefly at one aspect of this difference with respect to the notion of 'conscious belief', but there are several aspects of it. One or another such aspect has been taken by various philosophers to be incompatible with the ordinary assumption of either an everyday realism about mental life, or the idea that the phenomena of self-knowledge do indeed involve the awareness or detection of some set of facts. The assumption is that either realism or a substantial epistemology has got to go. What I hope to show is that, if we begin with a kind of commonsense (or perhaps simple-minded) realism about mental life, then the various

ways in which self-awareness and self-understanding make a difference to the character of one's mental life are just what we should expect. Seeing more deeply into *why* we should expect this will take us further into the question of why there should be any apparent peculiarities of first-person access in the first place; and why the language of "special access" saddles us with a misleading, too purely epistemic, idea of the responsibilities of the first-person position.

2.1 SELF-INTERPRETATION, OBJECTIVITY, AND INDEPENDENCE

The difference self-consciousness makes is bound up with a related first-person difference, which has received less attention within philosophy of mind than in certain forms of social theory (particularly with a Hegelian provenance). In any event, there is a general idea, found in both Continental as well as Anglo-American philosophy, that the way in which a person conceives, for example, his own emotional state is partially constitutive of what that state is. Similar to what I have claimed about the altered character of a belief that becomes a conscious one, it is claimed that, for example, for someone to interpret his own response as, say, either righteous indignation or as mere peevishness *constitutes* his state as being of a different kind. The first-person interpretation of an emotional state is supposed to play a role in constituting the identity of the state that is not shared by interpretations "from the outside." This claim is often part of a larger argument against the application of ordinary notions of objectivity or realism to certain aspects of psychological and social life. For instance, Charles Taylor has argued in a series of papers that the ordinary notion of representation cannot apply to various important situations of self-understanding.

[B]uilt into the notion of representation in this view is the idea that representations are of independent objects. I frame a representation of something which is there independently of my depicting

it, and which stands as a standard for this depiction. But when we look at a certain range of formulations which are crucial to human consciousness, the articulation of our human feelings, we can see that this does not hold. Formulating how we feel, or coming to adopt a new formulation, can frequently change how we feel. When I come to see that my feeling of guilt was false, or my feeling of love self-deluded, the emotions themselves are different. . . . We could say that for these emotions, our understanding of them or the interpretations we accept are constitutive of the emotion. The understanding helps shape the emotion. And that is why the latter cannot be considered a fully independent object, and the traditional theory of consciousness as representation does not apply here.

(Taylor 1981, pp. 100–101)

We can see how a similar question might arise with respect to the claim that 'conscious' as applied to belief indicates a difference in the character of the state itself. It was said that the relation here is not like that of an object to the observation of it. And one familiar way to spell out *that* relation would be to say that, for instance, a tree is a real independent thing whose existence and qualities are not dependent on being observed. It has full, real, objective existence. So, then, what about the facts about one's mental life? Are they not real and objective facts as well, or is even ordinary objectivity a misguided aspiration when *self-knowledge* is in question? Against this suggestion we might remind ourselves of the fact that we commonly take the question of what someone *else's* belief is to be an ordinary objective matter. (Obviously we can't be thinking of 'objectivity' in terms of 'mind-independence' here. Mindless entities don't have beliefs, yes; but this doesn't make the facts of someone's mental life any less real.) Thinking doesn't make it so in the "external world"; no more does it do so in the "inner world." Hence, one might insist, *what* my particular mental life is at any moment is something independent of what I think about it, how I interpret it, or indeed whether I am aware of it in any way at all. If we

are commonsense realists about wrens and writing desks (where this is intended as a vernacular commitment, with minimal philosophical baggage), then we are, or ought to be, commonsense realists about our beliefs and other attitudes as well.

In addition, apart from any issues in the metaphysics of mind, we might also be concerned to preserve some sense of objectivity here in order to make sense of various features of the ordinary phenomenology of self-knowledge. Here we may think of such qualities as the *effort* involved in self-reflection, the struggle to get something right and the characteristic risks of being wrong. Even though introspective awareness does not base itself on observation of behavior, and even after we have weaned ourselves from the picture of observation directed to an interior, there remains the sense that one's reflection is *answerable* to the facts about oneself, that one is open to the normal epistemic risks of error, blindness, and confusion. Doesn't this require the idea of a "fully independent object" and the notion of "consciousness as representation" that goes with it? I can't make sense of my own efforts at truthful self-interpretation if I take my interpretation to be constitutively self-fulfilling, making itself true.

Unlike the claim of conceptual dependence we examined in Wright, the claim here—that one's state of mind is in some way conceptually dependent on how one interprets it—does explicitly restrict itself to a *first-person* phenomenon. No one else's interpretation of my mood is granted this constitutive role, thus it is intrinsic to the claim itself that it describes some asymmetry between self and other. But it is nonetheless an idea with many strands, not commonly distinguished. For one thing, we must distinguish the kind of phenomenon described in the quotation from Charles Taylor, which concerns the role of specific self-interpretations, from another familiar claim about the role of general conceptual capacities. It is hard to deny, for example, that a shallow or impoverished vocabulary for emotional self-description makes for a shallow emotional life; and, conversely, that richer conceptual resources make for correspondingly enriched possibilities of emotional response. A person whose conceptual universe of the emotions is lim-

ited to the two possibilities of feeling good and feeling not-so-good will certainly fail to be subject to (and not just fail to *notice*) the range of responses possible for some other person with the emotional vocabulary of Henry James. This sense of conceptual dependence is not only properly first-personal, but also restricts itself to the range of phenomena we have identified as posing the philosophical problems of self-knowledge; roughly, psychological phenomena identified under our ordinary psychological concepts (and not what might be described as their "subpersonal" components). That is, there is no temptation to invoke any such conceptual dependence when it is a matter of knowing other sorts of facts about oneself. Someone who cannot distinguish shame from embarrassment (or better, someone who recognizes no such distinction in principle, seeing only various forms of discomfort) will not be our best candidate for either emotion, whereas someone who cannot distinguish a heart murmur from heartburn could nonetheless unambiguously have either the one condition or the other.¹ My being in a particular medical (including neurological) condition does not depend on my conceptual capacities for understanding my condition. Why should it be different in the case of my envy or gratitude?

One basic reason for this difference is that someone's envy and gratitude are themselves attitudes, modes of understanding the world as well as oneself. We should expect that, unlike a medical condition, a particular mode of understanding the world will only be available to someone with particular conceptual resources. We presume a background of such resources when we routinely talk about anyone's exercise of some intelligent ability. We unhesitatingly refer to the child as sorting blocks by color. But then we look closer and find that the groupings are really only into what we would call lighter and darker colors. So maybe it's not *colors* that are being sorted. Or we find that the child cannot or will not sort anything other than blocks (not sticks, for instance), or only blocks of a certain size; nothing else is treated as

¹ See Williams (1993): "What people's ethical emotions are depends significantly on what they take them to be" (p. 91), where he cites Gibbard (1990, esp. chapter 7).

having a "color." So maybe the child isn't really sorting anything at all. When the background of conceptual capacities is cast into doubt, so is the original attribution of the activity or attitude that is the expression of those capacities. By contrast, although someone's ulcer may well be the symptom or by-product of his resentment, we can identify the condition of his stomach lining without making any assumptions about his conceptual capacities. This is so because the ulcer itself is not an attitude or way of seeing the world. (This is but one of the complexities and ambiguities entailed by the idea of "mental illness," and the whole medical model of emotional disturbance.)

Of course, to say that the identity of one's state of mind depends on one's general conceptual and descriptive resources is not to say that one's interpretation of it makes it what it is, nor even that, for example, one must conceive of oneself as feeling precisely *shame* in order to be in a state of shame. The idea of this variety of conceptual dependence refers to conceptual *capacities* and their implications for a person's emotional life, and does not itself entail anything at all about the consequences (causal or logical) of any particular interpretation one may adopt. That is, the claim about conceptual resources does not tell us that any particular self-interpretation made by the person must count as *right*, or even that altering the conception of his state has any consequences at all for its identity. For all that has been said so far, it could be that someone's sophisticated vocabulary for self-interpretation coexists with, or even contributes to, chronic illusion on his part about his actual state of mind (a Jamesian theme of its own.) And not only may his self-interpretations fail to be constitutively self-fulfilling, we could also imagine that his shifting self-conceptions make no difference at all to the underlying mood he is presumably trying to capture.

2.2 SELF-FULFILLMENT AND ITS DISCONTENTS

To move, then, to the level of particular self-interpretations, it may be claimed that conceiving of oneself in a certain way is *necessary* for

being in a certain emotional state. And there are two versions of such an idea. The stronger idea has it that if a person is to count as being in a state of, say, envy or indignation, she must see herself under the concept of that emotion. This is normally an idea entertained only about some range of the more social, sophisticated emotions, and does not suggest itself with respect to various states of fear or rage we take ourselves to share with those humans and other creatures who don't go in for self-interpretation in the first place. But even with respect to the conceptually complex responses of envy or guilt we may have our doubts about such a claim of necessity, since such an idea amounts to the claim that one could not be in a state of envy unknowingly, or while failing to understand oneself that way. Hence, a more modest claim about necessity on this level would claim that, for instance, being envious requires that the person have a certain range of thoughts about herself and her situation, which may or may not include the specific thought that she is envious. For at least a range of cases, such an idea seems undeniable, and not without its importance for the moral psychology of understanding oneself. What is less clear is its relation to the idea of 'self-constitution' and the related doubts we've encountered about the idea of '[self-] consciousness as representation'. It doesn't seem that this more modest claim of necessity should have any such epistemic or ontological consequences.

Rather, the language of 'self-constitution', in Taylor and elsewhere, suggests that the logical relation in question is one of sufficiency and not necessity. The idea is that adopting a different formulation of one's state constitutes it as different, that is, suffices for a new description to be true of it, and this is importantly unlike one's interpretations of other people and other things. This thought can also be read in more than one way: either as claiming that the self-interpretation constitutes one's state so as to *conform* to that interpretation, thus making itself true; or more modestly, as claiming that adopting a new interpretation of one's emotional state suffices to constitute the state as *somehow* different, now requiring a different characterization from anyone. These ideas are not always distinguished, and it is not obvious that either one of

them by itself should threaten any ordinary ideas about truthfulness or objectivity about ourselves, or the phenomenology of self-knowledge as the effort at the "representation of something independent." Here again we're presented with an apparent tension between a common-sense realism about mental life (or a "substantial epistemology" for self-knowledge) and some special feature of the first-person point of view. Taylor and other writers who pursue the idea of 'self-constitution' have drawn attention to genuine and important asymmetries between the understanding of oneself and others, but they and their philosophical opponents seem to share the assumption that the idea of such asymmetries is incompatible with the "substantiality" of self-knowledge.

Of the two versions of 'self-constitution' involving logical sufficiency, the more radical one claims that, at least for a certain range of cases, the person's own interpretation of his state suffices for its being that very way. On such a view, interpreting myself as, say, ambivalent, mistrustful, or ill at ease makes it the case that I am correctly characterized in those very terms. In cases like these, we may feel that "thinking makes it so" because there's simply nothing to choose between, say, *taking oneself* to be ill at ease and really being so. One condition is as distressing as the other, after all, and perhaps this is because the one simply is the same thing as the other (or a sufficient condition for it). But what makes for such cases, and how special are they within the general field of the psychological? Consider a nonpsychological case with a similar self-fulfilling character. If someone sees his marriage as a failure we may well feel that in virtue of that fact alone the marriage is indeed, to that extent, some kind of failure. To say it is "to that extent" a failure expresses the sense that for one partner in a marriage to feel it is a failure is, for logical reasons, a *constituent* of its failure, and perhaps a decisive one. No marriage can be a happy or successful one if one of the partners sees it as a mistake or a trap. Thus, certain possibilities of *simply* being wrong, just mistaken in the apprehension of the case, are not available here. At the same time, of course, there will always be logical room for hysteria and overreaction, and things

don't have to be as disastrous as the person takes them to be. But even someone's exaggerated reaction makes it the case that not all is well. And the fact that this conceptual dependence will be common knowledge between the two people creates its own depressing and intriguing possibilities for unhappiness. The chronically or neurotically hurt and dissatisfied person knows that here he cannot be merely or simply *wrong*, and hence he knows that retreating to this position is a permanent possibility open to him, and one that he cannot be brought back from in the ordinary ways (e.g., by someone trying to show him that he is mistaken, that in reality things are quite different from how he sees them). Hence, the logically self-fulfilling character of the belief can be exploited to produce consequences that take it from a partial truth to something closer to the whole truth (when the breaking point is reached after the long sulk, the unreachable withdrawal, and the partner capitulates with "Fine then! You're quite right!" Bargaining relations in general provide rich opportunities for the constituting self-assertion of failure).²

Naturally, not every self-interpretation will work this way. In thinking about what makes for the self-fulfilling character of such descriptions, both psychological and nonpsychological, it is hard not to be struck by their "negative" or undermining character. If taking oneself to be ambivalent or mistrustful is a decisive constituent of actual ambivalence or mistrust, this must be because the contrary states of wholeheartedness or trust are themselves tacitly defined by the *absence* of certain thoughts. Any condition defined in such a way can be undermined by the presence of thoughts it logically excludes (as in the familiar paradoxes of deliberate spontaneity or unselfconsciousness). Certain basic forms of trust are defined by the absence of corresponding doubts. If I come to doubt my trust, and conceive of myself "negatively" as lacking trust, or as ambivalent, then I *am* mistrustful or ambivalent,

² In his *belief*, he cannot be simply mistaken; hence, the danger is the tacit assumption that he must therefore be *right*. But the original error lay in his casting it in the form of a question for belief or disbelief in the first place. As we will see, this problematic does indeed have correspondences in the case of self-understanding.

at least to the extent of being prone to such thoughts. We may note, as a piece of metaphysical unfairness, that the converse "positive" claims do not hold. A person will not be wholehearted about his work just in virtue of his conceiving of himself that way, nor will his marriage be successful just in virtue of his interpretive say-so. The possibilities for self-deception and plain deception are all too familiar here. In some cases, some such "positive" interpretation of one's situation may well be a necessary constituent of the state in question, but it will never be sufficient. Thus, as far as the capacity for 'self-constitution' goes, it will always be easier to constitute oneself in compromising and undermining ways than to constitute oneself as unified and wholehearted.

There is an asymmetry of entailments here, whereby it is only for the "compromising" states that their self-nomination has a self-fulfilling character. But what sort of unfairness is this really, and where should we lay the blame? Conditions like wholeheartedness or confidence are normally understood absolutely, as the complete absence of any contrary attitudes, whereas a condition like ambivalence is understood to be partial, a matter of degree. (What, indeed, would it even mean to speak of total ambivalence, undiluted by anything else?) If we conceive of the two opposed conditions in this way, then it will be a matter of logic alone that the absolute one is difficult to attain and easy to undermine, whereas the partial one remains as a kind of permanent possibility, easy to fall back into and difficult to emerge from.³ Interpreting oneself as wholehearted will not suffice to make one so, because wholeheartedness is a complete state: it is the absence of any ambivalence, whether conscious or not. By contrast, seeing oneself as ambivalent will suffice to make one ambivalent, because ambivalence is a partial state, defined by the presence of *any degree* of doubt, and not itself incompatible with the simultaneous presence of contrary thoughts. Add to this the fact that conceiving of oneself as ambivalent

³ Recall Wilde's Lady Bracknell: "I do not approve of anything that tampers with natural ignorance. Ignorance is like a delicate exotic fruit; touch it and the bloom is gone."

is itself a kind of self-doubt, and we get the result that only with respect to the "compromised" condition can thinking make it so. On the other side, there is no similar bootstrapping oneself into the states of wholeheartedness or trust one may associate with one's better self. As far as this sort of 'self-constitution' goes, the only possible direction is downwards.

But just because conditions like ambivalence and mistrust *are* partial or "mixed" states, this unfairness may be more apparent than real (or more self-inflicted). For it remains an open question precisely *how far* we have deviated from the condition we have defined as a "complete" one, as well as just how much this matters. It may not be very far at all; indeed, in a given case it may be just far enough, perfectly appropriate. The asymmetry depends entirely on one condition being defined as the total absence of any competing thoughts. There would be no asymmetry, no metaphysical unfairness, if both states were conceived of either absolutely or partially. And it only seems depressing, or unfair, because we forget that the fragility of the one state and the virtual inescapability of the other depends on the fact that the inescapable state is conceived of as a partial one, and thus, as far as any a priori argument can show, may approach asymptotically to the fragile "positive" state, so closely as to make the difference between them negligible. Exiling oneself from wholeheartedness may well be too easily a matter of self-fulfilling interpretation, but it may also be relatively shallow. Despite the imagery of contamination here, the depth of one's ambivalence may on this or that occasion be no greater than that of a passing thought.

If there is, for such logical reasons, a self-fulfilling character to self-interpretations of this kind, we shouldn't expect it to apply to more than a special range of cases. These will be, roughly, "partial" conditions (like ambivalence), which are defined as the contrary of some "absolute" state, which is defined so as to be incompatible with any compromising doubt about itself. This is not a trivial range of cases, but at the same time we certainly do not conceive of most of psychological life in this way. 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.

2.3 THE WHOLE PERSON'S DISCRETE STATES

I mentioned a more modest version of the idea of self-constitution, and this is the idea that for a person to alter his conception of his own mental life logically suffices for *some* significant difference in the truth about his state, though perhaps not such as to make that very interpretation self-fulfilling. There are a number of ways in which a person's reconception of his state of mind will require an altered description of his state, and for reasons that are first-personal, not shared by anyone else's conception of him. There is a sense in which such an idea accords a special "privilege" to the person's self-conception, since it is only his own conception of his state, and no one else's, that is claimed to have this logical character. But the idea of "privilege" here should not prevent us from seeing that this status given to the person's own conception does not depend on his interpretation being *true*, let alone true because it is self-constituting. One reason for this is simply that even someone's false conception of his state is part of the very person we want to understand, and so, for example, is at least *relevant* to understanding the emotion in question ("what must his envy really be like if he's inclined to misdescribe it in this way?"). Even someone's fairly gross misrecognition of his desire or fear will nonetheless be an important indication of the nature of his attitude itself. But we can see a more important reason for allowing even false self-conceptions to make a difference to what they're directed upon, once we drop the pretense that an emotion or other attitude is something like an atomistic particular. For, consider two people who both feel gratitude toward some benefactor. We might think of one person as more naive, since he sees his gratitude as something simple and free of any ambiguity, whereas the more sophisticated or cynical person sees his gratitude,

and perhaps all gratitude, as bound up with resentment and aggression toward the benefactor. These two people may both be grateful, but they will undoubtedly feel and think and act differently in the expression of their gratitude, and we would expect the histories of their emotions to follow different courses. Any description of them which left out of account the differences in how they conceive of (their) gratitude would not simply be incomplete, but would be seriously misleading. Admitting all this, however, need not prevent us from imagining the case as one in which the sophisticate is seriously wrong, or misguided about himself. There is still room for the idea of accuracy and truthfulness in this domain, and for the attendant risks of error and illusion. His interpretation of his gratitude as resentful does not constitute it as such, any more than the naive person's self-understanding makes it the case that his gratitude is innocent. (And if we do see a self-fulfilling aspect to the "sophisticate's" compromising understanding of his gratitude, this will be, I suspect, under the influence of the metaphor of contamination discussed in the previous section.) But in both cases we can see that while we retain the applicability of notions of error and accuracy, it is at the same time true that a proper account of the person's state, either from the "inside" or the "outside," cannot be indifferent to his own conception of it, cannot just dismiss it as one more flawed opinion.

Retaining the possibility of being wrong does not mean that we abandon the appearance of a self-other asymmetry here. A false conception of one's state can constitute a difference in its total character, and still be false for all that. Someone may see his pride as sinful, but if there is no such thing as sin (really), then surely his conceiving of his pride this way cannot constitute it as such. Nonetheless, it will remain true that the presence of this self-interpretation suffices for his pride to be of an essentially different nature from someone else's pride, or from his own pride before he came to see it that way. And this will be true for reasons specific to its being *his* interpretation of his pride, and not someone else's. One reason for this anticipates a point to be

developed shortly, concerning the "outward" as well as "inward" direction of the self-interpreter's gaze. For the person to see his own pride in these terms means for him to see the things he is proud of in certain corresponding terms. For him to interpret his pride as sinful is for him to see his wealth, his power, his beauty as essentially *unworthy*, and to see their allure as so much temptation, appealing to spiritual weakness. The more settled and unambiguous his disparaging estimation of these things is, the closer we as interpreters would come to incoherence in our own description of him, if we blithely persist in speaking of his "pride" when by our own lights this attitude has lost any connection with the ordinary requirements of seeing the object of one's pride as valuable, admirable, distinguishing. Self-interpretations are bound up with the rational-conceptual grounds of the particular attitudes; hence, past a certain point of apparent irrationality in the state we may ascribe to the person (e.g., "perverse pride"), it will be the retention of the *ascription* of pride that becomes incoherent. But well before this point is reached, it remains true that the specific terms in which the person understands his own state play a role in making it the kind of state it is. Hence, contrary to what is usually assumed, the hermeneutic privileging of self-interpretations (whether individual or social) does not require the assumption of their truth. Any outsider who wishes to understand or even to describe this person's pride at all accurately must include the fact that he interprets it in these terms, that he experiences and lives out his pride under these particular concepts.

The outsider must include these facts not only in anticipation of their likely influence on the person's pride or gratitude, but also for the logical reason that the condition he seeks to describe is a condition of the whole person. That is, the very object of the outsider's interpretation includes the person's second-order as well as first-order attitudes. The interest taken in someone's gratitude, whether by himself or by another person, is not an interest in some discrete state. Rather, it is an interest in the total orientation of the person toward his benefactor. We would have no use for a notion of a state of gratitude about which it could sensibly be asked whether 'it' remained the same when it was

regarded by the person himself as dominated by feelings of resentment, unaccompanied by any desire to acknowledge the benefit received, or as the expression of a kind of shameful neurotic dependence. The original ascription of a state such as gratitude is already intended to take into account such aspects of the person's total outlook. This shows that there is a purely logical dimension to the idea that a difference in the person's own interpretation of his attitude makes a difference to what his attitude actually is, that the self-interpretation is "partially constitutive" of its object. This is due to the fact that, in the ordinary case, the total state of the person we want to characterize includes the reconception itself.

2.4 BELIEF AND THE ACTIVITY OF INTERPRETING

However, in speaking of 'self-constitution' Taylor and others clearly have something other than this logical claim in mind as well. The general idea is often illustrated by reference to the situation of a person articulating his emotional state, and a kind of causal language is invoked to describe the phenomenon. Adopting a new description of one's state is said to *alter* it: "The understanding helps shape the emotion," changing how we feel.⁴ And indeed, the failure to get clear about the difference between causal and logical aspects of 'self-constitution' contributes to the appearance of mystery about the whole phenomenon, a matter of word magic or "the power of naming." Presentations of the idea of the special status of self-interpretations will sometimes make it appear as if the mere act of redescribing something (e.g., some aspect of one's own state) were being credited with the power to transform it. Not only would it remain mysterious just what sort of process

⁴ "Formulating how we feel, or coming to adopt a new formulation, can frequently change how we feel" (Taylor 1981, p. 100). The reference to what "frequently" happens is itself enough to suggest that it is not a logical matter at issue here.

this is, but such a picture has voluntaristic implications from which Taylor himself is concerned to distance himself.⁵

These unwelcome implications are a consequence not only of unclarity about the combination of causal and logical considerations, but also of a peculiarity in the description of the particular cognitive act involved. Taylor and others use a host of terms for the self-interpretive activity that is said to be constitutive of our emotional responses: we adopt new formulations, vocabularies, or languages for our emotions; we interpret or articulate them differently; we "see" them this way or that. What is striking about this list is that, while these are all cognitive activities, the basic idiom of "belief" is consistently avoided throughout the discussion. And its avoidance highlights the fact that the remaining favored terms ("interpreting," "describing," etc.) all refer to what are plainly *actions* of one sort or another, and hence which *can* be performed for more than one kind of reason. The reason an activity like "describing" can be undertaken "arbitrarily" is that one can describe or interpret something this way or that, without assuming any commitment to the way things actually are. I can *describe* the Vermeer before me as a Rembrandt, perhaps as an exercise I am given to do, and perhaps even learn something thereby, without any change of mind as to who did the painting. Or I can *see* the figures in some diagram as either receding or advancing, and switch back and forth at will, without altering my judgment about anything. In this way, "describing" and so on is like "saying," which need not even involve the pretense of expressing one's actual thought. By contrast, "believing" does not describe an action that can be arbitrarily undertaken, precisely because it does involve such a commitment. A belief as such is answerable to one set of reasons, reasons connected with its truth, and is not an action performable in response to a request or as an exercise. This is why insofar as the idea of self-constitution is cast in terms of

⁵ "(Of course, this is not to say that we can change our emotions arbitrarily by applying different names to them. We are not talking about a process that could be arbitrarily undertaken)" (Taylor 1977a, p. 70; see also pp. 64–65).

genuine activities like interpreting or describing, it will persistently raise the question of voluntarism and arbitrariness. For as far as mere description goes, a person can *describe* his own state any way he may choose, for a host of different kinds of reasons. And if *this* sort of activity is said to have the power to alter the state itself, then it is hard to see this as other than a special ability each person is credited with, to be exercised at will.

It will serve not only to demystify this last version of the idea of self-constitution, but also situate it within some of the ordinary problems of self-understanding, to show how any genuinely transforming role for the interpretation of one's state depends precisely on an understanding of 'interpretation' and the like which does *not* denote a voluntary capacity. (And conversely, to the extent that we *are* talking about an active or arbitrary capacity, it will be one without any special power to alter anything.) Verbs such as 'describe', 'interpret', and the like are fated to equivocate between a use that expresses one's genuine sense of how things are, with the same kind of commitment as belief, and a different, noncommittal use denoting an ordinary activity. Favoring verbs in this latter sense serves both to dramatize the idea of self-transforming redescription, and to obscure its genuine basis. Yet, at the same time, the presentation of the idea also requires the other, fully cognitive or "committed" sense of these verbs to play a tacit role. When Taylor speaks in the first passage quoted of "com[ing] to see that my feeling of guilt was false," he is surely talking about changing his beliefs about it, coming to believe that it was false or baseless. Yet, to believe this about one's feeling of guilt would require the same sort of reasons one would require for believing anything else and is not something to be performed at will.

In one way, it is just what we should expect that when someone comes to believe that his feeling of guilt was false, the emotion itself changes. Not only does his understanding of it change (he may now think of it as compulsive or neurotic), but what he feels changes, too. He may now admonish himself for the guilt itself and feel a mixture of relief and embarrassment. But in such a case the identity of what he

did feel before he changed his mind about its nature is not constituted or even affected by the understanding of it that he later comes to accept. This change does not challenge the status of the original emotion as a "fully independent object"; only it has now been replaced by a different one. And insofar as we want to talk about "objects" here, it remains true that no object is so independent that it remains unaffected regardless of whatever else goes on in the world (including the "inner world").

The fact that mere beliefs about my emotions can alter what I feel *would* be surprising if the emotions themselves were not attitudes directed toward something. Coming to believe that some fear of mine is unfounded will normally change my emotional state, replacing fear with something else, perhaps relief. My fear was *about* something and dependent on my beliefs about that thing. This cognitivity is also the reason why *noncognitive* states, such as physical pain, are (regrettably) considerably less sensitive to our beliefs or to our understanding of them. One needn't claim that emotions simply *are* a species of belief in order to view them cognitively and as falling under rational criticism. A familiar fact about some emotional states is that they do *not* alter when the beliefs on which they are based are sincerely denied by the person in question—for instance, phobias that the person is aware of as such and which survive the person's recognition of them as baseless fears. (It is, of course, equally familiar that we often criticize such a state as irrational.)

So, if we can think of altering the interpretation or articulation of an emotion as involving a commitment akin to changing one's *belief* about it, and we view the emotion itself under its aspect as an attitude, we have something belief-like on both sides of the relation. In this way we may hope to shed some light on the "self-altering" character of certain self-interpretations by consideration of what are called 'second-order' attitudes, for example, beliefs about one's own beliefs. Now there are certain familiar situations of self-reflection in which determining the character of one's state of mind exhibits a peculiar shiftiness. In such cases, the more one tries to focus directly on one's current

thought and feeling (e.g., about last night's quarrel, movie, phone call), the less definite or constant one's state of mind seems to become. If we picture what one is doing here as tracking down an inner state, then it can seem like a search whose object mysteriously changes just as one's introspective gaze is directed upon it. But even if this is a misleading picture, something like this general phenomenon with regard to beliefs is what we should expect. If a person is at all rational, his first-order beliefs will indeed be sensitive to his second-order beliefs about them, and they will change accordingly. He may, for instance, discover that some set of his beliefs is inconsistent, or suspect that a particular belief of his is the product of prejudice or carelessness, or, at the limit case, that it is just plain false. His first-order beliefs will then normally change in response to his interpretation of them. Here the misunderstanding involved in putting such facts of change simply in terms of offering new *descriptions* of one's emotion becomes clear. For a new description of my emotion or belief is powerless to alter it unless I *believe* the description. Clearly, it is the actual believing that is crucial to this change, and not the activity of naming or describing.

2.5 THE PROCESS OF SELF-CREATION: THEORETICAL AND DELIBERATIVE QUESTIONS

Such redescrptions of one's state of mind may be arrived at in more than one way, and only for some of them will descriptions of psychological life function differently in the first-person and the third-person cases. There is more than one spirit in which a person may reflect on his psychological state, and these involve corresponding differences in how such understanding may contribute to an altered state of mind. For example, with respect to knowledge of one's own intentions, philosophers sometimes invoke a distinction between certainty that is based on evidence or discovery, and certainty that is based on a decision made by the person. According to this distinction, uncertainty about what one intends to do is normally a matter of one's having not

yet fully *formed* an intention, and this uncertainty is ended by a decision about what to do rather than by a discovery of an antecedently formed intention.⁶ The question expressing this uncertainty will *not* indicate a situation in which there is something I intend to do but I don't yet know what it is. Rather, the question expresses the fact that my intention itself is uncertain. This is practical and not theoretical uncertainty, and the resolution of it is a decision, rather than a prediction of what I will do. Ending my uncertainty about, for example, what I will wear is indeed coming to know something; and, like other things I know, I can tell it to someone else so that they know it, too. But in other ways, although we rightly speak of knowledge here, it is not purely a theoretical or epistemic matter. My knowing what I will do next is not based on evidence or other reasons to *believe* something, so much as it is based on what I see as reasons to *do* something. Hence, a person's statement of intention is not to be challenged by asking for his evidence. When I make up my mind about what to do, and tell someone else, I do indeed provide him with a reason to expect something, a very good reason if I'm not too vacillating, or a liar; but what I possess myself is not an expectation, based on evidence, but an intention, based on a decision.⁷

The question, "What am I going to do?" may seem to have *only* a practical and not a theoretical application, that is, never to express inquiry into some antecedently formed intention. If this is so, it will be due only to the difficulty in imagining a situation in which a person would both need to seek to learn what his intention is, *and* have good

⁶ For this distinction, see Hampshire and Hart, (1958), and a later discussion by Grice (1971) and Anscombe (1957). See Velleman (1989) for a very different picture, according to which intentions are theoretical predictions to which we have a standing motive to conform our actions.

⁷ Here I have drawn on Hampshire (1975), invaluable for thinking about these issues. It is sometimes argued that knowledge of one's future actions is indeed a matter of prediction, since it must rest on ordinary inductive evidence, relating to one's abilities, the empirical features of the world one relies on in planning, etc. But the claim made here is not that inductive knowledge is irrelevant in making up one's mind, but that after all such evidence is in, it remains for the person to decide how he will go.

enough reason to believe that he *has* some settled intention in the matter. The difficulty here is a function of the fact that even when we can imagine a situation of theoretical blindness to one's own intention, it will be hard to see this as something like mere ignorance, a gap in awareness which reflects no (practical) conflict in the intention itself. The problems with the idea of blindness to oneself of this sort do not depend on Cartesian assumptions about the mind's self-transparency. Even within a psychoanalytic explanation it will normally be the case that the contrary thoughts and attitudes which explain the subject's blocked *awareness* of the intention will themselves be reasons for ambivalence *in* his overall intention; that is, the intention itself will not be a wholehearted one. Ignorance in such a case will not be mere ignorance, not only because it will be irresistible to look for a *motivation* of sorts to explain it, but because the motivation we then impute to the person must qualify the original ascription of the intention (as conflicted or partial).

With respect to states other than intention, in any case, it's clear that the same words may express either the aim to identify a certain state of mind, or the desire to make up one's mind about some matter. Thus, for instance, saying "I don't know what I want" may either express a divided consciousness containing certain definite though repressed desires, or it may express someone deliberating about what's desirable, choosing between certain options. In the latter sort of case, the uncertainty about one's desires is an uncertainty or indefiniteness in the desires themselves, and resolving the uncertainty is a matter of forming one's desire.⁸ And with respect to one's emotional life, a person may want to know what his true feelings about something are, or he may be engaged in making up his mind, coming to some settled response he can respect, or at least make sense of; that is, his inquiry may be either a purely theoretical one about his psychological state as it is, or part of the process of forming his feelings.

⁸ Hampshire describes the situation of such a person, in which "the conclusion of his considerations will be a decision from which a definite desire emerges. He now

What we're calling a theoretical question about oneself, then, is one that is answered by discovery of the fact of which one was ignorant, whereas a practical or deliberative question is answered by a decision or commitment of some sort, and it is not a response to ignorance of some antecedent fact about oneself. When self-reflection concerning one's emotional response is of this latter sort, the declaration "I don't know *how* I feel about that" is equivalent to "I don't know what *to* feel about it"; that is, how to settle and sort out the various conflicting elements of one's immediate reaction. The corresponding theoretical question would be of the form, "I don't know what it is that I do feel," where this is equivalent to the kind of question that is a frequent occurrence in the literature of romantic love, "What is this that I feel?" The person asking this might think to himself, for instance, that previously he never thought he had any particular feelings for a certain other person, though sometimes he felt self-conscious in her presence; but now both these feelings to which he is attending and his past behavior make him believe that he *has* some definite emotional response, but he isn't sure what to call it. Perhaps when he sees her with other people, or just with certain others, he finds himself feeling something that could only be jealousy, but at first he can see no reason why he should feel anything like *that*. Naturally, if this inquiry remains a *purely* theoretical one for him, separate from the question of what he *is* to feel, his emotion is likely to be inapt or fixated, whatever it turns out to be. This would be to reduce the emotion to an interior occurrence to which he is passively subject. This, too, is familiar in the literature of romantic love. Goethe's Young Werther is a subtle and discerning observer of the movements of his own mental states. His interest in them, however, is primarily a theoretical, contemplative one, separate from questions about the world that those states of his are presumably directed upon.

When the articulation or interpretation of one's emotional state plays a role in the actual *formation* of that state, this will be because the

knows what he wants to do, because he has now formed his desire, and not because he now knows how a pre-existing desire is to be characterised" (1975, p. 52).

interpretation is part of a deliberative inquiry about how to feel, how to respond. And there one's attention will be directed at least equally outward, toward the object of one's response, as it is directed toward oneself. The idea of "deliberative" reflection about one's response is meant to denote something more than simply the normative appraisal of it, the sort of reflection that would terminate in some settled assessment of it. For the mere appraisal of one's attitudes, however normative, would apply equally well to past as well as to current attitudes, and indeed may have just the same application to another person as to oneself. In itself, such assessment is not an essentially first-person affair. Rather, "deliberative" reflection as intended here is of the same family of thought as practical reflection, which does not conclude with a normative judgment *about* what would be best to do, but with the formation of an actual intention *to do* something. Similarly, in the sorts of cases mentioned where seeing one's feelings of guilt as false, or one's anger as childish, "helps shape the emotion," this will be because the "seeing" in question is not purely theoretical or descriptive (even where such description includes evaluative assessment), but is rather an expression of the ordinary deliberative reflection about how *to* feel.

This distinction between theoretical questions and deliberative ones introduces a new dimension to the issues we've considered concerning the objectivity of the mental and the substantiality of self-knowledge. 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. Much remains to be said to clarify just what this role is, and the implications it has for such problems as "being objective toward oneself," in the sense pertaining to moral psychology rather than metaphysics and epistemology. But we should not confuse the introduction of the agent in self-reflection with either abandoning ordinary realism about the mental or denying a substantial epistemology for self-knowledge. It is indeed essential to our nature as persons that we are "self-interpreting animals" (in Taylor's phrase), and that the exercise of this capacity plays a crucial role in making us who we are.

And the 'self-constitution' in question here is genuinely substantial and productive and not merely a matter of something like 'logical construction'. But rather than seeing this as undermining our ordinary ideas about the reality of mental life, we should notice at this point that for *any* concept of the mental rich enough to include attitudes toward one's attitudes, this sort of mutual responsiveness is just what psychological health would involve. Hence, the question of the "independence" of the person's state of mind from his interpretation of it now takes on a very different character. For when a person's emotional state is independent in *this* sense from his other attitudes toward it, what this means is that the two sets of attitudes are cognitively isolated from each other, and thus that, for example, one's deliberative reflection on what's desirable leaves one's actual desires unaltered. Although it is a familiar enough condition in itself, this is a form of impairment that is not demanded by either the epistemological or the moral virtues of objectivity.

2.6 RELATIONS OF TRANSPARENCY

We will gain a clearer view of the interplay between these two types of inquiry by considering a related claim about how a question about one's own belief must present itself, from the first-person point of view. In the end this will also help put in proper perspective the issue of the role of the person as agent in the formation of his attitudes. Ordinarily, if a person asks himself the question "Do I believe that P?", he will treat this much as he would a corresponding question that does not refer to him at all, namely, the question "Is P true?" And this is not how he will normally relate himself to the question of what someone else believes.⁹ Roy Edgley has called this feature the "transparency" of one's own thinking:

⁹ The question, "Do I believe P?" might be better phrased as "What do I think about X?", where X is some issue or possibility rather than a particular proposition represented

[M]y own present thinking, in contrast to the thinking of others, is transparent in the sense that I cannot distinguish the question "Do I think that P?" from a question in which there is no essential reference to myself or my belief, namely "Is it the case that P?" This does not of course mean that the correct answers to these two questions must be the same; only I cannot distinguish them, for in giving my answer to the question "Do I think that P?" I also give my answer, more or less tentative, to the question "Is it the case that P?"

(1969, p. 90)

And more recently, Gareth Evans made a similar observation in connection with a remark of Wittgenstein's directed against the idea of self-knowledge as involving an "inward glance":

[I]n making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me "Do you think there is going to be a third world war?" I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question "Will there be a third world war?"

(1982, p. 225)

To claim that one question is "transparent" to another is not to claim that one question reduces to the other. The fact that answers the question about the war is different from the fact about a particular person's belief. As Edgley points out, the correct answers to the two questions need not be the same. But nor is it right to say, as he does, that the two questions are indistinguishable from within the first-person point of view. After all, it isn't as if, although the answers to the two questions are in fact distinct, I must remain somehow in the dark about this, or that I cannot see them pointing in different directions. It will be common knowledge, among anyone with the concept of belief, that al-

by P. This is both somewhat more natural and more clearly at home in a context of deliberation about some matter.

though one believes something *as* true, the fact believed and the fact *of* one's belief are two different matters. From within the first-person perspective I acknowledge the two questions as distinct in virtue of acknowledging that what my beliefs are directed upon is an independent world, and they may therefore fail to conform to it. So, rather than reducibility or indistinguishability, the relation of transparency these writers are pointing toward concerns a claim about *how* a set of questions is to be answered, what sorts of reasons are to be taken as relevant. The claim, then, is that a first-person present-tense question about one's belief is answered by reference to (or consideration of) the same reasons that would justify an answer to the corresponding question about the world.

If this serves as a clarification of the content of the idea of transparency, it raises the question of the *kind* of claim Edgley and Evans are making, and what reason there may be for believing it. Is it true as a matter of empirical fact that, as we may put it, the self-directed question and the world-directed question are always answered in the same way? Or is the claim rather that, for conceptual reasons, we cannot make sense of the idea of answering a question about one's present belief without "attending to precisely the same outward phenomena" as one would in answering the world-directed question? With respect to attitudes other than belief, it seemed earlier that we could well imagine situations where the two questions were not treated equivalently. Someone may want to know whether it is resentment that he feels, or whether any resentment is called for in this case, whether it is what he is to feel. Or he may learn of his own desire in a way that approaches the purely theoretical or behavioral, and is quite different from any reflection on what in the world is good or worth desiring. Such a division between the two sorts of consideration may well represent a failure of sorts (of rationality, willpower, or something else), but nonetheless any equivalence between them is not something guaranteed by the logic of the first-person, but looks more like a kind of normative ideal. With respect to belief, the claim of transparency is that from within the first-person perspective, I treat the question of my belief

about P as equivalent to the question of the truth of P. What I think we can see now is that the basis for this equivalence hinges on the role of deliberative considerations about one's attitudes. For 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.

When we unpack the idea in this way, 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, deciding and declaring myself on the matter, and not confront the question as a purely psychological one about the beliefs of someone who happens also to be me. This is not to say that one normally arrives at one's beliefs (let alone one's fears or regrets) through some explicit process of deliberation. Rather, what is essential in all these cases is that there is logical room for such a question, about regret as much as about belief, and that the actual fear or regret one feels is *answerable* to such considerations. I may confess that my fear is beyond my control, and that I can't help being afraid of something where, by my own lights, there is nothing to be feared. But so long as I am to understand my condition as *fear* of any kind, even irrational fear, I cannot fail to accept the relevance, the force of the deliberative question "Is there anything to be feared here?" That may not be the only question for me in my situation, but if I cannot make sense of the point of *that* question directed at my state, then I cannot make sense of my state as fear of any kind, however irrational.

In characterizing two sorts of questions one may direct toward one's state of mind, the term 'deliberative' is best seen at this point in contrast to 'theoretical,' the primary point being to mark the difference between that inquiry which terminates in a true description of my state, and one which terminates in the formation or endorsement of an attitude. And so to speak of the person's role in forming his attitudes is not to invoke a kind of willful or wishful capacity for self-creation. A person adopts this role insofar as he can answer questions of the sort "What am I to

believe here?" and thereby come to believe something, or answer a question of the form "Is this what I really want?" in terms of considerations of what is worth wanting, and thereby come to clarify the structure of his actual desires. The fact that we do have this capacity should not be controversial, for it amounts to the idea that part of what it is to be a rational agent is to be able to subject one's attitudes to review in a way that makes a difference to what one's attitude is. One is an agent with respect to one's attitudes insofar as one orients oneself toward the question of one's beliefs by reflecting on what's true, or orients oneself toward the question of one's desires by reflecting on what's worthwhile or diverting or satisfying. This is not the only possible stance one may take toward one's beliefs or other attitudes, but it is an essential one, and it is hardly the same thing as the free or arbitrary adoption of beliefs for reasons of convenience, fear, or fashion. There is a role for the agent here insofar as we may speak of a person's responsibility for his attitudes, and we shouldn't expect this sort of responsibility to be any simpler or immune to damage or evasion than the other kinds.

Both the shiftiness and ambiguity of self-interpretation noted by Taylor and the "outward-looking" character of first-person belief reports noted by Edgley and Evans have their source in the primacy of a deliberative rather than a theoretical stance toward one's own state of mind. This suggests that the way forward on these and other issues concerning the special features of the first-person may be as much in the area of moral psychology, broadly construed, as in epistemology or metaphysics. The phenomena of self-knowledge, not to mention the wider spectrum of asymmetries between the first- and third-persons, are themselves based as much in asymmetries of responsibility and commitment as they are in differences in capacities, or in cognitive access.

Conforming to the idea of transparency between self-directed and world-directed inquiry thus appears to be less a matter of the logic of self-reference and more a matter of assuming a certain stance toward oneself and one's attitudes. As such, then, we should expect it to be something that will apply differently to different kinds of states of

mind, with different degrees of stringency, with different possibilities for and consequences of its failure or deliberate violation.

The succeeding chapters attempt to flesh out the idea that the familiar asymmetries between the first- and third-persons, whether interpreted in terms of self-constitution, or the inapplicability of ordinary notions of objectivity, realism or substantiality to self-knowledge, have both a broader application and a different basis than their confinement to more epistemological contexts would suggest. Showing this will enable us to put some of the more familiar problems of self-knowledge within a wider perspective of differences between possible relations to oneself and to others, and the temptations, philosophical and personal, to try to model one such relation on the possibilities that properly belong to the other. In particular, what we've been calling the Spectator's picture of self-knowledge shows itself to be an intellectual expression of the ordinary, nonphilosophical pressures to adopt a purely theoretical stance toward oneself in particular situations of life, pressures stemming not only from motives of evasion or "bad faith," but also from the motives of moral objectivity toward oneself. The following chapter takes up the idea of the contrast between deliberative and theoretical stances toward oneself, and the ineliminability of the demands of either one of them, beginning with the relatively straightforward case of belief.