

TOPICS IN PHILOSOPHY

1. Stephen P. Stich (ed.), *Innate Ideas*, 1975
2. John Perry (ed.), *Personal Identity*, 1975
3. Amélie Oksenberg Rorty (ed.), *The Identities of Persons*, 1976
4. Robert J. Swartz (ed.), *Perceiving, Sensing, and Knowing*, 1976
5. Amélie Oksenberg Rorty (ed.), *Explaining Emotions*, 1980
6. Brian P. McLaughlin and Amélie Oksenberg Rorty (eds.), *Perspectives on Self-Deception*, 1988

# **PERSPECTIVES ON SELF-DECEPTION**

---

---

Edited by  
BRIAN P. McLAUGHLIN  
AND  
AMÉLIE OKSENBERG RORTY

UNIVERSITY OF CALIFORNIA PRESS  
Berkeley Los Angeles London

## THE PECULIAR EFFECTS OF LOVE AND DESIRE<sup>1</sup>

BAS C. VAN FRAASSEN

Who can untie this most twisted and intricate mass of knots?  
—St. Augustine, *Confessions* II-10

To write about self-deception, I must choose a very personal style, since I bring no professional qualification to the subject, nor much learning. How self-deception is even possible is of course the first philosophical question about it. Every approach to this subject leads quickly back to the inconsistency borne in its name. Yet logical puzzles about how one could possibly believe one thing and know its opposite seem to me almost a travesty of the perplexities we feel about the phenomenon, or cluster of phenomena, we refer to as self-deception. The threat they pose, I have come to think, lies not there but in our sense that their possibility undermines any defense against skeptical doubts—not of the global kind we tend to dismiss as philosophical *vieux jeux* but of a much more intimate sort. My idea here is to look closely at some of those phenomena, after only a few pages of general reflection.

FREDERIC MOREAU AND MME. ARNOUX

1. Which phenomena, exactly, should we look at? This question cannot be answered by a definition of self-deception; at the preliminary stage all we can have is a rough demarcation. If I say that you deceived yourself, I imply that you had a false opinion, and that you arrived at it in a way that makes you culpable. For if I deceived someone else, he or she would have a false opinion, and would have arrived at it in a way that made me culpable. In central cases of deception the deceiver knows that the opinion in question is false, or at least, is of the contrary opinion. Hence in the clearest cases of self-deception we must also be able to say, "You really knew all along; or at least, if you had faced the issue honestly and explicitly, you would have

had to acknowledge the truth." But not all cases are like this. Suppose that a certain bridge is dangerous, and I do not know this, that I really have no evidence that makes it less likely than not, for me, that the bridge is safe. But suppose in addition that I think it would be very good for me if you believed the bridge to be safe, and I successfully persuade you to believe this falsehood. Then I have certainly deceived you, although I did not know that the information was false. Similarly I expect that there can be cases of self-deception in which the person cannot correctly be said to have known or even recognized the truth as the most likely. The main requirement for it to be (self-) deception, as opposed to unintentional (self-) misleading, is that he be culpable of persuading himself by epistemically unfair means or practices. This leaves us unsure of the boundary with wishful thinking, but it may be correct to call that a limiting case of self-deception.

An accusation does not, for its effectiveness, depend on the verifiability or falsifiability of the charge. This was proved at the stake and in the House Committee on Un-American Activities. If we harbor radical doubts about whether there is a coherent concept of the self-deceiver, we doubt at least that it is in principle possible to give nontrivial, adequate truth conditions for attributions of self-deception. Although an accusation always *purports* to state a solid sort of fact, such a radical doubt would not remove the possibility of charges and accusations of self-deception. It would, however, undermine such accusations (you cannot harbor the doubt and also take the accusation seriously) while explaining the language-game as a (mere) mechanism of accusation. The learned would explain the vulgar game, but their explanation would prevent them from taking part.

Such a position has the usual drawback that it is exceedingly difficult to forsake the game. I shall turn to examples below; meanwhile it will probably be agreed that we often want to say of others that they do deceive themselves, and of ourselves that we did but do no longer. Those claims may sound patently false to the audience: consider "For years I deceived myself, thinking I did not love my daughter, but really all the neglect and harsh words and the misery I caused her sprang from love and were for her own good . . ." But the automatic reaction ("I think that is false!") is equally incompatible with the radical doubts outlined above. And what do you say when someone just does not come to an inevitable conclusion, and hypotheses of dim wit or dissimulation are ruled out? "To me the evidence is clear, and you have the evidence also; the facts are staring you in the face; I see it must be more comfortable to reject my conclusion, but you are deceiving yourself if you do." What else is there to say?

This second example points to a different way to exploit the idea of accusation. An accusation can consist in a moral or quasi-moral evaluation of commonly admitted facts. To deny the accusation can then take the form of

admitting those facts, but rejecting the evaluation. The obvious examples are: to admit killing but deny murder, to admit falsehood but deny deception. And the denial can take the form of either advancing a contrary evaluation of the same sort, or denying the applicability of that *sort* of evaluation at all.

This too has a difficulty: in the case of self-deception, the (quasi-) moral evaluation would have to be of the formation of belief or the weighing of evidence—or so it would seem. There are indeed two traditions in epistemology that differ on this, but the one for which it would make sense (*voluntarism*) has, I think, never been a popular one. For most writers on the subject, weighing evidence and determining what is credible is construed on the model of arithmetic—or at best, Archimedean statics. You may not be very good at sums—but that is stupidity or ignorance. It would be very hard, I think, to explain self-deception this way if we could be relevantly guilty only of deliberately careless or incompetent arithmetic.

I do not wish to continue in this vein of exploring general possibilities for an account of self-deception in the abstract. Let us go to the facts themselves, for a while. Of course, let us approach them a little systematically, or we'll just get lost. We need to choose something, tentatively, to go on.

2. Let me hazard a first guess. There is something theoretical about the ideas of thought, opinion, and belief, and also something nontheoretical. If I want to accuse you (or myself) of self-deception, I can make the charge *prima facie* irrefutable by skillful trade on the consequent ambiguities.

In the first place, attribution of thought or thinking need not carry any implications of conscious mental activity at all. (I am not suggesting that there exists unconscious mental activity: let us leave such hypotheses to psychologists.) For example, if you ask me, "How are you?" and I reply, "Happy," I do imply that I am happy, and was happy already when you asked me, but not that there were conscious feelings or thoughts of a relevant sort preceding my answer. I was happy already; now I realize it, as I answer you. Even if I look at the sky and say, "I think it will rain," or look at you and say, "You clearly think this will spoil our walk," I attribute no conscious mental activity preceding my words either to me or to you. I merely attribute opinions. Think of how irrelevant it would be if you replied, "Not at all; that is false, for I was thinking about something else entirely, about dragons as a matter of fact." For it is possible that both are true: that you *are* thinking about dragons, and that you *do* think rain will spoil our walk.

For there is also conscious thought, which is, or is like, talking to oneself *sotto voce*. It happens as you walk along the street, deeply engrossed; it happens in little pauses between the sentences I am now writing, as I sculpt the next, though many of these sentences are just written without such a prelude. And since this activity (let me call it thought "*s.v.*" for "*sotto voce*")

is so much like speech, it offers itself at once as a base from which to make the usual, standard inference from what a person says to what he thinks (in the sense of: what he opines).

The function of thought *s.v.* is, I think, mainly preparation. I prepare for writing the next sentence or paragraph by trying out phrases in my mind; also for writing a paper, or part thereof, by trying out schemes and outlines *s.v.* Suppose now that as I walk along the street, deeply engrossed, eyes unseeing and shoulders hunched, the subject that so engrosses me is my relations with *Mme. X*. She has told me this morning that I am merely deceiving myself if I think that I still love her, that I only have to look at my own actions and gauge my own inclinations to perceive the truth. In my mind I am speaking to her eloquently and at great length, explaining how I have always felt, how she misperceived my earlier ardor and even more my present preoccupation with work and desire to travel. I hear her voice, I see her laughing at me, with more amusement than bitterness, clearly showing that she had escaped these bonds sometime before. I suspect her of incredible infidelities, of really loving her husband after all; I swear my affections have never faltered. And now, tell me if there is one single test by which these exclamations *s.v.* are not to be counted as the purest, most sincere, and most privileged expressions of belief, if anything could be?

If there is to be any inference at all from what I say sincerely, without qualification or reservation, without constraint or possibility of gain or loss, to what I really think or believe, then such an inference must surely be correct *a fortiori* now. For now I am my own sole audience, and my only aim is to rehearse a problem crucial to my own life and to my action and happiness henceforth. This thought *s.v.* is apparently preparation—that is its normal function—for speech out loud, which, if performed, may have the most far-reaching consequences for my own life. And yet, here I am, sounding like a cheap lawyer, distorting the truth to suit the impressions he is paid to create, on behalf of a client without integrity, before a judge so corrupt that he will let go unchallenged the most improbable of claims. Not only that; I half realize I will not say or write these things, except perhaps for those accusations that, if true, will go some way to justify my own actions, or make them look better.

The normal function of thought *s.v.* is in such a case abrogated. It is then a pretended speech in my own defense, before a pretended tribunal consisting of *Mme. X* (now compliant, now skeptical, as my rhetorical moves require for their setting), myself, and the world. Later, looking back on these sad episodes, I will again think *s.v.*, appear before the tribunal of my mind, and tell stories that explain what really happened. In some I will shine, and in some I shall be as guilty as if I had constructed and run Dachau single-handedly. Depending on how traumatic the episodes were, the number and

diversity of mutually contradictory reconstructions in memory, performed at the tribunal in my mind, may be quite extensive.

One crucial point about these pleas and stories that makes it possible for them to live in my mind is that so many of their lines attribute to me and to *Mme. X* thought that is not *s.v.* and attribute intentions, motives, and feelings in the same way, without necessarily implying any conscious mental activity. The point is that I can attribute thought and feeling with total impunity as far as details are concerned. For it is entirely consistent to say that you or I thought or felt this or that yesterday, while admitting we had no conscious awareness of any such thought or feeling and engaged in no action that bore explicit witness to it.

The paradigm inference to what a person thinks is from what he or she says, when there is no contraindication of gain from the effect of the saying itself. As soon as we *doubt* that self-deception is really possible, we must take thought *s.v.* to be very strongly qualified as a base for such an inference. If you say to me that you (too) love *Mme. X*, the effect you want to have on me may have more to do with what you say than your belief concerning its content. But what if you say it out loud, alone, in the middle of a forest? *A fortiori*, what if you think it *s.v.*? But (here comes the catch) if thought *s.v.* is such a very good base for that inference, and if we can evaluate it “from outside” (in memory, in our own case) as deceptive, rhetorical, lying, flouting all ethical standards for debate, then we can conclude that the person is engaged in a deception practiced on no audience but himself. These shoddy arguments, dismissals and distortions of evidence, and comforting opinions with no claim to warrant, we must then conclude to be really *his*. So, to sum up the argument: if you doubt the possibility of literal self-deception, this doubt makes it impossible to explain away the very phenomenon that, if taken seriously, gives us overwhelming reasons to say that self-deception is regrettably pervasive in situations of stress.

We have an alternative, though I don't think it is very palatable. It is to reject as a mistake of the vulgar, that paradigm inference to what a person thinks or believes. This surely results in an almost complete skepticism about inferences from thought *s.v.*, let alone from speech however sincere, etc., etc., to what anyone (even oneself) really thinks or feels. I say “almost complete skepticism,” because two paragraphs ago I had to say “as far as details are concerned.” Of course, the story in the large is made up of details, but it is less easy to be skeptical about less detailed matters. This is a matter of degree, but there are limits. In André Gide's *Strait Is the Gate*, Alissa writes in her journal about Jerome:

I strive to act according to reason, but at the moment of action the reasons that made me act escape me, or appear foolish; I no longer believe in them.

The reasons that make me fly from him? I no longer believe in them. . . . And yet I fly from him, sadly and without understanding why I fly.

We have already, perhaps, concluded that the reasons she gave were not the real reasons; although she was sincerely persuaded of them, she was deceived. But did Alissa really love Jerome? Did Jerome love Alissa? Did Juliette love Jerome? Perhaps in the end we are disinclined to say *yes* to any of these. But the disinclination is not of equal strength; we are under the impression that in trying to answer such questions, we are weighing various relevant factors, and they are not the same. In any case, no opinion based on the text can imply that Jerome loved Juliette. Conscious avowal, thought *s.v.*, and actions and behavior almost uniformly speak against it. No, it would be too farfetched; the novel has strained our credulity too far already.

Well, if it is so farfetched, why have I wasted more than three words on it? Because it can't be ruled out, except by bludgeoning it with common sense, and refusals that are just a bit too loud to be mere denials. Reasonable skepticism about what human beings really thought or felt, even ourselves, has its limits—but they are not theoretical limits.

3. The story of *Mme. X* that I sketched above does not exist. What I mean is: there is no such story. One recurrent reason for the misgivings I feel with the philosophy of mind is that it often uses as data, examples that do not exist. An example exists if it is made up, you will say; and that is right. An example could be a real happening or a story. But a cartoonlike sketch of a story is neither. Both in real life and in real literature, the observer finds himself in a context so rich that—despite the clear limitations on what he can observe—he has a basis for conclusions about thought and emotion. Cartoonlike sketches, however, do not generally give him such a base, but only *assert* that he has such a base. Examples so sketched are like Harlequin romances: they give the reader simply the conclusions he is supposed to reach. Any philosophical conclusions reached by such means are, I think, sheer moonshine, just as are conclusions about love and marriage based on Harlequin romances. They are only the opinions of the author, thinly disguised in dramatic form. There is no saying what conclusions we would reach if we came upon a factually similar episode in real life or literature. Possibly there is also a good way of proceeding by sketched examples: to ask the reader to take them as pointers to real episodes in his or her own and acquaintances' lives or in literature. But then the conclusions drawn by the author from his or her sketches must be carefully checked against the readers' own, genuinely known instances. Moreover, they must be ones the readers *can* check—that rules out most science fiction examples, and many others as well.

I want to point to a real episode in literature, to which the *Mme. X* story (as I continue it below) bears some distant relation. In Gustave Flaubert's

*Sentimental Education*, Frederic Moreau falls in love with Mme. Arnoux, and although she is unattainable, there is never anyone who can supplant her in his heart. This sentimental education of Frederic covers eleven years, from 1840 to 1851; near the end he can say to Rosanette, "I have never loved anyone but her." Then many years go by:

He travelled.

He came to know the melancholy of the steam boat, the cold awakening in the tent, the tedium of landscapes, and ruins, the bitterness of interrupted friendships.

He returned.

He went into society, and he had other loves. But the ever-present memory of the first made them insipid; and besides, the violence of desire, the very flower of feeling, had gone. His intellectual ambitions had also dwindled. Years went by; and he endured the idleness of his mind and the inertia of his heart.

Towards the end of March 1867, at nightfall, he was alone in his study when a woman came in. (p. 389)

It is Mme. Arnoux. The scene that follows puts the lie to everything Frederic has pretended. Or does it? They speak of love as of a lost treasure; he feels a violent desire for her, but does not want her for fear of disgust later. "Besides, what a nuisance it would be! And partly out of prudence and partly to avoid degrading his ideal, he turned on his heel and started rolling a cigarette."

Much can change in sixteen years, must change; this scene does not mean that he did not love her *then*. But what about that contention—which we are surely meant to understand as his—that the ever-present memory of his first love made the others insipid? Did that story begin to play a role in certain defenses, hiding the disinclination for a more exacting emotional life? And looking further back, was that love for Mme. Arnoux, never consummated and never pursued beyond very strict bounds, not present mainly in the "official" role in which it helped define his relations with others, such as Rosanette and Mme. Dambreuse? In reflections such as these we can see the mechanisms of accusation roll ponderously into action—the first and primary function of our attributions of self-deception to someone else. The doubts raised have as *one* obvious resolution the conclusion that Frederic deceives himself.

I am not in a position to resolve these doubts with any certitude and propose that we continue for now, naively, with that most difficult question: could there be self-deception in emotion? Suppose we continue with the story, unfortunately not of such literary merit, of my affair with *Mme. X*. Imagine that the scene I have already described happened long ago. Years later (*She traveled. She came to know the melancholy of steamboats, cold awakening in tents, . . .*) she let me know she wished to see me again. I wrote that I

had never forgotten her, that all loves had been insipid beside the memory of hers, that we could meet upon her return, a month hence. One week before the meeting, I fell violently in love with a woman I met on the Metro. I was delirious; she was not out of my thoughts for a moment. A few months later it was hard to recapture this feeling; it had ended peaceably enough. Meanwhile, one may imagine with what mixed feelings I approached the meeting with *Mme. X*. It was strange to meet her again: at the same time as if we had never left each other, as if no time had elapsed, and also as if we were strangers, doubtful that we had any way of gauging each other's feelings. As it happened I did not actually have to reveal to her how inextricably my heart was now involved with someone else. Before the end of the first evening, she said ruefully that the many weeks that had elapsed since she wrote to me had undoubtedly been a mistake; one should always act on one's great impulses, but she had been afraid. Meanwhile, she and Ricardo, a traveler on the same steamboat by which she had come home, had realized how uniquely suited they were for each other, romance blossoming into love . . .

What shall we say of this narrator and his paramour? Did they not really have the emotions they avowed? Let us consider only his infatuation with the woman he met on the Metro. By every test one can imagine, in speech, thought *s.v.*, behavior, decisions made day by day, one has to admit that the evidence is overwhelming that he really had those feelings, that he wanted her, experienced tenderness and compassion, cared for her, had no greater joy than to see her joy—you, reader, may supply whatever criteria I omit. So his immediate feelings were real, if any are. He was perhaps mistaken in thinking (as he surely did?) that they would last, but that is not the point at issue. Nevertheless, this infatuation was as effective a defense against the very possibility of reconciliation with *Mme. X* as can be imagined. And so it was with her feelings for Ricardo, which, we should perhaps add, lasted not much longer.

Imagine, in contrast, a slightly different scenario. Suppose that instead of becoming preoccupied with another woman, our narrator had stayed totally preoccupied with the coming meeting, but had contemplated day and night *Mme. X*'s past treacheries, real or imagined, and developed baroque hypotheses about her present plots and motives and sexual involvements. Suppose that he had convinced himself that even as he thought about her so ardently and tortuously, she was having a shipboard affair with someone on the journey home, some fascinating Latin, with a name like Roberto or Ricardo—all without the slightest shred of evidence. We would say that he was deceiving himself, just as much as a hypochondriac who refuses to acknowledge the evidence of his good health and sees pathological symptoms everywhere. And we would know why. Despite his letter (but how shall we explain that?) there was nothing he wanted less than a reconciliation with *Mme. X*, or the exacting future this might bring in train. (But how are we to explain his

making all the arrangements to meet her, and speaking already of subsequent meetings?) The two scenarios are not really so different. He was deceiving himself.

4. Now we have two stories, one real by Flaubert and the other only sketched, in which the main character was apparently deceived about his own emotions. In the latter we concluded that he was deceiving himself when he thought and said that he wanted to meet *Mme. X* again, that he still felt very strongly about her, that he was happy at the prospect of meeting her. And also of course when he thought that he was in love with the other woman, that being in her presence was an end in itself, the delight of being entranced with someone for whom he cared deeply. (I preserved certain similarities with *A Sentimental Education* to draw on your memories of something closer to reality.) So this character was deceived on the very subject to which epistemology has traditionally given him privileged access. But is *our* situation (qua philosophical spectators of the collaboration between writer and reader in the preceding section) really so simple?

It will not have escaped you that at certain points the argument was distinctly weak, and the weaknesses glossed over by bits of persuasive rhetoric. By what process did we (you, reader, and I, alternately character, narrator, and critic) reach this conclusion about the character's self-deception? It is essential that we inspect this process closely, because we wish to have it as an example of how, in reality, such conclusions are reached about others and ourselves. So let us look again, for a moment, at the preceding section.

I hope that you did not become impatient with me for tailoring the story of *Mme. X* to my didactic purposes as I went along. To be able to do that is surely the exact and only reason why a philosopher makes up his own examples rather than drawing on real literature and life? At the beginning of the *Phaedo*, Socrates (who has been writing poetry while awaiting execution) explains that poetry and philosophy are totally different because the poet makes up stories and fables. The reader immediately recalls stories of Socrates' and notices in fact that a very few lines before, broaching the nature of pleasure and pain, Socrates has sketched a fable of a double-bodied (two-backed?) beast. I was only following Socrates' practice rather than his preaching therefore when I sketched a story in the service of philosophy. My tailoring of the story was never capricious, but provided the vehicle for an argument that moves by its own internal necessity; isn't that true? And I was quite honest about it; little would have been gained anyway if I had done all the tailoring beforehand and dished up the whole story at once, at the outset. You would have seen through that anyway.

The point I wanted to lead you to was that the emotions the character felt for the woman he met on the Metro were subjectively no less and no more real than were (in the alternate plot) his suspicions and dreadful certainties

about *Mme. X*'s betrayals. They had also exactly that same efficacy, which makes the accusation of self-deception spring to our lips. In neither case can he (allow himself to) "know what he is doing"—i.e., sabotaging the possibility of reconciliation—but he is doing it nevertheless, and the facts are there before his eyes as clearly as before ours. But *it is the reader who says this*, the accuser who adds an explanation to the events depicted, who imposes an interpretation on what happens, who gives primacy (in his or her reading) to the defensive function that he or she (the reader) attributes to those feelings and thoughts. If that was not clear to you at the time, was it perhaps because I, the writer, switched so naturally from role to role, speaking at certain points even with the voice of a reader of the story I had just held up before us?

That the reader could have reached different conclusions is evident. The sudden infatuation on the Metro could be seen as a fear reaction: the way the relationship with *Mme. X* had ended is still in his mind, the memories of agony, reproach and self-reproach, the sense of loss and defeat, the mourning and grief over lost dreams. There must inevitably be fear—of failure, of renewed rejection, of a return to those tortures.

O, the mind, mind has mountains; cliffs of fall  
Frightful, sheer, no-man-fathomed. Hold them cheap  
May who ne'er hung there.

And if in fear he grasped for some safety, some harbor or firm ground, that was not a sign that he did not deeply want reconciliation. The accusing reader, in the preceding section, discarded or ignored such and similar alternative explanations, and concluded instead that this man deceived himself about wanting to make a reconciliation possible.

This accusing reader is not omniscient. He or she—more specifically, you—may have motives that do not bear close scrutiny. So let us be objective. Let us examine one another, you the reader and I the writer, with particular attention to what happened as we journeyed together through that story of *Mme. X*.

This accusing reader, who ended section 3 with "He is deceiving himself," is not omniscient. What motives lie hidden from our scrutiny? Have you, accusing reader, been betrayed by another's inconstancy; do you now have a stake in saying angrily that this man in my story was betraying and deceiving himself and everyone around him? Why were you so ready to accuse him of self-deception—implying that no, he was not really in love with this siren of the subways; no (in the alternate plot) he was not really the tortured subject of painful doubts, but only rationalizing his unwillingness to be receptive to *Mme. X*'s approach? Were you perhaps basing your verdict less on evidence than on an identification with *Mme. X*, seen (at least until late in the section) as the person wronged?

And what of me, the writer? Did I help you to this verdict by obligingly presenting, tailor-made, only those bits of evidence that fitted exactly the "real" story you constructed as a gloss on my own? Was I in fact disingenuous in *this* section too, when I outlined the reasons why a philosopher must sometimes make up his own stories rather than rely on literature or reality? (You noticed, of course, how defensively verbose that paragraph was?) A novel would not so compliantly have provided only the evidence tailor-made for one explanatory hypothesis—and reality would not have either.

Indeed, now that the doubts have sprung up, we must notice that the narrator of that story was also its main character (for the story is in the first person) as well as the writer of the section. We must surely see now how convenient it would have been for this protagonist to believe afterward that he had not really wanted a reconciliation with *Mme. X*. This conclusion, after all, allows him to discard such soul-rending possibilities as that he really wanted her terribly, and that a reconciliation would have finally led them to the love fate had reserved for them, could they only have accepted it—that his infatuation on the Metro had sprung from fear, or stupidity, and that he was unworthy of or unable to accept the great gift life had once more held out to him. And we notice sadly that the conclusion of that accusing reader, "He deceived himself" (referring not to the final diagnosis but to the infatuation begun on the Metro) gives the main character precisely that conclusion, so convenient for his later peace of mind.

But when the story purports to be a memory of one's own past, then protagonist, narrator, reader, writer, and critic are all one and the same. Where shall I stand, when the text of my life deconstructs itself at every turn of the page?

5. All these rhetorical questions may have exhausted you a little. But you saw well enough, didn't you, that they were rhetorical? That is to say: I was really presenting you with a new story, the story of how a writer and reader had collaborated in drawing the conclusion of self-deception, in a process that was a far cry from an argument moving by its own internal necessity. Indeed, this fictional writer and reader (for of course, nothing like that really happened when I wrote, and you read, section 3!) were deceiving themselves. The writer purported to be writing a story that could show us, by its verisimilitude, the phenomenon of a person self-deceived with respect to his own emotions. This scenario, once created, could be held up in place of a bit of reality, a neutral object there for inspection by everyone, and everyone would see the same thing. This purport was false; the compliant reader was deceived. Or rather, he deceived himself, because not only the writer's carefully selective presentation of data but also the reader's readiness to fit these data into a scenario of treachery were needed to arrive at that conclusion. (*You see*

what I meant about philosophy of mind done by the method of sketches of examples?)

This new example was one of self-deception in the interpretation of a story that was (perhaps, or perhaps not) about self-deception also. In thought *s.v.* every part of this could have happened, with writer, narrator, main character, and reader all one and the same person, reflecting on his own past. Recall that person, walking along the street early on in the *Mme. X* story. Totally engrossed in a self-justificatory speech *s.v.*, he plays his own lawyer, judge, jury, general public, and on top of all that, the voice and eyes of a compliantly listening *Mme. X* present at this tribunal. Perhaps he walks along another street, years later, totally engrossed, his thought *s.v.* now duplicating more or less our section 3. Now he plays, in effect, the role of writer, narrator, character, etc., in the story of his past so far. Here he finally concludes that he was deceiving himself when he came to the conclusion that his affair of the Metro was a case of self-deception about his own emotions. He concludes it was motivated by the desire to feel sure, in retrospect, that he had not really wanted the reconciliation with *Mme. X* and that therefore he had not really lost anything when the (apparent?) attempt at reconciliation was a failure.

We see therefore that the phenomena of self-deception are extremely complex, and not really touched by sophisticated little discussions about how to believe both *p* and not-*p*. The past is a kaleidoscope into which I look in memory. The hand that turns the kaleidoscope is perhaps more skillful at influencing the changing patterns than I would like to admit. Under these conditions, there is some reason for despair: not only about our ability to know where we have been and what we have done or even seen but also about the very possibility of finding significant, nontrivial truth conditions for such a statement as "He deceived himself." After all, if he did not know the truth, belief in a falsehood was not deception but error. One may object here that the deception lay in the idea that he had adequate evidence for this falsehood, while really he had better evidence for the truth. My story about stories inside stories was meant in part to preempt this reply. If *he* could not infer from his own felt emotions, actions, desires, speech, and thought *s.v.* that he was (or was not) really in love with the woman of the Metro and/or did (or did not) really want a reconciliation with *Mme. X*, then we never have sufficient reason for belief in such propositions. The truth about what evidence I have, or how it should be counted or weighed in the balance—a tortuous enough topic in the epistemically more hygienic case of scientific methodology—seems no easier to know than the conclusions themselves. If he tells me afterward that the evidence was unambiguously and overwhelmingly in favor of the conclusion that he really wanted the reconciliation but had been taken by surprise, on his way there, by a love so deep and beautiful that he was helpless in its grasp—then I doubt him just as much as if he had simply said that the conclusion was true. I just don't believe it. "He is culpable; he

was deceiving himself and everyone around him," I say, and no attempt to shift the discussion to confirmation theory can budge me. (What *can* budge me, especially if he and I are closely connected or identical, is accusations of self-deception in the way I arrived at *that* conclusion.)

In the first section I did my best to explain the difficulties with a skeptical position about self-deception. The skeptical doubts about self-knowledge we have now come to seem worse than the ones I discussed there. They also have to do with the way in which thought, belief, and emotion are partly theoretical and partly nontheoretical, of course. Beside the problems we have now about the very possibility of defining the conditions under which "He deceives himself" is true, there are also strong doubts about whether (if there are such conditions) we could ever have adequate reason for thinking they obtain. This is a problem that goes beyond the philosophical *problématique*: if I have come to the conclusion that I was deceiving myself in some respect, why should I think I won't come—in a rational fashion—to a contrary opinion after thinking some more? The skeptical doubts of *this* sort are neither based on general considerations nor general in nature. They arise if we take it for granted that self-deception makes sense and is possible, but pay attention to the whole horizon of doubts surrounding any such story. Once the possibility of self-deception is taken seriously it undermines all stories (about oneself, but also one's own stories about others), including those that attribute self-deception. We are in a quandary, personal as well as philosophical.

#### DELL'OPINIONE, REGINA DEL MONDO

6. I want to make a new beginning, and draw on another source altogether. Today, of course, the whole of Western literature occupies itself with romantic or at least sexual love, whereas once upon a time it dealt entirely with war. Compare, if you will, Homer with Barbara Pym, *The Song of Roland* with *À la Recherche du Temps Perdu*. To our writers today, it seems, love is the only battlefield. (Perhaps this is not altogether accurate; I am no expert. In philosophy facts never matter too much anyway, only the ideas; isn't that so?) However, in those battles where sword rang against sword, the heroes were all men broad and bold, not much given to introspection. Whatever truth there is in all this supports the impression that self-deception is a new and recent concept, having come to light only when writing turned from the spectacle of glancing blows to those more intimate ones of glance intercepted by glance. But—and here is my point—that impression totally ignores another literature, turned even further inward, focused on battles fought in loneliness. For once upon a time, too, not all literature was secular. To recall this quickly dispels the impression that the idea of self-deception is novel, the phenomenon only recently noticed, or the philosophical question—*how is self-deception even possible?*—new.

My knowledge of this other literature, surely as old as any, is even more limited, and I must proceed somewhat differently from before. Let me begin by citing the totally conscious, explicit description by Pascal of how we deceive ourselves:

The nature of self-love and of this human self is to love only self and consider only self. But what is it to do? It cannot prevent the object of its love from being full of faults and wretchedness: it wants to be great and sees that it is small; it wants to be happy and sees that it is wretched. . . . The predicament in which it thus finds itself arouses in it the most unjust and criminal passion that could possibly be imagined, for it conceives a deadly hatred for the truth which rebukes it and convinces it of its faults. It would like to do away with this truth, and not being able to destroy it as such, it destroys it, as best it can, in the consciousness of itself and others; that is, it takes every care to hide its faults both from itself and others. . . .

It is no doubt an evil to be full of faults, but it is still greater evil to be full of them and unwilling to recognize them, since this entails the further evil of self-delusion. (*Pensées*, fragment 100, pp. 347–348)

Pascal is as close to us as Descartes, the Port-Royal Logic, and the probability calculus. But reading Pascal we hear the echoes of St. Paul: "For if anyone thinks himself to be something, whereas he is nothing, he deceives himself" (*Letter to the Galatians* 6:3). We also hear echoes of St. Augustine examining his own blindness before conversion, and also afterward:

I did not know these things at that time, nor did I advert to them. They beat upon my eyes on every side, and yet I did not see them. (*Confessions* III–7)

Why am I hurt more by abuse cast upon myself than by that cast in my presence with the equal injustice upon another? Am I in ignorance of this also? Does it amount to this, that I deceive myself, and do not do the truth before you in my heart and on my tongue? (*Confessions* X–37, 62)

If we found ourselves inextricably entangled in the treacheries of the mind, so did Augustine: "Who can untie this most twisted and intricate mass of knots?" (*Confessions* II–10).

Let me try to offer an explanation of how religious writers engaged in such self-scrutiny arrived at our questions—what is self-deception and how is it possible at all?—and of what sort of answer they appear to give.

7. We are very safe if there really is no limit to hypotheses of self-deception, and if there is no final arbiter between them. In George Crabbe's poem "The Parting Hour," analyzed by J. Hillis Miller, the returned sailor remembers his capture by pirates, his wife and his children in Spain, and the life he led there once he had gained his freedom by marrying and converting to

Catholicism—the life there that he had fled, to return eventually, aged worn, and crippled, to Protestantism and reunion with his first love.<sup>2</sup> But the dreams and longings of his old age are exactly of this Spanish wife and of the sunshine in which he lived and loved with the vigor of his youth. We feel the shameful waste and loss of a life lived at each point in alienation from the present, oriented to future or to past but always in the desire for somewhere he is not. I mention this story to conjure up the unbearable sense of loss of someone who gave up a country, family, and peace in the arms of someone who loved him, for an uncertain future, and the unbearable regret after the choice has been made and become irreversible. But unbearable? Nothing is unbearable if you can convince yourself that the loss you regret is only the loss of an illusion. If this man could look back to those Spanish eyes and "realize" they would not have waited for him, convince himself that the memories of happiness are really of an illusion, of a blindness to the bitter lines around that Spanish mouth, of a time in which he deceived himself (or about which, in his recent memories, he indulged in self-delusion)—then there was no real loss after all. And once he tries, if he tries, he will soon have ample evidence to sustain this "realization." The hypothesis that you were deceived *can* bring consolation rather than despair; and this haven, it seemed above, is always accessible.

The religious writers to whom I have now directed our attention deny us this sad refuge of skepticism: they assert that there *are* limits, there is an end to deception. They once were lost but now are found; there is a truth to which they once were blind, but now they see. They describe graphically the state of blindness, and blame it on themselves, as a delusion for which the deluded person is responsible. And they also describe graphically the undeceived state, and what they saw once undeceived. Their philosophical problem is the paradigm problem of self-deception: how could they have been blind, or have made themselves blind, to what was there plain and clear for anyone with eyes to see? Their accounts carry an absolute, if implicit, rejection of the idea that all such states, the later as well as the earlier—any state at all of subjective certainty and conviction—are equally vulnerable to the hypothesis of self-deception.

I say the descriptions are graphic; I do not mean that they are philosophically transparent. Nor can I say that their attempts to solve the problem have an immediate or satisfying finality. Augustine writes:

In all that bitterness, which in accordance with your mercy resulted from our worldly deeds, when we sought to know the reason why we should suffer such things, darkness confronted us. Groaning, we turned away, and we said "How long shall these things last?" Often we said this, but even as we spoke, we did not give up our worldly ways. For as yet there shone forth nothing certain, which, such ways forsaken, we might reach out to and grasp. (*Confessions* VI–10, 17)

Ponticianus told us this story [of a conversion], and as he spoke, you, O Lord, turned me back upon myself. You took me from behind my own back, where I had placed myself because I did not wish to look upon myself. You stood me face to face with myself, so I might see how foul I was, how deformed and defiled, how covered with stains and sores. I looked, and I was filled with horror, but there was no place for me to flee to away from myself. If I tried to turn my gaze from myself, he still went on with the story that he was telling, and once again you placed me in front of myself, and thrust me before my own eyes, so that I might find out my iniquity and hate it. I knew what it was, but I pretended not to; I refused to look at it, and put it out of my memory. (*Confessions* VIII-7, 16)

[T]he day had come when I stood stripped naked before myself. (*Confessions* VIII-7, 18)

At crucial points the description of the experience becomes metaphorical, but the point is clear: there was a truth about himself that he did not see before and does see now. When he did not see it he was mistaken about himself, and this mistake was no mere mistake but was due to a refusal "to look" and to willful, wish-fulfilling acceptance of a rosier picture maintained against and in the face of the facts he did not face.

8. It is also characteristic of these writers that what they see in their past is a situation so desperate that the despair would be unbearable if it *were* faced (alone). Accordingly they hypothesize mechanisms of self-deception in analogy to avoidance, flight, and escape characteristic of other, more familiar failures of courage. The mechanisms must, however, be means of mental rather than physical flight, and must be drawn from familiar types of mental activity. Their hypotheses relate accordingly to distortions of imagination, opinion, and desire or value. This is very clear in Pascal, who may be regarded after all as a pioneer of decision theory. Let me quote from him first to establish the view, similar to Augustine's, of the real truth about the human situation, and then to show the ways, less metaphorical than Augustine's, in which he suggests explanations. First, the undeceived view:

When I see the blind and wretched state of man, when I survey the whole universe in its dumbness and man left to himself with no light, as though lost in this corner of the universe . . . I am moved to terror, like a man transported in his sleep to some terrifying desert island, who wakes up quite lost and with no means to escape. Then I marvel that so wretched a state does not drive people to despair. (*Pensées*, fragment 198, p. 88)

One needs no great sublimity of soul to realize that in this life there is no true and solid satisfaction, that all our pleasures are mere vanity, that our afflictions are infinite, and finally that death which threatens us at every moment must in

a few years infallibly face us. . . . Nothing could be more real, or more dreadful than that. (*Pensées*, fragment 427, p. 157)

Next we find him wondering how it is *possible* for people to be in this situation, and yet not in despair. He is inevitably led to suspect deception and pretense:

Thus the fact that there exist men who are indifferent to the loss of their being and the peril of an eternity of wretchedness is against nature. With everything else they are quite different; they fear the most trifling things, foresee and feel them; and the same man who spends so many days in fury and despair at losing some office or at some imaginary affront to his honour is the very one who knows that he is going to lose everything through death but feels neither anxiety nor emotion. It is a monstrous thing to see one and the same heart at once so sensitive to minor things and so strangely insensitive to the greatest. It is an incomprehensible spell, an supernatural torpor. . . . Man's nature must have undergone a strange reversal for him to glory in being in a state in which it seems incredible that any single person should be. Yet experience has shown me so many like this that it would be surprising if we did not know that most of those concerned in this are pretending and are not really what they seem. (*Pensées*, fragment 427, p. 159)

He echoes here St. Augustine's "Whence comes this monstrous state?" (*Confessions* VIII-9, 21). Elsewhere (in the passage from *Pensées*, fragment 198, which I quoted above) he says, "Then these lost and wretched creatures look around and find some attractive objects to which they become addicted and attached." The first hypothesis is therefore of a distortion of desire or value judgments, which results in a disproportionately high valuation of something trifling. This distortion of judgment is not possible without a certain blindness or myopia as well, i.e., without ignoring much of our total situation and focusing on what is near and obvious alone. He gives a simpler, more familiar sort of example, designed perhaps to show how distortion of value judgments and of perception (in some wide sense) cannot be disentangled: "How is it that this man so distressed at the death of his wife and his only son, deeply worried by some great feud, is not gloomy at the moment and is seen to be so free from all these painful and disturbing thoughts? There is no cause for surprise: he has just had a ball served to him and he must return it to his opponent" (*Pensées*, fragment 522, p. 214).

Nor is it possible to disentangle imagination and opinion; they are distorted and distorting, in unison. In a long and searing commentary on the disguises sustained by imagination, Pascal writes scornfully: "Imagination decides everything: it creates beauty, justice and happiness, which is the world's supreme good. I should dearly like to see the Italian book, of which I know only the title, worth many books in itself, *Dell'opinione regina del mondo*. Without knowing the book, I support its views . . ." (*Pensées*, fragment 44,

p. 41). Imagination, he says here, is the dominant faculty, "master of error and falsehood, all the more deceptive for not being invariably so."

9. Let us try to sum up, in a preliminary way, how these writers view the phenomena. *First*, as I said, they identify a certain appreciation of our situation in the world as the correct one, the truth, not vulnerable to further charges of self-deception. This *appreciation* (a term I shall use to mean both perception and evaluation) is one they have come to themselves, and forms the point of view from which they evaluate the others (including that of their earlier selves). *Second*, it is possible for anyone to see the situation this way, and it is really surprising that people don't see things as they really are, for "they beat upon [their] eyes on every side." *Third*, they agree also that it is actually very difficult to see the truth—they do not assume that our minds or personal histories are transparent to us. The surprise must be about the apparent difficulty: how is it possible not to see the truth, and how can it be that it is in fact so rarely faced? The surprise is also with the contrast: how can a person attend so closely to (comparatively) small issues and be oblivious of the large ones? The *normal, usual* state is one of ignorance and deception about oneself.

But this must also mean that normal, rational functioning can go on well enough in a state of self-deception. When Pascal examines the work of the imagination, he does not uncover pathological functions: as he describes it, imagination has a role in the normal, rational processes of formation of opinions and decisions.

I think the first point is a very important one. Not only is it the stopping point that keeps these authors from sinking into the quandary that caught us in the first part of this essay ("Frederic Moreau and Mme. Arnoux"), but perhaps there is a secular, "relativized" version that can help us in our quandary too. This secular version would be: the attribution of self-deception makes sense only from the point of view (one appreciation of the facts) of someone who does not share it. That entails, correctly, that I could never rationally and truly say, "I am now self-deceived about so and so in such and such a particular way." It does allow, also correctly, that I could say that I have been so self-deceived, or even that I think I must now be deceiving myself about something, though I don't know what. (The "preface" paradox notwithstanding, I can say that one of my present beliefs must be false, though I don't know which. It is also possible to say that my present appreciation of the facts must be mistaken, although I can't yet see how, and that from the correct appreciation it will appear that I was now guilty of having persuaded myself in some culpable fashion.)

To insist on this will not remove the doubt whether my present conclusions about myself are ever invulnerable from charges of self-deception. But it does mean that when my appreciation of the situation is consistent (even in the broad sense of being free from glaring inequity in my treatment of different

subjects), I shall have a certain measure of safety. For in such a case I can be sure that the charge of self-deception cannot be warranted by appeal to features of my present appreciation, but only by a point of view that I do not (yet) share. We glimpse here the possibility of a position that does not "deconstruct itself."

Let me elaborate a little. The worst fear is that I would always be in a position in which I can be convicted *on my own grounds* that there is serious reason to doubt my conclusions about myself. We have removed this, leaving only two other fears that are simply not theoretically removable (and we must live with them). The first is the fear that if I examine my own appreciation more closely, I will find inconsistencies or glaring inequities. The second is that I will at some later time encounter in someone else, or even in myself, an appreciation according to which I was not simply mistaken but culpably mistaken in my present one. These two fears the religious writers also overcame, but admittedly not by reason alone.

10. I omitted from the list in section 9 the important fourth point in the story told by these religious writers. That is that the truth, the real state of fallen humanity, is unbearable. This is also the beginning of their answer to the puzzle: how is it possible to be so deceived? For if the truth is unbearable, it is not possible to live while facing it. And here they can link up with more mundane cases again, for it is appropriate to look at less extreme ones. Such is Pascal's widower, so distressed at the death of his wife and son, yet at present not gloomy at all. "There is no cause for surprise: he has just had a ball served to him and he must return it to his opponent." This person is not unaware of the facts, but he is not currently facing them. We can quickly marshal other cases: escape into daydream, or recourse to magic and superstition, or a headlong plunge into work, career, infatuation, parties . . . any of the many effective creations of illusion that "make life worthwhile again."

What is needed is an account of normal, rational functioning of opinion, desire, and decision which gives substance to this view. For it is at first sight really startling: it says that self-deception proceeds by exactly the same means as rational changes of mind, mood, and behavior.

In Pascal's *Pensées*, not solely in its famous wager, we find the beginnings of decision theory. Let me sketch this sort of analysis of one's situation, as we might today. A person faced with a decision has first of all an *opinion* about the current state of nature: what the relevant facts are like. Second, he *imagines* a set of alternative actions. Unfortunately, his imagination is limited, so some possible actions open to him may fail to be included. The opinion above may also have such limits: he imagines various ways in which the world may be, and these may not exhaust all possibilities, logically speaking. Thus imagination enters in two ways. Let  $S_1, \dots, S_n$  be the states of nature he imagines as possible; his opinion consists in judgments of the form " $S_k$  is

$Q(k, m)$  as likely to be actual as  $S_m$ "; the number  $Q(k, m)$  is called his *odds* for  $S_k$  against  $S_m$ . In addition, let the set of actions he imagines as possible be  $A_1, \dots, A_q$ . His imagination steps in to supply possible outcomes of these actions; let these be  $C_1, \dots, C_r$ . His opinion then contains further judgments of form "If the actual state of nature is  $S_n$  and I do action  $A_m$ , then  $C_n$  is  $Q(k, m, n, p)$  as likely to occur as  $C_p$ ", thus entering further odds. Now, why should this situation be unbearable to him?

Well, one of the actions imagined as possible is to kill his mother, and one of the outcomes imagined may be to be guilty of not avenging his father and, moreover, to be torn apart by the Furies. Such was Orestes' situation. In addition, the factual judgments constituting his opinion entailed that he could not avoid the latter outcome without performing matricide. Given his values, moral judgments, and desires, his love for mother and father alike, action and outcome are unbearable to contemplate. Besides imagination and belief or opinion, therefore, there is a third ingredient, or sort of ingredient: value, moral sense, desire.

Fallen humanity's situation is unbearable in a different, if related, way. The state of nature includes much about him, namely, his own character and past actions: everything that is settled so far. He is not entirely definite about which is the actual state, though some seem likelier than others. In addition, he evaluates some as despicable or regrettable. In addition also, looking at his possible actions and their outcomes, and hence, at the accessible future states of affairs, he sees again much that is despicable, regrettable, and frightful. But even more, the limits of these three imagined sets, *sub specie* his opinion, are such that he has "nowhere to turn": he sees no escape from the regrettable, despicable, and frightful. Hence his situation is unbearable.

That is, it is unbearable as long as he or she is objective. He seeks frantically for an escape from the inescapable, within his own power, and objectivity is lost. He *does* find a way to escape, a way fully allowed by the above abstract theory. The situation in which he finds himself is unbearable, after all, in virtue of the collaboration of three ingredients. It is unbearable in part because transformations of the situation by means of an action are not envisaged as leading to one more bearable. So "normal" escape is not possible. But the situation could be transformed in a way other than through action: namely, through changing the imagined sets of possibilities, or the opinions concerning them, or the values and desires that define their significance for him. Of course, one cannot do most of this while knowing what one is doing. For in the light of one's present opinions and values, different opinions and values are mistaken. Hence from one's present point of view, the result of the transformation would be a delusion or a perversion.

It will be objected at once that such an escape would be pathological and easily distinguished from escape through action. Certainly, many such "sub-

jective" transformations of the situation must have recourse to the irrational. Yet the objection does not stand. For each such subjective change can be, and under certain circumstances will be, the rational and correct response. If we postulated moreover that no one is ever in a problem situation without a rational solution, it would even *follow* that "normal" escape through action is impossible only if some such subjective transformation is rationally required.

Let us consider examples. Sometimes one is trapped by limited imagination: there *are* ways out, but the agent was not imaginative enough (and perhaps no one so far had ever been!) to see them. Salvation may come through the imaginative construction of a richer world-picture, a richer set of possible states of nature or of actions. But from the preceding point of view, two classifications of this escape are possible: alchemy or science. Think of a secular reaction to an imagination that calls up a paradise in the hereafter or atonement through sacrifice, or anyone's reaction to the purchase of a lottery ticket. The imagination has suddenly hit on something else to be brought into the picture. And the lottery ticket may win a million dollars. The foolish dabbings with wooden wings and smelly little engines may lead to the first powered flight, followed by the era of true aviation. The escape mechanism is the same.

Of course, it is not enough for imagination to provide a richer set of alternatives, if the new ones are allocated no credence. All the neighbors laughed when the poor man bought the lottery ticket, or those dabblers invested their energy and resources in heavier-than-air flying machines. The agents themselves must have found it a little less incredible—though in an unbearable situation, even a little credence may suffice for such a wager. The transformation of opinion by something other than minimal adjustment to the evidence must be *per se* within the bounds of rationality, if my current argument is to succeed. I think it is, and have argued for it elsewhere.<sup>3</sup> But it is easy to see that self-deception can also result in this way—as well as foolish hopes and superstitions of all sorts.

The most powerful transformation of all is the emotional, which effects a transvaluation of all values. The situation is unbearable because all outcomes of all envisaged actions are just different forms of misery. But why are they misery? Because they satisfy no desire, and harbor threats of more deprivation later. But which desires, what satisfactions? Suppose that as the situation is first defined, only one sort of outcome is acceptable: a reconciliation with *Mme. X*, a return to the love and trust and affection of the past. When it seems no avenue is open that leads to this, suspicion, blame, and anger may spell sudden salvation. For suddenly there is a new *sort* of alternative (or a new cross-classification of the same outcomes as envisaged before), which gives satisfaction: e.g., the alternatives in which the protagonist rightly and

strongly rejects her, and bitterly lets her know how she deserves rejection. The very same action that before was ranked with cutting off one's own hands, now (in anger) ranks as the way to satisfaction.

When is this transformation through emotion to be lauded as rational and when to be depreciated as irrational defense or self-deception? We could insist on truth as a criterion: the situation has been correctly changed exactly if, as a matter of fact, *Mme. X* deserves rejection. It is rational to revalue the options by anger exactly if the resulting valuation is objectively correct. Quite apart from the difficulty of bringing factual truth and objective value together, there is also this problem: the criteria of rationality should be such that self-policing is in principle possible. Irrationality cannot be due to ignorance. Self-deception cannot be a matter of factual accident, outside the agent's ken. The religious view has no difficulty here: the real facts, the real values may stare one in the face—and the criterion of evident truth, admitted or denied, can replace that of truth. Thus abstractly stated, this view can also be secular. But there will be many gaps—many cases in which the truth is not evident. And there is the general problem of distinguishing evident truth from great but wrong certainty, certainty that is *especially great while one is in the grip of emotion*. From the point of view held preceding the anger, pity, tenderness, fear, or other emotion, the subsequent valuation is of course wrong, and from the one that accompanies the emotion, right. If criteria of rationality must allow of self-policing, they cannot distinguish in such a case.

We are driven again to the conclusion that the distinction between the case of a "correct" change in the emotive perception of the situation and that of a self-deceiving, rationalizing, or self-justifying change, must come from outside. Such an admonition, as, "But don't you see how she is suffering? Don't harden your heart!" or its contrary, "But don't you see how she has been leading you up the garden path again? Take hold of yourself!" reflects a definite emotive and evaluative perception of the situation. However, only giving in to the emotion of tenderness or pity on the one hand, or of anger on the other hand, can effect the admonished change. For the protagonist has to come to see *Mme. X* as deserving sympathy, or rejection, respectively—and this change in emotive perception *is* emotion. It is not a result of emotion that could equally be achieved by deliberation alone; it *is* emotion.

#### ON FAITH IN THINGS UNSEEN

11. *We have now an answer, albeit a little sketchy, to the first philosophical question: how is self-deception possible?* I am not unhappy with the answer; indeed I am quite convinced that the normal functions involved in rational opinion and decision clearly allow for this possibility. This answer presupposes a nonparadoxical identification of the phenomenon, something like: to

be of a false opinion, and to maintain and support this opinion in a fashion that makes one culpable of wrong-doing (such as biased selection, presentation, or weighing of evidence) of the sort characteristically involved in deception. And such an identification in turn presupposes an exterior viewpoint: the evaluation of someone as self-deceived in some way must be from within an appreciation of the situation that differs from his or her own (though it could be one's own somewhat later).

As we discussed the phenomenon of self-deception, however, the "secondary" problem of skeptical doubt became more and more pressing. Are there no limits to self-deception? Does the very possibility of self-deception not automatically rule out the possibility of self-knowledge? Well, perhaps knowledge is too theoretical or idealized a concept to worry us. But the problem still remains. For if self-deception is always possible, consider the plight of the poor subject who has begun to wonder if he has not perhaps come to some opinion by deceiving himself. He begins to investigate more closely, to weigh the evidence, to imagine new hypotheses of sorts not hitherto entertained. Unfortunately, it is exactly in these processes that self-deception always finds its means. For one selects, but not explicitly or consciously, what aspects to attend to more closely; the weights to give to different bits of evidence are not "written into" that evidence but must be supplied by oneself; and the imagination—ah, the imagination! So active in directions pleasing to oneself, or alternatively frightening, servant of wishes and fears, itself the flawed mother of the flawed queen of the world! Under these conditions, how shall we ever find a spot of safety, secure from distrust of one's own opinion?

The religious writers, as I pointed out in the second part, assume or recognize such a secure vision of the world, the single truth against which all other views are measured. But it is admittedly not reached by reason; the certainty found is one of illumination. I accommodated this momentarily by noting the important role, played in attributions of self-deception, of an exterior viewpoint, different from that of the accused. This does not solve the skeptical problem; at best it underlines it. Yet we can't very well offer faith as a solution—faith, the sum of our hopes and evidence of things unseen, is not a *philosophical* solution to skepticism. And apart from that, the religious writer surely has his own skeptical problem. Should an angel appear and speak to us, what credentials establish him as an angel? And if words spoken with the voice of an angel strike certainty in the heart, so to speak, who is to say whether his own certainty is so come by? For, despite Descartes' proofs, we have no a priori certainty of the sources of our certainties.

12. Faith, I said, is not a *philosophical* solution to the skeptical problem. But what is a philosophical solution to a philosophical problem? What we can hope for is a wider perspective, in which the initial double bind appears

to be due to demands or presuppositions that we need not share. And that sort of solution, I do think, is to be found in Augustine. I am now referring not to his *Confessions* but to the first work he wrote after conversion, the work called *Against the Academics*.

13. To describe Augustine's response to the Academic philosophy—the Pyrrhonist skepticism of the Third Academy, handed on through Sextus Empiricus and Cicero, and popular in the late Roman empire—I should first describe the skeptical problem as he found it. I shall begin with the story of the Stoic "criterion" and the critique it received in the Academy, then state the problem faced by those critics.<sup>4</sup> At first sight, Augustine's reaction seems naive and inadequate; but I think that in the final analysis, it proposes a form of voluntarist epistemology that deserves to be taken seriously.

How shall we distinguish the true from the false? This problem arises acutely already for simple judgments of perception: "That is Helen standing on the shore," "There is a snake in that corner." To say that sense perception is the final arbiter in questions of what is true does not solve the problem—it only pushes it back, for we are often mistaken in just such simple judgments. Chrysippus provided the definitive version of the Stoic solution. He draws a distinction between sensation and judgment. The former is a change in the soul, produced in a way that is not subject to the will. A judgment is something we form, in general—through by no means always—deliberately. This is where error may arise, and the function of the will in Chrysippus' account is to explain perceptual error. Some judgments, however, are forced by sensation—these are called *irresistible perceptions*, and that irresistibility is the mark of truth. The criterion of truth lies exactly in those judgments that are not formed deliberately, in which the will plays no part, but which are forced irresistibly upon us.

There is good prima facie reason to accept the account. Is there any way that I could, at this moment, believe, deliberately form the judgment, that there is an elephant sitting on the table before me? No. Certain simple perceptual judgments are forced upon me by my present sensory experience, and I *cannot* believe something with which they conflict. This does not imply that belief is not subject to the will. Larceny presumably is; and if I am sitting in a warm bath in my own bathtub, in my own home, I cannot steal. The preconditions are not met. Our present experience may be relevant to belief in just that way. However, Chrysippus intends something more: certain judgments are irresistible, made so by sensation. *Seeing is believing*. And finally, most important, irresistibility is the mark of truth.

The New Academy, under Arcesilas and later Carneades, attacked the Stoic doctrine in a number of ways. Because the conclusion for which they argued was that belief is never rationally compelled, and is really better not to have, they do not seem to have considered the question whether, during given

experiences, some beliefs may be impossible simply because some preconditions of belief are not met. Certainly they comment indirectly, because they insist that, given any belief, they can find reasons to show that a belief in the opposite is at least as rational. Doing so produces "equilibrium" in which agnosticism recommends itself. So apparently their contrary view is here that *the feeling of irresistibility comes not so much from sensation by itself, as from sensation in the context of a limited range of contemplated possibilities*. The irresistibility disappears when you think deeply enough about all the possible ways in which the judgment could be false.

Let us put this in some perspective. The Stoics, noting the irresistibility of certain judgments at certain times, attribute this to sensations. But it is clear that if I were in a different state, or had had different previous experience, the same sensations could force different judgments. After talking and thinking about snakes all evening, I "see" one on the veranda the moment I open the door. Of course, at the moment of judgment, it is not within my power to change the preceding evening. But which alternatives I presently contemplate it is within my power to determine. By considering various new possibilities and inventing new hypotheses, judgments that were at first accompanied by that feeling of irresistibility may lose it. And when they do, the preconditions for forming opposed judgments may be met. Having seen this, however, the idea that such irresistibility is the mark of truth loses its charm. False judgments are at times accompanied by such feelings every bit as strong as those that marked the true. The rationale behind the skeptics' attack is that which judgments we arrive at in experience depends in part on the contemplated hypotheses and possibilities. Any sort of optical illusion will bear this out, but ropes seen as snakes will do as well.<sup>5</sup>

Nevertheless, Arcesilas is left with a serious problem. It seems that, in immediate experience, seeing *is* believing. If you see a horse rushing toward you at full gallop, you jump aside. What better evidence that the sensation *produced* belief? If we lived in epistemic equilibrium, we would not form those impetuous judgments of perception, but we would die. So not only do we apparently rely on our ability to arrive at true beliefs, but we ought to rely on it—suspension of judgment would automatically lead to suspension of action. Thus Arcesilas ends up with the problem of giving an account of practical action. It is clear how he would react to the argument—by inference from other skeptical reactions to other points. For the skeptics readily admit that they have the same feelings of conviction as other men do, and that they naturally act on them. They deny only that in doing so, they implicitly disown skeptical doubt. For even if the feelings are momentarily irresistible, this need not lead one to *endorse* the judgment. There is no reason to say that seeing *is* believing, when the only evidence we have is that we act just as we would if we did believe, immediately, what we saw. The alternative hypothesis, which produces skeptical equilibrium here, is that we rely not on our ability

to arrive quickly at true perceptual beliefs, but instead on an ability to correct our impetuous mistakes as quickly.

So endorsement, not action, is the criterion of belief for Arcesilas. The problem remains acute, however, for if we do not act on our beliefs, what do we act on? If belief is not the guide to life, what is? Carneades, the head of the Third Academy, developed the constructive alternative answer to these questions. He explicitly developed an account of *probability as the criterion for the conduct of life*. (See Sextus Empiricus, *Against the Logicians* I, pp. 166–173: “These were the arguments that Carneades set out in detail . . . to prove the non-existence of the criterion (of truth); yet as he, too, himself requires a criterion for the conduct of life and for the attainment of happiness, he is practically compelled on his own account to frame a theory about it.” The basic idea is that what appears to us in experience gives greater or less probability to various possible judgments that occur to us (the irresistible appearance is replaced by the probable appearance at the bottom of the epistemic ladder), and we act on these probabilities—endorsement never being involved in the step from seeing to acting.

This one-sentence summary does not do justice to his account, for he did not merely degrade the epistemic status of irresistible perceptions from indubitable truth to probability. In addition he took away their criterial role; for the immediate judgments have, for him, the lowest level of probability, and are relied on least. It is tempting to compare his view with such a view as Richard Jeffrey’s that the epistemic subject does not give probability 1 to any contingent proposition, and acts merely to maximize expected utility. (Jeffrey himself has done so.) What Carneades did have was a rough division into levels of probability, or perhaps of evidential support, and the recognition that within each of these levels, distinctions of more-and-less are possible. In any case, as a consistent skeptic, Carneades must have proposed this merely as a contrary hypothesis, to be used in the method of equilibrium, against the Stoic position.

In that alternative account, Carneades envisages us as giving least credence to the judgments made spontaneously in response to experience. We immediately subject them to the test of corroboration. Someone smiles across a crowded room; I rise a little hesitantly and say, “Isn’t that Peter?” When he comes near, I just extend my hand and say, “Why, hello, Peter!” But later even well-corroborated judgments can be submitted to further inquiry “as in a court of law”—for “no presentation is ever simple in form but, like links in a chain, one hangs from another” (op. cit., pp. 176, 182). One ends by giving assent (*syncatathesis*) to judgments that survive an active inquiry, similar to the practice “at assembly-meetings when the People makes inquiry about each of those who desire to be magistrates” (loc. cit.). But it is clear that this assent—so described by Sextus—is not the endorsement that Ar-

cesilas says we should withhold; it is merely a practical assent, an acknowledgment of willingness to act.

14. To understand Augustine, I think, we should be subversive and ask whether belief is possible for the skeptic. It is clear enough that Arcesilas and Carneades thought that people do reach full belief—but are foolish to do so. Are these negative evaluations of belief compelled by their account of endorsement and practical assent, or merely concomitant?

When St. Augustine was converted to Christianity, he immediately wrote *Against the Academics*, attacking the skeptical philosophy of the New Academy. What is most noteworthy, but little noted, is how much of the skeptical philosophy Augustine accepts as correct. To be sure, he argues against it that we can have certainty, but of eternal truths (he gives the arithmetic example that  $2x = x + x$  for all  $x$ ) and egocentric ones (I am alive, I exist, I think). For perception and the natural world, he agrees to all the skeptical arguments. There, no belief is rationally compelling, nor any evidence. But he adds that we do have beliefs about the empirical world, takes it for granted that this is rational, and concludes that rational belief is shot through and through with leaps of faith. Is this really a consistent position?

To examine this question, we must look into the Academics’ reasons for counting full belief irrational. A moderate skepticism can be summarized in four tenets:

- (1) seeing is not believing: experience does not force belief (in the sense of endorsement);
- (2) nothing gives any empirical proposition the automatic right to acceptance, or would compel all rational beings to believe it;
- (3) nothing gives any accepted proposition the automatic right to continued acceptance, or immunity from revision;
- (4) rational action and decision can be based on the probable, while judgment as to truth is suspended.

There is nothing in these four tenets that implies that having beliefs is irrational. It is implied that if you are concerned only to have your behavior and deliberation qualify as rational, you need never have a full belief. But neither need you ever reach an aesthetic judgment or have a capacity for aesthetic appreciation—to give only one other sort of example—so this does not suffice to make belief irrational.

The fifth, hidden tenet that characterized the skeptics and led them to that conclusion was a certain view of life, happiness, and wisdom. For note that (1) through (4) also establish that no empirical proposition is absolutely

beyond doubt. Hence if you form a full belief, then you court the danger of being wrong, of being contradicted by further appearances. This is a perturbing experience, all too common in ordinary lives, and the basis of all unhappiness—what is it to be unhappy but to be deceived in one's expectations? Wisdom consists, not only for the skeptics but for their rivals in antiquity, in guarding against the very danger of unhappiness, in the ability to lead a happy life, that is, a life free from perturbation, whether by actual *contretemps* or the apprehension thereof. Needless to say, Augustine did not share this view of life—it may not be unfair to say that this disagreement is the exact mark of the passing of the Pagan world, the coming of the Christian era. Yet this was the reasoning that led to the further tacit principle:

- (5) if something is not beyond reasonable doubt, then it is irrational to believe it.

So we may conclude at least that it is consistent for Augustine to accept the main skeptical arguments against empirical beliefs while holding that one can rationally have such beliefs. We will require from him, however, two things: reasons for having (empirical) beliefs *at all*, and an account of the epistemic process that makes the role of belief explicit. As to the first, I have found no direct answer in his writings, and I do not feel that I can rigorously extrapolate one. But Augustine does give quite an explicit answer to the second, and provides indeed an account of belief that is as definite as Carneades' account of practical assent. I may sum it up in four points.

(1) The will plays a crucial role in perception (*Trinity* XI, 2, and *De Musica* VI, 5, as well as sundry passages in *Against the Academics* and *On the Freedom of the Will*). In his analysis, perception has three distinguishable aspects: the content (object seen), the experience or act (the seeing), and the attention, which is the part played by the will. In other words, sensation is, as the Greeks said, a change in the soul—but there is no perception of the object sensed unless attention is focused on it, so as to make the information available information. Hence Augustine says that the action of the will “binds together” the act and the content. All this is quite metaphorical or analogical (which is typical of the way we describe mental phenomena), but I think that the point is a good one: deliberate selection, to a large extent under our control even if *usually* automatic, is an essential part of perceptual experience.

(2) The criterion of belief is endorsement or explicit assent, and this is subject to the will: “for what else is it to believe but to assent to the truth of what is propounded? Consent being a matter of the will” (*On the Spirit and the Letter*, 54). Not unexpectedly he is quite clear that to say this does not mean that consent is necessarily easy, or that one believes the moment one wants to believe—these being points that theologians typically discuss at length.

(3) Reason(ing) is subject to criteria of evaluation closely linked with, or analogous to, those of moral judgment. This point is one that emerges already in the skeptics' discussions, and especially in the description of the Carneadean threefold process of putting into question, resolving doubts, and giving assent. In this process of inquiry and decision one can be timid, prudent, courageous, impulsive, impetuous, steadfast, stubborn, or swayed by every breeze. Augustine goes further, for he argues that the very process presupposes some measure of traditional moral virtues:

Reason is the gaze of the soul, but . . . the gaze itself cannot turn [the soul] toward the light unless these three things endure, namely: *Faith*, by which it believes that the thing on which the gaze is to be fixed is of such a nature that when it is seen it will beget happiness; *Hope*, by which it trusts that it will see, if only it gazes intently; *Love*, by which it yearns to see and enjoy. (*Soliloquies* I, 13)

We have here an implicit criticism of Carneades' account; for why should one engage in scrutiny and inquiry? Why worry about strengthening the evidential support for propositions before acting on them, unless one believes that this will be beneficial? We may put it paradoxically, but pointedly, as follows: *if the only connection between probability and rationality is that it is a principle of rationality to act on probabilities, then there is no significant connection.* A principle without a point has no status.

(4) All belief is shot through and through with leaps of faith. This view, amply illustrated in his *On Belief in Things Unseen*, is clearly a necessary consequence of Augustine's acceptance of the main skeptical arguments. But since he simply takes for granted that much of what ordinarily passes for rational belief is rational, he can also conclude that these leaps do not make the belief irrational.

Yet the leap of faith is the paradigm of going beyond where rationality leaves off. To reconcile these two points, we must clearly distinguish *rational* from *justified*, and *justified* from *compelled*. Consider an argument in which these are not distinguished: “justification of belief is through evidence; but of course you cannot be justified by evidence in believing more than what that evidence justifies; therefore, it is irrational to go beyond the evidence.” I think everyone will recognize this argument as a fallacy, and locate the crucial ambiguity. The evidence may certainly give us good reason to believe a hypothesis even though it does not rule out all contrary hypotheses. Just as in practical decisions, the most prudent decision is not necessarily the moral one to make, so also in epistemic decisions, the most prudent policy of belief formation may not be the wisest. Rationality is at most bridled irrationality, and the process of rational inquiry does involve venture, commitment, enterprise, willingness to take a chance. If our belief, and belief change, is by and large rational, then rationality *is* shot through and through with leaps of faith.

15. So we are finally left with the conclusion that disagreement with the skeptics is in the view of life. To disarm the skeptics' final move, one must say that, on one's own view of life, genuine epistemic *engagement* has its own value, and is to be preferred to a life of utilitarian calculation and prudence—what Bradley called a shopkeeper's life of always a little bit more, a little bit less. There is no escape from skepticism in theory; skepticism has no theoretical limits. This is a conclusion I also asserted in section 2. Reasonable skepticism about what anyone really thought or felt has its limits—but they are not theoretical limits. At some point, reactions to skepticism, whether refusing or giving in to its seductive offers of momentary peace and safety, equally become matters of decision, attitudes, and will. Self-deception for example, is possible, and so certainty about what one felt or really thought, or about how one arrived at one's opinions, may in any given case be an illusion. But this is finally only one instance, if the most extreme, the closest to home, of that danger which lurks everywhere. There is no need to counsel us to live dangerously; we do. The question is how to live with that danger—seeking to remain in the safest position possible, the skeptic's solution, is no more *theoretically* justifiable than any other. Of course, the converse of skepticism is faith—to give in to skepticism is to refuse faith, to refuse skepticism is a leap of faith. This is as true in personal matters (Is it love? Does he trust me? Do I really feel that, or am I pretending even to myself?) as in ideology or religion.<sup>6</sup>

#### COURAGE AND THE END OF PHILOSOPHY

16. Thinking about self-deception we felt ourselves powerless, not against global skepticism—which allows life to go on as normal, except for philosophers—but very specific, disabling, local skeptical doubts. Telling and recognizing the truth about oneself, about one's own opinions, emotions, and attitudes, appeared insuperably difficult. And that takes the thrill out of every joy—for joy lives on unreflective certainty—and increases the sting in every guilt and anguish. Unfortunately for the philosopher, local, even little skeptical doubts soon threaten to swallow the whole world. But perhaps that is not so unfortunate, if in refutations of skepticism we can find defense against the local sorts as well.

There are two sorts of refutation of skepticism—I think we should call them the idealist and the voluntarist. The former supply a defense against global skepticism only, for they purport to demonstrate that the doubt is incoherent, and the demonstration hinges on the global nature of the doubt. Such are, in our day, the refutations offered by Donald Davidson and Hilary Putnam. The voluntarist, of which I have followed Augustine's version in some detail now, characterizes capitulation to skepticism as a failure of will,

of courage, of faith. G. E. Moore's, which looks at first sight so different—"look, here is one hand, and here is another"—is also of this type. How can we resist the slide into doubt, except by reaffirming faith? And faith lives in every moment of everyone's epistemic life. I can *see* my hand, but I can only *believe* in it as the means for lifting the cup, breaking the bread, touching a hand. The belief is verifiable; indeed, the very next moment may put it to the test, so to affirm this faith is not the empty posturing of the metaphysical world-maker. But the point is easily taken too far, as by Jean-Paul Sartre's communist Brunet in the prison camp in *Iron in the Soul*:

"Schneider" he said without raising his voice. "It is *conceivable* that . . . the roof of this hut might fall on your head, but that does not mean you spend your time keeping a wary eye on the ceiling. . . . When you stretch out your hand to grasp your mess-tin, the mere fact of that gesture postulates a universal determinism. (p. 338)

No, it does not. Our opinions need only be very modest to sustain practical life. But, if the voluntarist story is correct, Brunet is pointing to a clear and inextricable feature—though of course the word "postulate" comes more naturally to him than "faith." Either way, Brunet's remark points to the role of the will, and thus leads to the diagnosis of disabling doubts as a failure of will.

17. Whatever happened to Frederic Moreau? He grew old, settled down to a life of tired idleness and reminiscences with his old friend Deslauriers. As they exhumed their youth, they asked each other after every sentence, "Do you remember?" In their memories, real happiness was reserved for scenes in boyhood, when they were just becoming young men, their derring-do when every act was still more symbolic than real, surrounded by a halo of limitless possibilities. Frederic's life had after all been only a sort of life, the lost dream of a love lived with Mme. Arnoux always putting the meager realities to shame. So now those realities are discounted: no, life had never lived up to those glories of the imagination, for which courage and fate had proved insufficient.

And what happened to our narrator, whose courage had failed crucially at every step with *Mme. X*? Here we enter the garden of forking paths; many things happened.

His love of Flaubert led him to take up the mannerisms of the Second Empire: a cane, a hat, a daybook to record anecdotes, memories, and fantasies. Late one afternoon we see him writing in his daybook: I remember still that day I walked along the street, my mind a court in which I fought furiously to answer *Mme. X*'s charge that I no longer loved her. Again I hear her voice, gently telling me that I merely deceived myself if I thought that I

still loved her. I hear myself answering with many passionate fallacies, before a tribunal scandalously biased in my favor. Then, years later, when she wrote that she wished to see me again, I continued this self-deception when I replied—surely believing what I wrote—that I had never forgotten her, that all loves had been insipid beside the memory of hers. But then . . . what a disaster. How could I have kept myself so much in the dark about what I really felt and wanted? To think that I escaped blindly into an infatuation with someone else—just to avoid admitting that I really had been living a lie all along.

Or was *that* story, about how I deceived myself about being all that time still in love with *Mme. X*, the real self-deception? Was it simply unbearable for me, after all, to think afterward that I had lost her love just when fate finally offered it again, by not being courageous enough? Why did I not, when she came back, affirm my faith, commit myself utterly, however dangerous it seemed to place my life again in *Mme. X*'s none too reliable hands? Ah, well, who is to say? In the end, age having taken its toll, I prefer to think back to days long before those. I remember the time when I was still a boy but almost a man, when every action seemed charged with significance, when every step in the rustling autumn leaves carried me forward into life and its limitless possibilities. And then those days a little later, my first days in Paris, lamplight reflected in the wet November leaves, my first glimpse of *Mme. X*, . . . In this old age, calm after the storms, I have anyway the final courage not to regret my lost life.

So he writes in his daybook, having propped his cane against the desk, smiling at this fantasy that he has just written, about how he will grow old and become even more like Frederic Moreau. Then he closes the daybook and leaves the study—it is New Year's Eve. A warm excitement lights up his eyes as the valet brings his coat. Walking down the steps, he smiles at the disappearance of those illusions, doubts, and dreams once fed by cowardice but now dispelled. And he wonders what a hypothetical reader would surmise in reading his real story: that he is stepping out to meet *Mme. X*, or the re-found love of the Metro, or someone else, met much later? Or something else altogether, since he had not even known what to hope for? But it does not matter; every New Year's Eve he has had to admit that the year before he could not have foretold the year to come. Once he had found the courage to believe, everything else was added onto him, even effortless certainty of what to believe.

18. The importance of the voluntarist refutation of skepticism is that its reflections are immediately relevant to local skeptical doubts as well. Yes, it is true that the possibility of self-deception undermines certainty about oneself; but no, it does not follow that the only rational course is to renounce all confidence in one's own appreciation of the facts. Yes, the facts are ambig-

ous, and confidence may ebb; alternative interpretations may leap like leopards from every tree, every step of the way. But no, there is no failure of rationality in the very phenomena of trust, confidence, and affirmation.

Philosophical solutions never give practical guidance. Once we look into just how the mere possibility of self-deception undermines certainty, it brings to light the indispensable role of will—courage, faith—to sustain any appreciation of what there is. But courage to believe what? Courage to hold to an opinion increasingly difficult to act on, or courage to sacrifice it? The solution is *not* that this question has a "right" answer, *but* that this is the right question. The guidance philosophy cannot give would, if such an a prioristic discipline could give it, remove the role of courage altogether.

## NOTES

Bibliographical note: The texts cited come in many translations, of which the following are representative: André Gide, *Strait Is the Gate*, trans. D. Bussy (New York: Knopf, 1965); Gustave Flaubert, *A Sentimental Education*, trans. A. Goldsmith (New York: Dutton, 1961); Blaise Pascal, *Pensées*, trans. A. J. Krailsheimer (Hammondsworth: Penguin, 1966); St. Augustine, *Confessions*, trans. R. S. Pine-Coffin (Baltimore: Penguin, 1961); *Soliloquies* and *On the Spirit and the Letter*, in *Basic Writings of St. Augustine*, ed. W. J. Oates (New York: Random House, 1948); *Sextus Empiricus*, trans. R. G. Bury (Cambridge: Harvard Univ. Press, 1967); Jean-Paul Sartre, *Iron in the Soul* (pub. in the U.S. as *Troubled Sleep*), trans. G. Hopkins (Hammondsworth: Penguin, 1958). The quoted lines of poetry are from "No worst, there is none," in *The Poems of Gerard Manley Hopkins*, ed. W. H. Gardner and N. H. Mackenzie (Oxford: Oxford Univ. Press, 1967), p. 100. The verse from *Galatians* is essentially the same in the three or four Bible translations I have consulted.

1. I want to thank Mark Johnston for many helpful comments and criticisms of an earlier draft.
2. George Crabbe, "The Parting Hour," in G. Crabbe, *Tales, 1812, and Other Selected Poems*, ed. H. Mills (Cambridge: Cambridge Univ. Press, 1967), pp. 136–147; J. Hillis Miller, "The Ethics of Reading: Vast Gaps and Parting Hours," in *American Criticism in the Poststructuralist Age*, ed. Ira Konigsberg (Ann Arbor: University of Michigan Slavic Pub., 1981), pp. 19–35.
3. Bas van Fraassen, "Belief and the Will," *Journal of Philosophy* 81 (1984):235–256.
4. The main sources are Sextus Empiricus, *Against the Logicians I*, and *Outlines of Pyrrhonism*; and Cicero, *Academic Questions II*.
5. See Sextus Empiricus, *Against the Logicians I*, 162–163, 170–171, 180–181, and 187–189.

6. The view of our epistemic situation which I have pursued and elaborated here is not meant to exist in isolation from more theoretical concerns. Epistemological problems, encountered in one area of reflection, soon reveal tendrils and shoots in many others. To end, then, by linking this discussion to the article cited in note 3, let me repeat from it a quotation from William James' "The Will to Believe" in which he contests W. K. Clifford's contention that it is wrong always, everywhere, and for everyone to believe anything on insufficient evidence. Clifford holds up the scientist as (professionally) living by this creed, but James insists that the scientist's two aims, to believe truth and avoid error, must be pursued each at the cost of the other. James concludes:

he who says 'Better go without belief forever than believe a lie!' merely shows his own preponderant private horror of becoming a dupe. He may be critical of many of his desires and fears, but this fear he slavishly obeys . . . a certain lightness of heart seems healthier than this excessive nervousness. . . . At any rate, it seems the fittest thing for the empiricist philosopher. (*Essays in Pragmatism* [New York: Hafner, 1948], p. 100)

## SELF-DECEPTION AS RATIONALIZATION

DAVID H. SANFORD

When the editors invited me to contribute to this volume, I had recently completed an essay on infinite regress arguments, which maintains that we often misperceive our own desire structures. As I started to develop an account of self-deception from this, a look at philosophical literature on the topic neither convinced me that I was on the wrong track nor taught me that just the same track had been traveled before.<sup>1</sup> Discussions similar to mine tend to treat as secondary what I view as primary. The following passage by Terence Penelhum captures what I regard as essential to self-deception:

The appraisal of the object furnishes justification, if there is any, for being pleased at the object (sometimes even for enjoying it). This generates a higher-level type of self-deception, viz., the invention of reasons which render the pleasures we already take more acceptable to ourselves or to others. A man may tell himself his new car is a bargain in order to justify his pleasure in it, and in order to lend some degree of weight to his arguments with his wife over the expense. In such cases the indulgence in a self-consciously irrational pleasure is a more honest state of mind—the thirst for rationality is a major source of lies.<sup>2</sup>

Penelhum provides no more details about this car-buyer, and I think the case can be filled out in several ways that involve different kinds of self-deception. However the case is filled out, it includes the following mistake: the car-buyer thinks he takes pleasure in his new car because it is a bargain, but the real reason for his pleasure is something else. I shall say that the car-buyer's belief that his new car is a bargain is an *ostensible* reason for his taking pleasure in a new car. While use of the term "ostensible" implies that appearance in the respect in question does not correspond to reality, in some contexts it is not self-contradictory to say, "In this case, the ostensible reason turned out to be the real reason." Also, when we normally call someone's professed reason "ostensible," we do not imply that he is self-deceived about his reason. He may only be attempting to deceive others while knowing perfectly well what he is up to. For the purposes of this essay, I should like *ostensible* to be understood in a narrower sense: one's ostensible reason always functions differently from the way one takes it to function. Professed reasons that are also genuine, and those that are professed with a conscious