Let me begin by presenting two opposed conceptions of the relationship between the realm of the mental and our knowledge of this realm, neither of which is the conception I shall be supporting.

One I will call, with no pretension to scholarly accuracy, the Cartesian conception. According to this, the mind is transparent to itself. It is of the essence of mental entities, of whatever kind, to be conscious, where a mental entity’s being conscious involves its revealing its existence and nature to its possessor in an immediate way. This conception involves a strong form of the doctrine that mental entities are “self-intimating,” and usually goes with a strong form of the view that judgments about one’s own mental states are incorrigible or infallible, expressing a super-certain kind of knowledge which is suited for being the epistemological foundation for the rest of what we know.

The other is the view that the existence of mental entities and mental facts is, logically speaking, as independent of our knowing about them introspectively, and of there being whatever means or mechanisms enable us to know about them introspectively, as the existence of physical entities and physical facts is of our knowing about them perceptually, or of there being the means or mechanisms that enable us to have perceptual knowledge of them. This is the view that implies that our introspective self-knowledge should be construed in terms of what in my first lecture I called the broad perceptual model. The core of the stereotype of perception underlying that model consisted of two conditions, one of them (call it the causal condition) saying that our beliefs about our mental states are caused by those mental states, via a reliable belief-producing mechanism, thereby qualifying as knowledge of those states and events, and the other (call it the independence condition) saying that the existence of these states and events is independent of their being known in this way, and even of there existing the mechanisms that make such knowledge possible.
I do not hold either of the two views I have sketched. My aim is to develop a view which rejects the second, and so rejects the claim that introspection conforms to the broad perceptual model, without falling back into the first. The view I support holds that there is a conceptual, constitutive, connection between the existence of certain sorts of mental entities and their introspective accessibility, while denying the transparency of the mental. It is a version of the view that certain mental facts are “self-intimating” or “self-presenting,” but a much weaker version than the transparency thesis.

Although I am equally opposed to Cartesianism and the broad perceptual model, it is only the latter that I shall be arguing against here. This is because it is only the latter that nowadays commands widespread support. But some of the reasons for rejecting Cartesianism may seem to count equally against the weak version of the self-intimation thesis that I am supporting, and in favor of the independence thesis.

Even before Freud it was well known to novelists and ordinary folk, if not to philosophers, that there is such a thing as self-deception about one’s motives and attitudes. And recent work in psychology has revealed that in certain circumstances people are systematically wrong in the claims they make about their reasons for actions, and that such claims are sometimes the result of “confabulation” (see Nisbett and Wilson, 1977). Of a different sort, but equally damaging to the Cartesian conception, are various phenomena that have been observed in people who have suffered brain damage. One of the most striking is “blindsight.” Here a person will emphatically deny that she can see anything in some portion of her visual field; yet when she is forced to make “guesses” about what is there, the ratio of correct to incorrect guesses will be much better than chance. Finally, there are philosophical considerations of the sort that figure in David Armstrong’s “distinct existences argument”—the argument that introspective beliefs, or introspective awarenesses, must be thought of as causal effects of the mental states that are objects of awareness, that causes and effects must be “distinct existences,” and that distinct existences, as Hume taught us, can be only contingently connected (see Armstrong 1963, 1968).

None of this implies, of course, that there are not plenty of circumstances in which people do have reliable introspective access to what is going on in their minds, or that it is not often the case that one can make introspective claims with a high degree of certainty. But what it all suggests is that to the extent that we do have reliable introspective access to mental states and processes, this is a conceptually contingent fact about us. In blindsight the person is blind to facts about his own mental condition which he would be aware of if he were normal. What the considerations I have mentioned suggest is that it is at least conceptually possible that such blindness should be much more widespread—that there could be people who are blind to a wide
variety of the mental facts to which normal people have introspective access. Of course, it will generally be agreed that introspective self-knowledge serves many useful purposes, and that creatures that are introspectively blind, or, as I shall say "self-blind," with respect to some kind of mental phenomena would tend to lose out, in the struggle for survival, to otherwise similar creatures who are not. So from an evolutionary standpoint, widespread self-blindness is not a real possibility. But it will be held that for each kind of mental fact to which we have introspective access, it is at least logically possible that there should be creatures in which such facts obtain, and who have the ability to conceive of them, but who are self-blind with respect to them. This I take to be a consequence of the independence condition that is built into the broad perceptual model of self-knowledge.

I should make it clear how I am understanding the notion of introspective self-blindness. To be self-blind with respect to a certain kind of mental fact or phenomenon, a creature must have the ability to conceive of those facts and phenomena (just as the person who is literally blind will be able to conceive of those states of affairs she is unable to learn about visually). So lower animals who are precluded by their conceptual poverty from having first-person access do not count as self-blind. And it is only introspective access to those phenomena that the creature is supposed to lack; it is not precluded that she should learn of them in the way others might learn of them, i.e., by observing her own behavior, or by discovering facts about her own neurophysiological states.

II

Judging from the frequency of its occurrence in philosophical examples, pain is the favorite mental state of philosophers. I think that there are in fact many ways in which pain is an atypical and misleading example of a mental state. But, following tradition, I shall start with it. Is self-blindness possible in the case of pain? Here, perhaps more than in any other case, intuition supports the answer I favor, namely "no." But the distinct existences argument, and the other considerations I have mentioned, may suggest that this intuition should not be trusted.

Examples abound of cases in which, allegedly, a person is in pain without being aware of the pain: the wounded soldier who in the heat of battle is oblivious to his pain, the injured athlete who doesn’t notice his pain until he is put on the bench, and so on. These may seem to support the possibility of self-blindness with respect to pain; if there are cases in which there is pain but no introspective awareness of pain, why shouldn’t there be creatures in which pain is never accompanied, and cannot be accompanied, by introspective awareness of it? But we should be wary of inferences from what can happen occasionally to what can happen as a matter of course: it may be true in Lake
Wobegon that all of the children are above average, but it can’t be true everywhere.

So let us try to imagine creatures who have intellectual, conceptual, etc. capacities comparable with ours, and who also have pain, but who are introspectively blind to their pains. Their only access to their pains is a third-person access—i.e., observing their own behavior, or their own inner physiology.

It must not be supposed that these creatures do not feel their pains. Pain is a feeling, and what they are self-blind to are, precisely, their feelings of pain. And surely it must not be supposed that their pains are not unpleasant. But can they be unpleasant to these creatures, who have no introspective awareness of them? Well, let’s try to suppose that when a pain is occurring in one of these creatures, the creature actively dislikes it and wishes that it would end. It is just that the creature is unaware of this active disliking and wishing. Does the disliking and wishing have any of its normal consequences? Let’s try to suppose that it does. When in pain the creature heads for the medicine cabinet, or calls the doctor’s office for an appointment. But what does the creature itself make of this behavior? It now appears that its self-blindness has expanded to cover much more than just its pain—it must be blind to its finding the pain unpleasant, to its wishing that it would end, and to the reasons for its behavior. To such a creature it should seem as if someone else had taken possession of its body. And, indeed, if we saw someone engaging in normal pain behavior, including visits to the medicine cabinet and calls to the doctor, all of this interspersed with wonderings out loud about what all this behavior is all about, one possibility we might consider is that there are two persons, or two “selves” or “personalities,” inhabiting one and the same body—one of which is in pain, aware of the pain, and acting accordingly, and the other of which is a prisoner in the body and a passive observer of what is going on. That would be a very strange state of affairs indeed, but it would not be self-blindness—for what the passive observer is “blind” to would not be his or her own pain, but that of the other inhabitant of the body.

One thing that seems clear is that at least some kinds of “pain behavior”—taking aspirin and calling the doctor are good examples—are intelligible as pain behavior only on the assumption that the subject is aware of pain; for to see them as pain behavior is to see them as motivated by such states of the creature as the belief that it is in pain, the desire to be rid of the pain, and the belief that such and such a course of behavior will achieve that result. Indeed, to say that a creature wants to be rid of pain presupposes that it believes that it is in pain. One can want not to have something while being agnostic about whether one has it; but one can’t want to be rid of something, to cease having it, without believing that one currently does have it.
So it seems that we cannot suppose that our self-blind creature has a desire to be rid of its pain, or engages in the kinds of pain-behavior that would be motivated by such a desire. Can we nevertheless suppose that it finds its pain unpleasant? Can it find the pain unpleasant, if it is in no way aware of the pain?

Perhaps it will be said: it isn’t that the creature finds the pain unpleasant; it is just that it is unpleasant. It has an intrinsic phenomenal character that constitutes its being unpleasant, and so makes it such that anyone who was introspectively aware of it would find it unpleasant. One feature of this view is that it makes the connection between a state’s being unpleasant and its subject’s wanting to be rid of it a purely contingent connection. This suggests some apparently alarming possibilities. Maybe all of what we take to be innocent pastimes produce in us states that are extremely unpleasant, but of which we are totally unaware. Wouldn’t that be terrible? Well, maybe not. Maybe this is a case of what you don’t know can’t hurt you. And likewise with our self-blind man: he is in extreme pain, his pains are extremely unpleasant, but there is nothing bad about this because he is unaware of his pains and, in this sort of case anyhow, what you know can’t hurt you. His pains hurt, but they don’t hurt him. But this of course is nonsense.

We have seen that if you take away the link with introspective beliefs from the total causal role of pain, you take away a lot else—namely the role of pain in explaining the behavior we take as indicating distress and the desire to feel better. But you don’t take away everything. Part of the causal role of pain consists in its being caused by certain kinds of things—bodily damage of various sorts—and its causing behaviors, such as winces, grimaces, and moans, that can be involuntary and do not have to be seen as motivated or “rationalized” by beliefs and desires. Certainly a state could play this role without its subject having introspective awareness of it. And such a state would be, if not bad in and of itself, at least indicative of something bad, namely bodily damage. This it would share with pain. But it seems obvious that a state that played this causal role, but did not play any more of the standard causal role of pain than this, would not be pain. Indeed, it would not be a mental state at all.

III

So self-blindness with respect to pain seems to be an impossibility. What goes for pain would seem to go as well for itches, tingles and the like. And it would also seem to go for such phenomena as having a ringing in one’s ears. Many people are plagued by ringing in the ears. It would sound more than a little ludicrous to suggest that countless others have the same condi-
tion but, mercifully, are totally unaware of it, being self-blind with respect to that particular mental phenomenon.¹

I now want to turn our attention to the sorts of experiences that occur in normal perception, and consider whether one could be self-blind with respect to these. Having such an experience will be a matter of things appearing (looking, sounding, feeling, etc.) certain ways, or of one's being "appeared to" in certain ways. We will suppose that our putatively self-blind person has normal perceptual access to things in her environment, and to the fact that she is perceiving these things in this or that sense-modality, and that she has normal access to her beliefs (the question of whether there can be self-blindness with respect to beliefs will be addressed latter). She makes judgments like "I see a tree," "I taste something salty," and "I feel something with silky texture." So she is not self-blind with respect to her seeing, tasting, etc. But she is, we are to try to suppose, self-blind with respect to the perceptual experiences involved in these kinds of perception.

Someone with normal perceptual access to things in the environment must at least be sensitive to the nature of her perceptual experiences, in the following sense: the perceptual judgments she makes about things in the environment must be a function of the nature of her perceptual experiences, of how she is "appeared to," together with her beliefs, in particular her beliefs about perceptual conditions. This is not to say that her perceptual judgments are inferred from the nature of the perceptual experiences. If they were, there would have to be knowledge of the perceptual experiences, and self-blindness with respect to them would be ruled out from the start. But normally the perceptual judgments are not inferred. Normally it is the having of the experience, not the belief that one has it, that issues in the perceptual judgment. Normally, too, the beliefs about perceptual conditions that contribute to the judgment are only tacit, except where there are indications that the perceptual conditions are not normal. Still, the person must be sensitive to a vast number of combinations of experiences and beliefs about perceptual conditions. If she has an experience "as of" red and believes that conditions are normal, she will judge that the wall is red; if she has that same experience and believes that the wall is being illuminated by red light, she may judge that it is white; and so on. The question is whether the person could have the appropriate sensitivity to combinations of experiences and beliefs without having the capacity to be aware of the experiences.

I think that if we are out to conceive of a case of self-blindness with respect to perceptual experiences, the best we can do is to suppose that our putative self-blind person is "hard-wired" to respond to various experience-

¹ I am not talking about the possibility, which I do not deny, that some people have it but do not notice it. What would constitute self-blindness is not not-noticing but the impossibility of noticing.
belief combinations by making the appropriate perceptual judgments. Such a person would in many cases be in a position to infer that she was having a certain kind of experience. Knowing what perceptual judgment she had made, and knowing what her belief about perceptual conditions is, she infers what her perceptual experience must be. So, for example, noting that she has judged that the wall is white, and noting also that she believes that the wall is being illuminated by red light, she infers that she must be having an experience as of something red. If the making of such inferences were automatic, she might seem like a normal person, making the right judgments about how things appear to her in a wide variety of circumstances. But she would give away her self-blindness by her inability to say what her visual experience is like if she is deprived of information about perceptual conditions. We rig things so that she cannot see whether perceptual conditions are normal, and tell her that we have done this, and then ask her about the color of the wall. Like a normal person, she is unwilling to make any judgment about what color the wall is. But unlike a normal person, she is unable to make any judgment about what color the wall appears to her.

I do not deny that this case is possible, but I do deny that it should count as a case of self-blindness—for I deny that the person, as described, has perceptual experiences in the sense we do. By hypothesis, she has states that play some of the functional role of perceptual experiences. In combination with certain beliefs about perceptual conditions, these states give rise to the same perceptual judgments about the world that certain of our perceptual experiences give rise to when combined with those same beliefs. But their role is not rich enough to give the subject normal perceptual access to the environment. And it is not rich enough to make them perceptual experiences of the sort to which we have introspective access.

I have been speaking as if beliefs about perceptual conditions were simply given to us. But such beliefs are themselves often grounded on our perceptual experiences. Sometimes, thrown into an unfamiliar setting, we simultaneously reason to conclusions about what sorts of things we are perceiving and to conclusions about what the perceptual conditions are. And the beliefs that influence our judgments are not limited to beliefs about perceptual conditions. Any of one’s beliefs about how perceptible things behave under various conditions, about how the appearance of things of certain kinds should change as one changes one’s situation with respect to them, and so on, can play a role in determining one’s perceptual judgments about how things are in one’s vicinity. Think, for example, of how one might move from a belief that one is perceiving some ordinary objects to the belief that one must be the victim of some hologram-produced or drug-produced illusion—or how one might move from a suspicion that the latter is true to a firm belief that things are after all as they appear. In such cases one is engaged in a certain kind of low level theorizing. And the data for the theorizing include facts
about how one is appeared to, i.e., about the nature of one's experiences. So while often perceptual judgments are not inferred from facts about experiences, often they are. And having normal perceptual access to things in one's environment requires having the capacity to engage in this sort of theorizing, and therefore requires having access to the facts about current experience which provide the data for it.

We tried to suppose that our putatively self-blind person was "hard-wired" to respond appropriately to experience-belief combinations, in a way that did not involve having introspective access to the experiences. But it is not credible that the low grade theorizing I have just been talking about, in which an interplay between one's current perceptual experiences and one's entire body of background beliefs gives rise to perceptual judgments about the world, could be the product of such hard-wiring. Perhaps one could say, stretching the notion, that we are "hard-wired" to respond rationally to facts about our experience in the light of our background beliefs; but if so, that has to mean that we are hard-wired to have introspective access to those facts about our experience.

Suppose that neurosurgeons set to work on someone who is blind in the ordinary sense, i.e., totally unable to see, and manage to get her into the state I imagined the putatively self-blind person to be in. They wire her brain and optical system in such a way that certain external situations produce certain internal states, and by an astounding breakthrough, they wire her brain in such a way that these internal states together with certain beliefs about perceptual conditions give rise to perceptual judgments about her environment—judgments that are normally true when her beliefs about perceptual conditions are true. But when she has no relevant beliefs about perceptual conditions, she is totally unable to say which of these states she is in, and at other times she can do so only by inferring what state she is in from what perceptual judgment she is making and from what she believes about what perceptual conditions she is in. If thrown into a situation in which the sort of low grade theorizing discussed above is called for if one is to know what the perceptual conditions are, she is as helpless as she was before the scientists got to work on her. My question is, would it be reasonable to suppose that the scientists had bestowed on her visual experiences that are phenomenally just like those we have, but to which she lacks introspective access? And it seems to me that the answer is: obviously, no.²

² After this was written it occurred to me that what the argument of this section implicitly invokes is the application of something like Stephen Stich's notion of inferential promiscuity (see Stich, 1978) to perceptual experiences. It would be wrong to say that perceptual experiences are inferentially promiscuous vis à vis their actual contents. For it is by no means the case that contents of perceptual experiences are invariably believed by the subject, and contents that are not believed will not be available as premises of inferences. What is inferentially promiscuous is not the putative information that is the content of the experience (e.g., that there is a red wall in front of one), but the information that one has
I want now to consider the possibility of self-blindness with respect to another aspect of our mental lives, namely the exercise of our wills. We have an introspective access, or at any rate a distinctive first-person access, to what we are doing or trying to do, and to the intentions with which we are doing or trying to do whatever it is. To what extent could we lack this?

If self-blindness is possible with respect to a given mental phenomenon, one would think that there should be something that would show, or at any rate provide good evidence, that someone was afflicted with such self-blindness, and that this should include that person denying, with apparent sincerity, that she was aware of the phenomenon in question in herself. There is an obvious difficulty in the supposition that someone might show in this way that she is self-blind with respect to all exercises of her will, i.e., with respect to the mental side of all of her voluntary actions. For either the behavior which seems to constitute her telling us that this is so is really an action on her part, or it is not. If it is not, it cannot be the required sort of evidence that she is self-blind. But if it is, then she is telling us that she is self-blind with respect to the very activity she is engaged in, which means that for all she knows "from the inside," and perhaps for all she knows simpliciter, this may not be an action on her part at all. And it hardly seems coherent to suppose that there could be a sincere assertion that says of itself that for all the speaker knows it is not an assertion of hers. "I am not saying this" seems to be a pragmatic contradiction on a par with "I do not exist"; and "For all I know I am not saying this" seems on a par with "For all I know, I do not exist."

To finesse this difficulty, let the suggestion be, not that someone could be self-blind with respect to all exercises of her will, but that someone could be self-blind with respect to some large class of such exercises—say all of those involved in actions other than certain speech acts. So a putative case of self-blindness will be one in which we observe what looks like someone engaged in some kind of action, and hear coming from the person’s mouth statements to the effect that she has no introspective awareness of what she

an experience with this content. It is the availability to one’s system of this information that determines, in conjunction with one’s background beliefs, whether one believes the content of the experience and, if one doesn’t, what if anything one believes instead. And this information is inferentially promiscuous in the sense that it will combine with any relevant beliefs that one has. Of course, to speak of it as inferentially promiscuous is already to imply that the information is believed—only information that is believed can figure as a premise in inferences issuing in new beliefs. My claim is that it is of the essence of having perceptual experiences that the system behaves as if such information is believed, and that that is sufficient for its being believed. This is not to imply that over and above the perceptual experience there is a separate state which is the belief that one has a perceptual experience with such and such content; it may be that the perceptual experience doubles as a belief about itself (just as, I suggest later, a first-order belief can double as a second-order belief to the effect that one has that first-order belief).
is doing, and that she can only gather this, as we do, by observing what she does.

But the coherence of the case is still in doubt. As I have defined it, self-blindness is not supposed to involve any cognitive deficiency. Nor is our person supposed to be deprived of information about the environment relevant to the execution of her action plans. So our self-blind person ought to be capable of having and carrying out action plans as complex as those within the repertoire of an ordinary person. But many action plans involve combinations of verbal and non-verbal behaviors, all aimed at the same goal. In such a case, if the agent is self-blind with respect to the total execution of the action plan, she will have to be self-blind with respect to the verbal behaviors it involves. So what we apparently have to imagine is that we have someone engaged in what looks like the execution of a rational action plan, with verbal behavior integrated with the nonverbal behavior, and that the only thing out of the ordinary is that the utterances that are integrated with the rest of the behavior are interspersed with ones that profess introspective ignorance of what is going on.

No doubt it is conceivable that we should observe such a thing. But it does not seem that if we did we should conclude that we had come across a case of self-blindness. Various alternative possibilities suggest themselves. One is similar to one that came up earlier in our discussion of pain. Maybe the body in question houses two persons, or two "selves" or "personalities," one of which controls all of its nonverbal behavior and some of its verbal behavior, and the other of which controls that part of the verbal behavior which consists in professions of introspective ignorance of what is going on. In that case, of course, the blindness is not self-blindness, and not introspective blindness; it is a matter of one of the occupants of the body being blind to what the other occupant is up to—which is compatible with the other occupant having normal introspective access to what he or she or it is doing and to the intentions with which it is done.

Let us refer to the person performing the bulk of the actions realized in the movements of this body as "the agent," and let us refer to the person who professes introspective ignorance as to what is going on as "the agnostic." If anything would make it reasonable to say that the agent and the agnostic are one and the same person, it would be the fact that what the agent is doing fits with, and is "rationalized" by, beliefs and desires that can independently be ascribed to the agnostic. Now unless we already know that the agent and the agnostic are one, we cannot take the overall behavior of the agent/agnostic body as evidence of the agnostic's beliefs and desires—that will give us evidence of the agent's beliefs and desires, but whether these are the agnostic's beliefs and desires is just what is in question. If the only access we have to the agnostic's mental states are her avowals of ignorance about the agent's actions, then of course we have no access at all to her be-
liefs and desires. But let us suppose that these avowals of ignorance are part of a larger discourse in which various beliefs and desires are also avowed, and that the various parts of this discourse cohere in such a way as to make it reasonable to conclude that there is a single subject responsible for all of it. Then, perhaps, we could get enough information about the agnostic's beliefs and desires to know that they rationalize the agent's actions to such an extent that it is reasonable to conclude that those actions stem from those beliefs and desires and, therefore, that the agnostic and the agent are one.

Now if we, as observers, are in a position to conclude this, so is the agnostic. For it is the agnostic's introspective knowledge of her own beliefs and desires, as manifested in her utterances, that provide the basis of our reasoning; so, necessarily, the agnostic knows at least as much about her beliefs and desires as we do. On the other hand, if she didn't have this introspective knowledge, neither we nor she could even begin the enterprise of finding out whether the agent's actions are hers. The question of whether there can be self-blindness with respect to beliefs and desires will be discussed shortly. But what we have seen here indicates that whether or not it is necessary *simpliciter* that one have introspective access to one's beliefs and desires, it is necessary that one have such access if one is to have any access at all, even an inferential access, to one's intentions, volitions, and actions.

So it is a necessary condition of one's having knowledge of one's own agency, even inferential knowledge of it, that one *not* be self-blind with respect to one's beliefs and desires. It is a further question, which is too complex for me to take up here, whether introspective access to one's own beliefs and desires is sufficient, as well as necessary, for knowledge of one's own agency. If it is, self-blindness with respect to agency is impossible. I am inclined to think that this is so; but I will content myself here with the conclusion that knowledge of one's own agency is incompatible with self-blindness with respect to one's beliefs and desires.

V

What I have just argued could be turned into an argument against the possibility of a creature that is self-blind with regard to its own beliefs and desires by being supplemented with an argument against the possibility of a creature, suffering from no cognitive deficiency, that is incapable of having even inferential knowledge of its own agency. But instead of pursuing this line, I will argue directly against the possibility of self-blindness with respect to beliefs and desires. Here I shall be developing arguments I have presented elsewhere (see Shoemaker, 1988 and 1991).

Remember that as I have defined self-blindness, it is supposed to be like ordinary blindness in not entailing any *cognitive* deficiency. The person who lacks sight can in principle be equal in intelligence, rationality, and conceptual capacity, to any sighted person. Likewise, the person who lacks access by
inner sense to some kind of mental state, and so is self-blind with respect to
that kind of mental state, can in principle be equal in intelligence, rational-
ity, and conceptual capacity to someone who is not self blind.\(^3\) What I shall
be arguing, in the first instance, is that if someone is equal in intelligence, ra-
tionality, and conceptual capacity to a normal person, she will, in conse-
quence of that, behave in ways that provide the best possible evidence that
she is aware of her own beliefs and desires to the same extent that a normal
person would be, and so is not self-blind.

I shall use the expression “rational agent” as short for “person with
normal intelligence, rationality, and conceptual capacity,” and I will use the
expression “self-aware agent” as short for “person who has normal intro-
spective awareness of his or her own beliefs and desires.” Let us see why a
rational agent will give every indication of being a self-aware agent.

First, consider what she will say. Rationality does not guarantee honesty,
and it does not guarantee openness. But one thing that will be true of a ratio-
nal agent, whether her intentions are honest or dishonest, is that she will an-
swer affirmatively to the question “Do you believe that P?” if and only if
she will answer affirmatively to the question “Is it true that P?” This is as-
suming, of course, that she has the relevant concepts and knows the meanings
of the relevant words—in particular, the concept of belief, and the meaning
of the word “believe.” Mastery of those concepts and meanings, together
with her rationality and intelligence, will be enough to make her appreciate
the logical impropriety of affirming something while denying that one be-
lieves it, or denying something while affirming that one believes it, and so
enough to make her give appropriate answers to questions about what she be-
lieves—true ones if her intentions are honest, false ones if she is out to de-
ceive. But it will also make her realize that “I believe that P,” said in a given
context, necessarily has the truth value that an affirmative answer to the
question “Do you believe that P?” would have in that context, and essen-
tially the same effect on the audience; and this should dispose her to some-

\(^3\) A qualification is needed here. It widely held that a person blind from birth would neces-
sarily lack some concepts, namely color concepts, which normally sighted persons have.
It does not seem obvious to me that this is right; but supposing it is, then it is not quite true
that someone blind from birth could equal a normal sighted person in conceptual capac-
ity. And, carrying the analogy along, perhaps there are some concepts that a self-blind
person could not have. But notice that “blind” doesn’t mean “blind from birth,” and “self-
blind” shouldn’t be taken to mean “self-blind from birth.” People who are blind but not
blind from birth clearly can have concepts of particular colors; and presumably someone
self-blind but not blind from birth can have whatever concepts any normal person has.
Insofar as my argument depends on the possibility of the self-blind person being equiva-
 lent in conceptual capacity to a normal person, what the argument directly shows is the
impossibility of a case in which someone has introspective access to the mental states in
question long enough to acquire the relevant concepts, and then becomes self-blind. I
think that if the impossibility of this is established, there will be little plausibility in the
view that the corresponding sort of self-blindness from birth is possible.

282 SYDNEY SHOEMAKER
times volunteer the statement “I believe that P” in circumstances in which she would respond affirmatively to the question “Do you believe that P?” if it were asked.

Similar remarks apply to desire. The rational agent who wants X and has normal mastery of language will, ceteris paribus, respond affirmatively to the question “Shall I give you X?” And given her mastery of the concept of desire, she will respond affirmatively to the question “Do you want X?” if she will respond affirmatively to the question “Shall I give you X?” So she will, unless she has devious motives, give correct answers to questions about what she wants. But since she knows, given her mastery of the relevant concepts, that it is true to say “I want X” just in case it is true to answer affirmatively to “Do you want X?”, and that these are similar in their effects on others, whatever motivates her to answer affirmatively to the question of whether she has a certain want can be expected motivate her to volunteer the statement that she has that want in cases in which the question is not raised.

What these arguments establish, at least in the first instance, is only a connection between the ability to give verbal expression to beliefs and desires and the ability to self-ascribe them. But there is another pair of arguments that establishes a more direct connection between the having of beliefs and desires and the self-ascription of them—or at any rate the behavior we would take as evidence of such self-ascription.

Again let us begin with the case of belief. A rational agent who believes that P will be disposed to use the proposition that P as a premise in her reasonings. Moreover, she will know that for any proposition whatever, if that proposition is true then, other things being equal, it is in anyone’s interest to act on the assumption that it is true—for one is most likely to achieve one’s aims if one acts on assumptions that are true. She will also know that to act on the assumption that something is true is to act as if one believes the thing; and she will know that if it is in one’s interest to act in this way it will normally be in one’s interest to make manifest to others that one is so acting—this will increase the likelihood that other believers in the truth of the proposition will cooperate with her in endeavors whose success depends on the truth of the proposition, and it will tend to promote belief in that proposition within her community and so to promote the success of endeavors whose success depends both on the proposition being true and on its being believed to be true by participants in the endeavors. Knowing that it is in anyone’s interest to act in these ways if a proposition is true, she will know that it is in her own interest to so act. So she can reason as follows: “P is true. So it is in my interest to act as if I believed that P and, in normal circumstances, to act in ways that would make others believe that I believe that P. Since the circumstances are normal, I should so act.” Assuming that she is rational enough to act in conformity with the conclusions of her own practi-
cal reasoning, and to know what will make others believe that she believes something, this should lead her to behave in the ways characteristic of someone trying to manifest to others that she believes that P, including saying “I believe that P.”

Similar considerations apply to wanting. Wanting that P be the case, unlike believing that P is the case, does not of course involve a disposition to use the proposition that P as a premise in one’s reasonings. But it does involve, in a rational agent, conforming one’s practical reasoning to the injunction “In the absence of reasons for not doing this, so act as to promote the likelihood of P’s being the case.” And the rational agent will know that in certain circumstances, namely when one is in the company of people who are cooperative and obliging, or willing to trade favors, presenting evidence that one wants something to be the case will increase the likelihood of its becoming the case. So when she is in such circumstances, the rational agent who wants it to be the case that P will be disposed to present evidence that she wants this, presumably including saying things like “I want it to be the case that P.”

There is a worry about this pair of arguments. Consider the argument about belief. I said there that the rational agent who believes that P would be disposed to use the proposition that P as a premise in her reasonings, and would have available an argument from this premise to the conclusion that she should act as if she believed that P, including saying things like “I believe that P,” and I took this to show that such a rational agent could not be self-blind. But it is natural to reply to this that while such an argument is available, it is not by the use of such an argument that we, who have introspective access to our beliefs, come to make statements about our own beliefs. And, it seems natural to add, if it is only by the use of such arguments that a person could rationally come to such putative belief reports, then that person would be self-blind, even if nothing in her behavior could give away her self-blindness.

But this worry rests on a misunderstanding. The reason for pointing out that such reasoning is available is not to suggest that it regularly goes on in us—obviously it doesn’t—but rather to point out that in order to explain the behavior we take as showing that people have certain higher order beliefs, beliefs about their first order beliefs, we do not need to attribute to them anything beyond what is needed in order to give them first order beliefs plus normal intelligence, rationality, and conceptual capacity. What the availability of the reasoning shows is that the first order states rationalize the behavior. And in supposing that a creature is rational, what one is supposing is that it is such that its being in certain states tends to result in effects, behavior or other internal states, that are rationalized by those states. Sometimes this requires actually going through a process of reasoning in which one gets from one proposition to another by a series of steps,
and where special reasoning skills are involved. But usually it does not re-
quire this. I see an apple and I reach for it. It is rational for me to do so, and
this can be shown by presenting an argument, a bit of practical reasoning that
is available to me, in terms of my desires and my beliefs about the nutri-
tional and other properties of apples. But I needn’t actually go through any
process of sequential reasoning in order for the beliefs and desires in ques-
tion to explain and make rational my reaching for the apple. And no more
does the rational agent need to go through a process of sequential reasoning
in order for her first order belief that P, plus her other first order beliefs and
desires, to explain and rationalize the behavior that manifests the second or-
der belief that she believes that P.

This argument strikes at the heart of the perceptual model of self knowl-
dge. From an evolutionary perspective it would certainly be bizarre to sup-
pose that, having endowed creatures with everything necessary to give them
a certain very useful behavioral repertoire—namely that of creatures with
ormal human intelligence, rationality, and conceptual capacity, plus the
ability to acquire first order beliefs about the environment from sense-per-
ception—Mother Nature went to the trouble of installing in them an addi-
tional mechanism, a faculty of Inner Sense, whose impact on behavior is
completely redundant, since its behavioral effects are ones that would occur
anyhow as the result of the initial endowment.

There is an additional argument against the possibility of self-blindness
which I can sketch only very briefly here. Briefly, the idea is that it is essen-
tial to being a rational being that one be sensitive to the contents of one’s be-
ief-desire system in such a way as to enable its contents to be revised and
updated in the light of new experience, and enable inconsistencies and inco-
herences in its content to be eliminated. In some cases, no doubt, the sensitiv-
ity responsible for belief and desire revision consists in the operation of au-
tomatic mechanisms that do not involve anything as cognitively sophisti-
cated as introspective awareness. Perhaps this is always true in the case of
lower animals. But in an important class of cases the rational revision or ad-
justment of the belief-desire system requires that we undertake investiga-
tions aimed at determining what revisions or readjustments to make—either
“external” activities of conducting tests or experiments, or “internal” ac-
tivities of constructing and evaluating arguments. Testing and argument
construction are voluntary activities, which are the results of beliefs and de-
sires that “rationalize” them. And here the beliefs will include higher order
beliefs, beliefs about the contents of one’s belief-desire system. One does not
test propositions at random, or evaluate arguments for arbitrarily chosen
propositions. Where it makes sense to engage in such activities is where, for
example, the outcome of the investigation will decide which of several ways

---

of rendering one's belief system consistent is the best way; and to know this one must know a good deal about what the contents of that system are. It is because one believes certain things, and has apparent reason to believe certain other things, and because not all these things can be rationally believed together, that one initiates the particular course of investigation one does. What rationalizes the investigation are one's higher order beliefs about what one believes and has reason to believe. Creatures without introspective access to their beliefs and desires would lack this resource for rational revision of these beliefs and desires, and would fall short of normal human rationality.

I want to make two final points in clarification of what I have said about knowledge of one's beliefs. First, if we distinguish, as surely we must, between "explicit" beliefs and "tacit" beliefs, all that I mean to commit myself to by my arguments is the view that a rational person who believes that P at least tacitly believes that she believes that she believes that P. The best account of tacit belief I know of is that of Mark Crimmins. He suggest that "A at-least-tacitly believes p just in case it is as if A has an explicit belief in p," where the right hand side of this can be paraphrased "A's cognitive dispositions are relevantly as if A has an explicit belief in p" (see Crimmins, 1992). My claim is that to the extent that a subject is rational, and possessed of the relevant concepts (most importantly, the concept of belief), believing that p brings with it the cognitive dispositions that an explicit belief that one has that belief would bring, and so brings with it the at least tacit belief that one has it.

My second point concerns my argument which starts from the point that a rational person who believes that P will be disposed to use the proposition that P as a premise in her reasonings. We should be clear that all of us regularly fall sort of the ideal of rationality here articulated. We do so, of course, when we repress beliefs, or are self-deceived about them. But we also do so in the more commonplace case in which we temporarily forget, or fail to access, something we perfectly well know—a good example of this is the case in which, having put one's watch in one's pocket so as to wash the dishes, one looks all over for it before remembering that one put it there. In such cases one has a belief that is, perhaps only temporarily, not available for use as a premise, or as a guide to action. What my argument shows is that if a belief is available as a premise, then the subject will be disposed, insofar as she is rational, to act in ways that manifest the belief that she has it. But this is only part of what the "self-intimation" of beliefs, and its connection to rationality, amounts to. The other part consists in the fact that while one can have beliefs that are not available, the very having of a belief that P involves being disposed to have, under certain conditions, an available belief that P. It

---

5 I take this example from Crimmins, 1992.
is a mark of rationality that one's beliefs will become available when there is occasion to use them, i.e., when one is faced with problems to which their contents are relevant. And here rationality differs in degree—some are more rational than others, and the same individual may be more rational about some matters than about others. But the minimal degree of rationality that is a necessarily condition of being a believer at all requires that there is at least a tendency of beliefs to become available under certain conditions. If states of a creature encoded information about the world, but did so in such a way that there was no tendency for this information to figure in the creature's reasoning or the guidance of its actions, those states would not count as beliefs. So an account of access to beliefs—and I think the same is true of desires and intentions—should have two parts. The first would be an account of how beliefs become available on appropriate occasions. I have said nothing about this, except that there must be a tendency for it to happen. The second would be an account of the relation between a belief's being available and the subject's believing that she has it. The latter is what my discussion here has mainly been about.

VI

If, as I have argued, it is of the essence of many kinds of mental states and phenomena to reveal themselves to introspection, what sort of account can we give of the relation of these to the introspective awareness of them? Take the case of belief, which I have just been discussing. A natural way of thinking is to say that the belief that P is one thing, the introspective belief that one believes that P is another, and that introspective awareness consists in the first giving rise to the second. Going with this is a way of thinking about how those mental states are realized-in the brain. The belief that P is realized in one neural state, perhaps a highly distributed one, the belief that one believes that P is realized in another, and the brain is so wired that, under certain conditions, the first neural state gives rise to the second. This is what suggests that the existence of the belief should be logically independent of its being accessible to introspection, and that self-blindness with respect to belief ought to be a logical possibility. How are we to avoid this conclusion without denying, what seems obvious, that to believe something is not the same as to believe that one believes it? Here again we have the "distinct existences argument."

When we speak of the realization, or implementation, of a mental state, we should distinguish between what I call the core realization of it and what I call the total realization of it. The core realization will be a state that comes and goes as the mental state comes and goes, and which is such that, given relatively permanent features of the organism, it plays the "causal role" associated with that state—it is caused by the standard causes of that state, and causes its standard effects, usually in conjunction with
other states. The total realization will be the core realization plus those relatively permanent features of the organism, features of the way its brain is "wired," which enable the core realization to play that causal role. So, for example, given the standard philosophical fiction about pain and C-fiber stimulation, C-fiber stimulation would be the, or at any rate a, core-realization of pain, while the (or a) total realization would be C-fiber stimulation plus those features of the way the brain is wired which make it the case that C-fiber stimulation is related as it is to bodily damage, to wincing and groaning, and to the desires and beliefs and distractions associated with pain.

This suggests a way, in fact two ways, in which a mental state and the belief that one is in that mental state could be different and yet such that it is of the essence of the mental state that under certain conditions it gives rise to that belief. One possibility is that while the two states have different core realizations, and so different total realizations, their total realizations overlap in a certain way. Suppose that the core realization of pain in us is C-fiber stimulation, and that the core realization of the belief that one is in pain is something we will call Z-fiber stimulation. While these are distinct states, it could be that the total realization of pain includes being such that under certain conditions C-fiber firing causes Z-fiber firing, and that the same thing is part of the total realization of the belief that one is in pain. If all of the possible total realizations of a state and of the belief that one is in that state were related in this way, it would be of the essence of that state that under certain circumstances it tends to give rise to that belief, and it might also be of the essence of that belief that in the absence of malfunctioning it is caused by that state. This would amount to the state’s being self-intimating, and the belief’s being, if not infallible, at least highly authoritative.

But this is not the only way in which we could get this result. Instead of the first-order state and the belief about it having different core-realizations but overlapping total realizations, it might be that they have the same core realization and that the total realization of the first-order state is a proper part of the total realization of the first-person belief that one has it. I suggest that this is how it might be when the two states are, respectively, an available first-order belief, say that the sun is shining, and the second-order belief that one has that first-order belief, say the belief that one believes that the sun is shining. What my earlier discussion suggested is that if one has an available first-order belief, and has a certain degree of rationality, intelligence, and conceptual capacity (here including having the concept of belief and the concept of oneself), then automatically one has the corresponding second-order belief. If it is possible to have the available first-order belief without having the second-order belief, this is because it is possible to have it without having that degree of rationality, intelligence, and conceptual capacity—which is perhaps the case with some lower animals. But on this conception, all you have to add to the available first-order belief, in order to get
the second-order belief is the appropriate degree of intelligence, etc. It is not that adding this pushes the creature into a new state, distinct from any it was in before, which is the core realization of believing that it has this belief. It is rather that adding this enables the core realization of the first-order belief to play a more encompassing causal role, one that makes it the core-realization of the second-order belief as well as the core-realization of the first-order belief. The more encompassing role will include its contribution to the monitoring of person's belief-desire system, including the motivating of investigations aimed at revising the system in the direction of greater coherence. The more encompassing role will also include its contributing to the motivating of behavior, including speech behavior, aimed at intimating to others that the person has that belief.

If, as I have suggested, believing that one believes that P can be just believing that P plus having a certain level of rationality, intelligence, and so on, so that the first-order belief and the second-order belief have the same core realization, then it will be altogether wrong to think of the second-order belief in such cases as caused by the first-order belief it is about. Here the relation of an "introspective" belief to the state of affairs it is about is altogether different from the relation of a perceptual belief to the state of affairs it is about. In the other sort of case, where the first-order mental state and the belief about it have different core-realizations but overlapping total realizations, there is a sense in which the first-order state can cause the belief about it; it causes it in the sense that its core realization causes the core realization of that belief. And presumably that will be the standard way in which such a belief will be acquired. But this causal relationship is altogether different from the sort that holds between a perceived state of affairs and the perceptual belief about it. Thought of as a relation between the mental states, rather than between their core realizations, the relation is an internal one, whose relata are not "distinct existences." Either way, our introspective access to our own mental states differs markedly from our perceptual access to things in our environment.

But, to sum up, I think that the fundamental difference between perception and introspection is the failure of the latter to satisfy the "independence condition." Perception and introspection are of course alike in being modes of non-inferential knowledge acquisition. But in the case of perception, the mechanisms involved are ones whose function it is to give us knowledge of an independent reality, one that was not made to be accessible to us and our faculties. In the case of introspection, on the other hand, the reality known and the faculty for knowing it are, as it were, made for each other—neither could be what it is without the other. The contrast is perhaps not quite as sharp as I have drawn it. From an evolutionary point of view, it may be that it is as much true that flowers developed the colors they have in order to be perceived by bees as that bees developed color vision in order to
see the colors of flowers; flowers need bees for pollination as much as bees need flowers for nectar. And as my next lecture will emphasize, the features of things we perceive include ones that constitutively involve relations to perceiving creatures. Not everything we count as sense perception conforms even to the broad perceptual model. But the central cases of it do—and it is the assimilation of introspection to these that I have been campaigning against.