NOÛS 35:4 (2001) 542–568

Self-Knowledge and Phenomenal Unity

CHARLES SIEWERT
University of Miami

1

How do you know what you're currently experiencing? How do you know that you feel pain when you do, or that something looks purple or round to you when it does? The question needs interpretation—but it is intelligible. And apparently, it's not inconsequential: knowledge of one's own experience seems to play an important and pervasive part in knowledge of the natural world, as well as in personal life, and clarity about what has been called "introspection" seems crucial to a sound methodology in studies of the mind generally. I would like to propose a way of understanding this question about self-knowledge and to develop an answer that makes central the notion that phenomenally conscious experience is *unified*.

We need first to recognize that the way in which we each know our own experience differs from the way in which we know others'. This is not to say that one discovers what kind of experience one has by means of some peculiar procedure. The point is rather: the *warrant* one has for first-person judgments about experience commonly *differs in kind* from that had for third-person judgments about experience.

To justify this we need make no assumption about the nature of this distinctive first-person warrant. And we need not claim for so-called "introspection" some special "privilege" such as infallibility, indubitability, or incorrigibility. We need only reflect on this. Often we would take ourselves to know what we experience in circumstances where, assuming that we do know, and that we have warrant for believing what we know about ourselves, it follows that the kind of warrant we have differs from the kind another person would ordinarily need, in order to know the same about us. This will be so, in those (very common) cases where what another would need to observe to acquire such knowledge is simply not there to be observed at the time. So I may be suffer-

© 2001 Blackwell Publishers Inc., 350 Main Street, Malden, MA 02148, USA, and 108 Cowley Road, Oxford OX4 1JF, UK.

ing from back pain, and know that I feel this pain, at such time as I have done nothing of the sort ordinarily needed for *you* to know that I feel pain.

We should reject arguments that say it can't be correct to speak of self-knowledge in these cases, on the grounds that one knows a given claim to be true, only if one can cite reasons or evidence so as to justify it without employing the claim itself. Admittedly, one is generally unable to offer a non-circular justification of what one claims to know about one's own experience. When one would claim to know that one feels pain, faced with the question "How do you know?", one would be likely to respond either with mere puzzlement, or with something like: "I know that I feel pain, because I *do* feel pain," or maybe, "Because *I* am the one who feels it." But maybe not all knowledge requires that the knower be inclined to offer a noncircular justification of what he knows.

This does not imply that when one knows and believes that one has a certain sort of experience, one knows but believes *without reason*. Nor should we infer from our inability to offer non-circular justifications for our first-person judgments about experience, that we have at most here a kind of knowledge *without warrant*. To do so would lead us to some absurd conclusions. Suppose that you have been sitting for a while in an airplane and begin feeling back pain. You get up, in order to relieve the discomfort, and your friend sitting next to you asks, "Is your back hurting you?" You lie convincingly, betraying no sign of suffering: "No, I just feel thirsty; I'm going to get a drink of water." Though you believe and know you don't feel thirst, but back pain, what your hearer has most warrant to believe about you in this situation is just the opposite. However, if we assume you have only a distinctive first-person knowledge that you feel pain, without distinctive first-person warrant for believing you do, what follows? Since you have no warrant at all for believing you feel pain, but do have available the misleading evidence you provide others, it seems you would have more warrant for believing your own lie, than for believing what you know to be true. So you would also have more warrant for denying you know that you're in pain than for affirming it. And in such cases we could never have warrant for believing we were convincingly deceiving others about our own experience. If saying one has no distinctive warrant for present-tense first-person judgments about experience yields such consequences, then it is clearly unacceptable.[1]

The sort of "deception" case just discussed helps focus the discussion of psychological self-knowledge in two ways. First, it provides us with a relatively neutral but still substantive starting point. For it does not beg deservedly controversial questions about the nature of self-knowledge—for example, questions concerning its similarity or dissimilarity to perceptual knowledge, and its vulnerability to error, doubt, or correction. And it requires no great assumptions about the nature of mind or experience. But even so, it shows that we do in fact have a distinctive type of warrant for first-person judgment about current experience. Second, it suggests a crucial test for any account of

this first-person warrant: not only should we say what this consists in and explain why we have it. We should do so in a way that explains how it can be that such warrant sometimes persists, even where there is counter-indicating evidence from the third-person point of view—as in cases where one convincingly deceives others about one's own experience.

In meeting this challenge it seems we should avoid saying that third-person evidence is *never* good enough to override whatever presumption there may be in favor of the accuracy of first-person belief. We may or may not think it is possible to be discovered in self-deception about the character of one's own phenomenal experience. Perhaps cases where one feels angry, but sincerely denies what one feels is anger, furnish instances of this. But in any event, I think we should allow that the desire to see oneself in some ways rather than others, inadequate care in self-description, and lack of awareness of false implications of one's claims, can sometimes mislead even judgments about one's own current experience. It seems one may, on reconsideration, be led to revise one's first-person attributions of experience. I may say I feel pain, but then, when challenged ("Oh come on, that can't really have hurt"), revise my claim saying, "Well, ok, the feeling is just mildly uncomfortable, not downright painful." And a student of philosophy might think he has introspective knowledge that he is immediately perceiving a red round sense-datum. But when led to reflect on the notion of a sense-datum, and the consequences of this claim, he decides it was false, and in fact he has never perceived a *sense-datum* in his whole life.

So a satisfactory answer to the question "how do we know our own minds?" should meet this challenge. It should provide an account of first-person warrant that explains how one can knowingly deceive others about one's experience. But it should not rule out the possibility that evidence may sometimes help defeat a particular first-person belief about current experience.

2

I want to propose an account of first-person knowledge of experience based on a certain conception of the phenomenal unity of experience. This calls for some explanation of what I mean by 'phenomenal unity'—hardly an obvious notion, and one that requires in turn some account of what I mean by 'phenomenal consciousness.' I have elsewhere undertaken to explain in some detail what I mean by that phrase.[2] Here I'll simply have to state some assumptions, some of them controversial, and hope they are clear enough to get my proposal about self-knowledge off the ground.

Phenomenal consciousness is a feature we can know with first-person warrant to be shared by episodes of silent speech, imagery generally, and sense-experience. One can distinguish it from other features by considering these conspicuous examples of its occurrence, in contrast with hypothetical scenarios in which certain of its instances would be missing, while other things nor-

mally found with these remain in place. (For example: you can conceive of a loss of conscious visual experience in part of your visual field—no stimuli in that region *look* any way to you—even though visual stimuli in this "blind" field cause you to make true, spontaneous, verbalizable *judgments* about them.) By contrasting positive examples drawn from our own experience, sharing the feature of phenomenal consciousness, with imaginary cases in which certain such experiences are missing—as in the "blindsight" just mentioned—we can, in thought, tease consciousness apart from its usual concomitants. To focus our theoretical attention fully adequately on phenomenal consciousness through such thought-experiments requires much more elaboration. Here I will have to refer the reader to my detailed treatment of these issues elsewhere, and trust that what I say now is enough to frame the current discussion. But it is essential here to add, that while sense-experience and imagery provide the least controversial initial illustrations of phenomenal consciousness, this feature is not, on my view, confined to such cases. For example: occurrent *thinking*—even when this is distinguished from imagery—is also phenomenally conscious.[3]

It is important also to recognize that experiences that share the feature of being phenomenally conscious also can be known with first-person warrant to *differ* in ways that only what is phenomenally conscious can differ: these are differences in the way it seems to have these experiences, differences in their *phenomenal character*. So, for example, the way it seems to me to feel pain differs from the way it seems to me to feel thirsty. And the way it seems to me to see the duck-rabbit as a duck differs from the way it seems to me to see it as a rabbit. And the way it seems to me to understand the remark, "He brought some coke to the party" in one manner (as about illicit drugs) differs from the way it seems to me to understand it in another (as about soda pop).

When I say it seems to me a certain way to feel pain or thirst, I don't mean that I *think that* I'm in pain or thirsty. Phenomenal consciousness is to be distinguished from the having of "higher-order thoughts." Furthermore, we should distinguish phenomenal consciousness from some putative "inner sense" that bears to higher-order judgment something like the relation visual appearances bear to visual belief or judgment. For on my view—and this will be important to my account of first-person knowledge—there are no such "inner" sensings. That is to say, there is no way in which your experiences themselves appear to you, accurately or inaccurately, as distinct from the way you judge them—truly or falsely—to be. [4]

Based on this conception of consciousness and phenomenal character, I will now begin to explain what I mean by 'phenomenal unity.' It is an aspect of the phenomenal character of experiences—of the way it seems to have them—that they are unified in a certain way. For example: normally, when it looks to me as if there's a spot on the left, and at the same time it looks to me as if there's a spot on the right, I not only can issue two judgments, one attributing the first, the other attributing the second of these two experiences. Also: the way it seems to have these two experiences, thus distinguished, so relates them

that I am also able to correctly make a judgment, combining and comparing them, such as this: *It looks to me as if there is a spot on the left brighter than one on the right*.

This notion of phenomenal unity can be brought out a bit more, if we contrast the example just given with a similar—apparently conceivable—case in which phenomenal unity of this sort is lacking. Suppose that in a split brain subject, each of his cerebral hemispheres simultaneously generates phenomenal experience: the right hemisphere a visual experience of a spot on the left, and the left hemisphere a visual experience of a spot on the right. Arguably, in such a case each hemisphere would affect the movements of a single human body in such a way as to make these legitimately regarded as the actions of *one person*, and the experiences of each hemisphere would both be experiences belonging to *that* person. However, the way it seems to have these two experiences would not relate them so as to enable the person to judge: "It looks to me as if there is a spot on the left brighter than a spot on the right." In this case, we may suppose, the two visual experiences, though simultaneously had by one person, are not phenomenally unified with one another.

This is not to say that you can make no judgments comparing your experience to someone else's—that is, to an experience that is not yours, and is not phenomenally unified with your own. I may think that the spot on the left looks brighter to *me* than it does to *you*. But this does not entail that my experience of the spot is phenomenally unified with yours, or that your experience is phenomenally unified with my thought, as *my* experience *is* thus unified with my thought.

One might suppose that here what differentiates the comparative judgments you can make about your own and the kind you can make about another's experience is just that the former are made on the basis of some kind of direct, non-observational link between the comparative thought and the experience compared, whereas the latter depend on some sort of less direct, observationally mediated connection with another person. However, we should note: not just any kind of direct, non-observational link will guarantee phenomenal unity. I might, in some sci-fi fantasy, imagine that I have a thought comparing my experience with yours, as a result of some connection neuroscientists have engineered between my brain and yours, a link unmediated by my observation of you or your brain. But this wouldn't necessarily secure phenomenal unity. Maybe, via some neuroengineered link, your experience could cause me to be struck by the thought that the light looks brighter to you than it does to me, or that you feel more pain than I do. But this doesn't entail that my visual experience or my feeling of pain would be phenomenally unified with yours. A crucial additional point: if two experiences are phenomenally unified, then the person who has one of them also has the other. If it seems to me as it does to have two unified experiences, both experiences must be mine. However, again, it is not clear the entailment runs in the opposite direction. For the split brain example above illustrates that we can apparently conceive of a

situation we might want to describe as one in which two experiences belong to one person (even simultaneously), without their being phenomenally unified.

In the initial example, the comparative first-person thought is also phenomenally unified with both of the unified visual experiences it is about. But it is not essential that two experiences be objects of such a higher-order thought for them to be unified with one another. It's not even necessary that the one who has the experiences be capable of forming such thoughts at all. To suppose otherwise would lead into an unacceptable infinite regress.

For first: there is no warrant for maintaining we have *nonconscious* higher-order thoughts of the sort supposedly required to do the job of unification, or that we can have such thoughts nonconsciously, when we don't have them consciously. It would be a mistake to suppose that my ability to report my experiences gives me warrant for maintaining that when I'm not consciously thinking about them, I must be doing so nonconsciously—at most it would be warranted to posit a dispositional state—a belief—here. And there is no reason to suppose that such dispositions are capacities to think nonconscious thoughts only. To the extent we are justified in positing higher-order beliefs where we have an ability to report on our experiences, these will be dispositions to have conscious higher-order thoughts about them.

So then, if we say phenomenal unity of experiences requires some actual or potential unifying higher-order thought that "knits them together" so to speak, this will be an actual or potential conscious thought. Now consider the phenomenal unity of one part of a temporally extended conscious higher-order thought with another part of it. That thought will presumably make the experiences it is about phenomenally unified, only if *it* is phenomenally unified as well. For, I take it, two phenomenally disunified states cannot constitute a single thought—higher-order or otherwise. (I would say, for instance, that my consciously thinking to myself, "If Bernie's drunk…", and your consciously thinking to yourself the very next moment, "…then he's not invited," will not together constitute a single occurrence of thinking: "If Bernie's drunk, then he's not invited"—because my thinking the antecedent is not phenomenally unified with your thinking the consequent.)[5] But then if *all* conscious experiences need to get their unity from a conscious, unifying (but also *unified*) higher-order thought, we would have to say what would make one part of such a thought phenomenally unified with another is also either an actual or potential conscious thought of a *yet higher order*. And in that case we have a regress, either of actual phenomenal thought of ever-higher orders, or of capacities to entertain such thoughts. And neither result is acceptable.

I have spoken of phenomenal unity in a way that may suggest some activity of synthesis—a "knitting together" of initially distinct elements: experiences. But I wish to remain aloof from this idea, and the conception of individuating experiences it seems to involve. I don't even assume that there is, in principle, some natural and unique division to be made of the precise number of experiences a person has over a given time, much less that there is

some temporal process whereby already extant but still disparate experiences are subsequently joined in phenomenal unity. And I am quite willing to entertain the idea that experiences can have other experiences as parts. Also, I do not require that for two of my experiences to be distinct, I must think or be aware of them as such, or attend to their distinctness.

However, I believe it is reasonable to count simultaneous visual experiences of different regions of space as distinct (as in my two spots example), and so similarly, I would count *its sounding to me as if a car is approaching from behind* as a distinct experience from *its sounding to me* (at the same time) *as if someone is speaking from in front of me*. Also, in my book, if experiences e1 and e2 are correctly classifiable as belonging to distinct sensory or imagery modalities, then they are distinct. (Even if my singing causes it to look to a nearby synaesthete as if a cloud of purple streams from my mouth, I would say *its looking this way to her* (that particular experience) is not identical *to its sounding to her as it does*.) And I would say, if I can distinguish different *temporal phases* of an experience—such as different temporal parts of an experience of silently uttering a sentence to myself, or of humming a tune in my head—then these can be counted as distinct experiences.

Experiences distinguishable in these ways, I want to say, may be phenomenally unified. And so phenomenal unity, as I understand it, normally relates the constituent experiences of a single visual "field" to one another, as well as those making up a single temporal "stream" of thought or imagery, and it joins field and stream, so that a person who has any one of these experiences also has any of those thus related to it.

Pulling together these remarks, I would characterize phenomenal unity as follows. Phenomenal unity is that relation U such that:

1) Necessarily, U holds only between phenomenally conscious experiences, and is an aspect of their phenomenal character. (There is a way it seems to have U-related experiences, to the person who has them.)
2) That U holds (that it seems to x the way it does to have U-related experiences), is first-person knowable.
3) It is not necessary that, if x has two U-related experiences, x can make first-person attributions of both (this is shown by the regress argument).
4) It is not necessary that, if x can make non-observationally mediated attributions of experiences e1 and e2, e1 and e2 are U-related (as is illustrated by the hypothetical case of interpersonal brain-link).
5) It is not necessary that, if, at time t, x has e1 and e2, then e1 and e2 are U-related (as is illustrated by the split-brain example).
However:
6) Necessarily, if e1 and e2 are U-related, and x has e1, then x has e2.
7) It is normally[6] the case that, if it seems to x as it does to have two simultaneous visual (or aural, or tactile) experiences—one of an object on the left, one of an object on the right—*and* x makes first-person comparative attributions of these experiences, then:

(a) U obtains between these two experiences; and
(b) U obtains between each of these experiences and the phenomenally conscious first-person thought comparing them.

8) Necessarily, U obtains between the temporal parts of any single phenomenally conscious thought.

I do not pretend that this is all that needs to be said to explicate the notion of phenomenal unity fully. I say just that it gives enough substance to that notion, to make the account of self-knowledge I will base on it worth taking seriously.[7]

3

I start with this observation. Often the first-person attributions of experience we make, or are inclined to make, are non-inferential in the following sense. Our being disposed to make them does not depend on our being inclined also to offer some further claims that we would take to justify our attribution, by constituting a non-circular argument for them. My dispositions to judge sincerely that I feel pain, or that it looks to me as if Sally is taller than Jane, do not depend on my being disposed to make *other* claims, which I would take to give me enough reasons or evidence to infer that the above judgments about my experience are true. Thus:

(1) I am disposed to make certain present-tense first-person attributions of experience non-inferentially.

Now, in such cases, when I can without inference attribute experience to myself using a first-person pronoun in the normal way ('I feel pain'), I can also report the occurrence of experience in a corresponding demonstrative way: '*This* is a feeling of pain'. A normal subject can think, not just "It looks to me as if there's a spot on my left," she can also think a thought about the same experience thus reported, whose expression would require some demonstrative reference to it, such as, '*This* is an appearance of a spot on my left' or, '*This* is seeming to see a spot on my left' or, '*This* is a visual experience of a spot on my left' or, '*This* is its looking as if there's a spot on my left.' These demonstrative reports *correspond* to certain first-person attributions, in that, if the first-person attribution is true, then so is the demonstrative experience report, and the attributed experience in the former is identical to the experience referred to in the latter.

Some (e.g., Sydney Shoemaker) would say that such demonstrative reference to experience is impossible.[8] I would disagree. But I would agree—and this is crucial—that one's understanding of such reference is not secured via some kind of "inner appearance" of the experience referred to (since I believe there is no such phenomenon). What makes understanding such reference possible, I believe, is not the "inner" perception of "inner" objects, but rather the

fact that the demonstratively expressible thought is phenomenally unified with the experience referred to in its expression, and this thought involves a direction of attention to the experience it is about. This is not just the same as attending to objects of experience (though perhaps one cannot attend to the experience without attending *also* to its objects). For example, I can attend to (and think about) the way things *look* to me, where that isn't simply a matter of paying attention to the *things* that look that way to me. Consider the type of attention involved in trying to make an accurate perspectival drawing of a still-life scene of vegetables on a table. This differs from the exercise of attention involved when one looks at the vegetables in the course of preparing them for cooking. In the former case, I would say, you attend to the way the vegetables look to you, while in the latter case you attend to the vegetables by looking at them. And this difference is manifest in the phenomenal character of the experiences. The way it seems to me to pay attention to the *look* of the vegetables from where I stand differs from the way it seems to me to focus my attention on the objects seen.

Thus:

(2) Normally, corresponding to first-person attributions of experience I am disposed to make non-inferentially are demonstrative experience reports that I am also able to make.

I want now to consider what is involved in my ability to understand the demonstrative reports of which I am capable, corresponding in this way to first-person attributions of experience. First of all, I want to introduce the premise that I do frequently understand reports of that type. I believe this is justified, since I think we can recognize that we have, in a certain sense, presumptive warrant for believing that we understand what we're saying. That is, if we think we understand something we say, and we have no reason for thinking we *don't*, then we have warrant for thinking we *do*. To suppose otherwise would, it appears, invite regress problems. (If, to be entitled to think I understand anything I say, I have to be able to give reasons for thinking I understand it, then I will have to be able to give reasons for thinking I understand the statement of reasons—and so on, *ad infinitum*.) Applying this to the case of demonstrative experience reports that correspond to the first-person attributions to which I am inclined:

(3) Since often I believe I understand a given such demonstrative experience report that I am able to make, and have no warrant for thinking I *don't* understand it, I frequently have warrant for thinking I *do* understand it.

Next we need to think about how these first-person attributions and the corresponding experience reports are bound up with the phenomenal unity of experience. Consider again: it seems conceivable that there could be a link

engineered between my brain and yours, such that, when it looks to you as if there's a spot on your left, I am caused to think, without observing you, "It looks to her as if there's a spot on her left," even though my thought is not phenomenally unified with your experience. We might call this a "non-observational link between thought and *experience without phenomenal unity*." But then, it also seems conceivable that a similar link could be engineered between the hemispheres of a split brainer. If one person's brain could have a non-observational access without phenomenal unity to another's, why could there not be such a link between one *half* of a brain and the other half of it? The split brainer's left hemisphere could generate the thought: "In my right hemisphere it looks to me as if there's a spot on the left," even though there is no more *phenomenal* unity between this left hemisphere thought and the right hemisphere visual experience that makes it true, than there is between my thought about your experience, and your experience, in the science fiction scenario.

So I don't want to deny that there could be cases in which one had non-observational access to one's own current experience, but without phenomenal unity. But now: is it perhaps true that (contrary to what I think) my own actual experiences that I am disposed to attribute to myself are not usually such as would be phenomenally unified with the thoughts I would express in corresponding demonstrative reports? Is it perhaps true, not just in weird or hypothetical cases, but usually, actually, that such corresponding demonstrative experience reports as I can make would express thoughts that are not unified with the experiences they are about? If that were the case, then it seems to me, I would really have no grip on the notion of phenomenal unity at all. That is, if phenomenal unity is not something I can reliably find joining such demonstratively expressible thoughts with the experiences they are about, then I don't know at all what it is. So assuming I do have some understanding of this notion of phenomenal unity, then the following holds. In all or almost all cases of demonstrative experience reports I *can* make that correspond to the first-person attributions to which I am disposed, if I *did make* the reports, *and* there are experiences to which they would refer, the experiences would be phenomenally unified with the thoughts expressed in the reports. I would say, not just that it rarely or never happens that such thoughts of mine would be disunified from the experience they are about, but that necessarily, if I have a grasp of the notion of phenomenal unity, this is so.

So now I will confine my attention to these cases of these demonstrative experience reports that are not weirdly dissociated from the experience, if any, they are about. I want to focus on the demonstrative experience reports I can make, which not only correspond to my first-person attributions, but are also such that, *if* they do refer to experiences, they refer only to experiences phenomenally unified with the thoughts they express. To say that a report is of this class is not to say that there is in fact a referent answering to it. The point is just that I am now setting aside whatever marginal cases there might be of corresponding demonstrative experience reports that express thoughts phenom-

enally dissociated with the experiences they are about, if any. Now I want to ask, regarding the remainder: if in fact there were no referent answering to one of these reports, could I still *understand* my use of the demonstrative in it? Could such reports, in this sense, suffer from an intelligible *failure of reference*?

To answer this, we should consider some ways in which other, more ordinary uses of demonstratives—uses where one intends to refer, not to experiences, but to objects currently in one's spatial environment—can (at least apparently) suffer such failures of reference. Arguably, failures of reference can occur where one's demonstrative report is based on inaccurate or nonveridical sensory appearances. I say, for example, 'This is pink' and understand myself to assert something, even though as it happens, there is nothing that my utterance of 'this' refers to because I'm suffering from withdrawal and hallucinating a little pink rhinoceros. But notice here, for me to intend to refer to something, and understand my use of the referring term in question, I need at least to have the (non-veridical) appearance of an object to refer to.

Might I fail to refer to an object, though intelligibly intending to use a demonstrative to refer to something in my environment, but *without* hallucinating? Perhaps accurate sensory appearances, combined with relevant delusions could underlie failures of demonstrative reference. I might, pointing at a region in which nothing at all appears to me, intelligibly and sincerely say something like: 'This is my invisible dog.'[9] But if I really do intend to refer to something in this way, there needs at least to appear to me as if there's a *place* for the purported object of demonstrative reference to occupy.

Now notice: neither these types of reference failure, nor anything analogous to them, is available where the relevant demonstrative *experience* reports are concerned. For first, here one cannot hallucinate a referent, since experiences don't *appear* to us at all, veridically or non-veridically: there is no inner sense, as distinct from first-person thought or judgment. So it can't falsely appear to me as if there's an experience in some place where there isn't one, the way it can look to me as if a little pink rhinoceros is cantering across the carpet. Secondly, ordinarily at least, I do not understand my demonstrative reference to experience by its appearing to me that there's a place, in which I suppose an experience (as opposed to what it's an experience *of*) to be located. Though I may believe a given experience of mine to occur in a (vaguely identified) place (I may think my thoughts occur in my head, and may think a feeling of pain or nausea is in my gut), I do not use my perception of my own body to distinguish a demonstrative reference to my *experience*, from a reference to what I take it to be experience *of*. For neither my perception of where my head (literally) is, nor my belief that my thoughts occur there, is essential to my demonstrative reference to my own thinking. And if my feeling of pain does not really occur in the part of my body that I feel so painfully (my gut, my foot, my chest), this does no damage to my reference when I say to myself: "*This* is a painful feeling." The fact that there is nothing I

refer to in the place to which I point when I say 'This is my invisible dog,' suffices to make my reference fail altogether. But suppose there is nothing at all that I refer to in the place I indicate by gesturing where I believe my foot is, when I say, 'This is a painful feeling' (maybe I am having a "phantom foot" experience). That is not enough to entail that there is nothing I refer to at all with my utterance of 'this.'

But there is a third way in which a perceptually based demonstrative utterance might fail to meet with a referent. It seems we might remove sensory appearance from a role in guiding a speaker's understanding of a demonstrative, if we turn our attention to a kind of judgment that qualifies as perceptual, on account of its causes, but from which perceptual appearance or experience has been subtracted. Consider a kind of "blindsight" judgment of the sort earlier alluded to briefly: there is no way stimuli in part of your visual field *look* to you, and yet the effect of certain types of those stimuli is to generate reliably accurate spontaneous judgments on your part that those types are there. Mightn't one express such blindsight judgment demonstratively, pointing into one's blind field and declaring 'That is a circle'? If so, then we may suppose that on some occasion this judgment is faulty (it's reliable, not infallible): one judges 'That is a circle' because of some activation of visual pathways, though there is in fact no circle indicated, nor anything at all rightly interpreted as referent of one's utterance of 'that.' Still, we may want to say the speaker understands what he means when he says, 'That is a circle.'

But even if we allow such a judgment to pass as genuinely demonstrative, still, if the speaker's understanding of 'that' is guided by a belief about the spatial location of its referent, it will not provide us with a model we can use to make sense of referent-less demonstrative experience reports. For we have seen that no such "locative" belief is essential to securing one's understanding of what one is referring to in demonstrative experience reports.

However, we may suppose that, when the blindsighter says, 'That is a circle' he understands his demonstrative in another way—by thinking of the referent of 'that' *as the circle that causes him to judge now that there is a circle*. Might I similarly suppose that when I think 'This is a painful feeling,' I think of the referent of 'this' as something like: *the painful feeling that causes me to judge now that there is a painful feeling*? If so, then perhaps we can make sense of referent-less demonstrative experience reports. For just as the blindsighter's judgment might not be caused by what he thinks it is caused by (and so his reference may fail), so my judgment of feeling may not be caused by a feeling of pain, as I think, and so my utterance of 'this' in 'This is a painful feeling,' though intelligible to me, may find no referent.

The problem here is that even if I do think of the referent of 'this' in 'This is a painful feeling' as the cause of my thought, my thinking of it in this way is not essential to my understanding of the relevant demonstrative. For I might also *not* think of the referent as cause of my thought. I might become a

convinced epiphenomenalist. And however wrong that doctrine might be, my holding it will not prevent me from understanding my reference when I say 'This is painful.' More plausibly (to my mind): I might deny that the referent of my demonstrative causes my thought about it—not on the grounds of epiphenomenalism—but rather, on the grounds that the referent of 'this,' the experience I refer to, and my thinking that this is a painful feeling, are *not wholly distinct events*, such as may be causally related. I would say the occurrence of my thinking that this is a painful feeling could not have happened without the experience it is about, for that experience is itself a constituent of the event which is my thinking—the thinking about the experience is not an event separable from the experience thought about. (As for the thought that *there is* a painful feeling, *that* thought need not occur to me at all, though I am prepared to recognize it as entailed by the (demonstrative) thought that does occur to me.)

One might try to counter this with the proposal that what is essential to the reference is not that I *think* of the referent as the cause, but that it actually *be* the cause of my thought that this is a painful feeling. Then one reasons: "If it does cause my thought, then, since an effect can conceivably occur without its cause, the thought could occur without the referent. And if that is so, the failure of such demonstrative reference is intelligible." But this won't do. The question is how I can conceive of the thought I express with 'This is a painful feeling'—the *thought* and not just a (real or imagined) utterance of this sentence—as something that might occur in the absence of a referent for 'this.' It does not help me find a way to conceive of this, to say: "The thought is an effect of the referent only if it could conceivably occur in the absence of its referent." If that is true, it simply shows that our difficulty in conceiving of the thought occurring in the absence of the referent is also a difficulty in conceiving of the thought as an effect of its referent. If we assume effects must be separable from their causes, and we have not yet found a way of making sense of intelligible, though failed, reference in the case of our demonstrative experience reports, then we should not affirm that the referent of such a report does cause the thought expressed in the report.

I conclude that the kinds of present tense demonstrative experience reports under consideration cannot intelligibly suffer from the reference failure that apparently can afflict analogous demonstrative assertions we take to be about things other than experiences, located in space. For I cannot "locate" the purported referent by "hallucinating" an experience: since experiences do not appear to us, we cannot hallucinate them. And I do not understand my reference to experience by either perceiving or conceiving of a place, which I suppose an experience to occupy (so that such reference will fail if there is no referent in that place). For even if there is nothing I refer to which is *in* such a place, this does not prevent there being an experience I refer to. Further, I do not understand the reference by supposing the experience to be the cause of my thought about it. For even if I did *not* suppose the experience to be such a cause, reference would not fail. Finally, it will not work to argue

that, to succeed in referring, my referential thought needs a referent as cause (whether I think so or not), and effects must be separable from causes. For until I have some way of rendering reference failure intelligible in the case of these experience reports, the principle that effects must be separable from their causes simply prevents me from affirming the premise: "My reference will succeed only if my demonstrative thought is caused by its referent."

Suppose we agree then: none of these ways in which demonstrative reports about one's surroundings can suffer intelligible reference failure, nor anything similar, can bring this to pass in the case of demonstrative experience reports. Still, mightn't one suggest that perhaps there is some altogether different way in which these may intelligibly lack a referent? Maybe in fact, this will seem easy to produce, once we allow (as I do) that there can be sincere, but false first-person attributions of experience. Suppose I mistakenly characterize some current feeling as painful, in such an attribution. Then won't there be a corresponding experience report, containing a use of a demonstrative I can understand: 'That is painful'? (Or more colloquially, 'That hurts!') But then, if I am not feeling pain, won't I use the demonstrative vacuously, without a referent, though not without understanding?

However, first, it is not clear that, in such instances, my use of the demonstrative fails to find a referent. By uttering 'that,' I would say, I still refer to some feeling I have, even if I misdescribe it as painful. What we need is a credible example of error in first-person attribution that prevents one from even referring to one's experience in the corresponding demonstrative report. And I don't see that there is one. We will be inclined to think there is, I believe, only if we imagine that one's understanding of the demonstrative in the experience report can be exhaustively characterized by forming a description from the first-person attribution to which it corresponds. So, only if we suppose that my way of understanding my use of 'that' in 'That is painful' can be expressed as: 'The painful feeling that I have,' or some such, may it seem that the failure of the description will imply a failure of the demonstrative reference.

But we can see that such descriptions fail to capture such understanding of the demonstrative as we can have in ordinary cases. For consider again the hypothetical case earlier described, as "non-observational access without phenomenal unity" in a split-brainer. Such split-brainers would make (with their left-brains) a first-person attribution, 'In my right brain it looks to me as if there's a spot on my left,' and I suppose nothing prevents them from making corresponding demonstrative reports, '*That* is its looking to me as if there's a spot on my left,' in which the demonstrative refers to their right-brain experiences. But in this case, their understanding of the demonstrative *is* correctly glossed by some description such as, 'The experience (in my right brain) of its looking to me as if...etc.' That is how they understand the demonstrative reference in the report 'That is its looking to me...etc.'

But it is clear I think, that there is a difference between the way the demonstrative would be understood in such cases, and the way in which one would understand which experience one was referring to in ordinary cases, where one

constructed demonstrative experience reports corresponding to one's first-person attributions. The way I would ordinarily understand which experience I was referring to when I said, '*This* is its looking to me as if there is a spot on my left,' differs from the way the hypothetical split-brainers would understand their references to their experience. So, I do not understand my reference, when I say, 'This is its looking to me...etc.' by thinking of the referent simply as 'the experience of its looking to me...' or 'the visual experience I have of...'. Otherwise, there would be no difference between my way of understanding which experience I refer to in such cases, and the way the imaginary split-brainers understand which they refer to. And there would be a difference.

But what is the difference in understanding? There is nowhere for it to lie, but in the crucial stipulated difference between myself and these hypothetical subjects. Unlike them, I understand my reference by attending to the experience referred to, in such a way that my thought about the experience is phenomenally unified with the experience it is about. The split-brainer, by hypothesis, cannot do this, because the corresponding demonstrative thoughts they would express are phenomenally dissociated from the experience they refer to. That is why they can only understand their reference by means of some description like "The experience of...I am now having."

Notice here that the point is not that I understand which experience I am referring to by thinking of the referent as 'The such and such experience that is phenomenally unified with my thought that I have such and such experience.' It is not by *thinking* of the experience *as* the experience phenomenally unified with my thought about it, that I understand what experience I refer to. No, it is rather by *attending* to the experience that *is* phenomenally unified with the thought, that I understand my reference. There is a difference, because otherwise we would have to say that those who had not grasped the concept of phenomenal unity to the extent needed to make judgments about the phenomenal unity of their experience would be unable to understand their use of demonstratives in experience reports corresponding to their first-person attributions, in anything other than the descriptive manner characteristic of the split-brainers in our story. But this is unacceptable.

But now, once we grant that the way we would ordinarily use for understanding demonstrative experience reports involves attending to an experience phenomenally unified with the thought expressed, we will see that what is peculiar to understanding such demonstrative reference is incompatible with intelligible reference failure. To see this, note the contrast with cases where sensory appearance underlies one's understanding of demonstrative reference. It can be said consistently: "Even if there is no object I refer to with it, I can still understand my use of a demonstrative, as long as it appears to me as if an object is there for me to refer to." On the other hand, it cannot be said consistently: "Even if there is no object referred to with it, I can still understand my use of a demonstrative in an experience report, as long as I attend to an experience phenomenally unified with the thought expressed in the report." For if

I attend to an experience phenomenally unified with the thought expressed in the report, then there *is* an object referred to. Thus, insofar as there do seem to be conditions peculiar to understanding demonstrative reference to one's experience, these offer us no way of making sense of reference failure.

What have we found? Understanding the use of demonstratives in reports of experience corresponding to first-person attributions of experience cannot suffer from reference failure in anything like the way understanding intended reference to items in one's surroundings can. Further, what is, in the ordinary case, peculiar to the former also cannot support an understanding of the relevant demonstratives in the face of such failure. Thus I conclude that such a demonstrative experience report, if I understand it, cannot be missing a referent. In short, no *likeness* between demonstratively referring to things in one's surroundings on the one hand, and to experience on the other, renders the latter intelligible when referent-less, and also nothing that is ordinarily *distinctive* of it can do this. Therefore, in ordinary cases, we simply cannot understand these demonstrative experience reports, if they lack a referent. So, if (as usually I am warranted in thinking) I understand such a demonstrative experience report, then *there is a referent for the demonstrative*: here reference would not fail, it would succeed. When I add this conclusion to the earlier point—that typically, if there is a referent for the demonstrative of which I am capable, then it is *phenomenally unified* with the thought expressed with the report—I arrive at the following:

(4) I have warrant for thinking, in the case of many or all such demonstrative experience reports I am able to make, that there is a referent of its demonstrative, which would be phenomenally unified with thought expressed by the report.

Now we need to turn our attention to a different form of error, other than reference failure—what we might call *predicative* failure. How susceptible are demonstrative experience reports to *this*? Again I propose to approach this by looking at ways in which demonstrative reports may suffer this shortcoming, when one intends to refer, not to an experience, but to something in one's surroundings.

It may happen that there is something I demonstratively refer to, all right, and I understand what I'm referring to—but my predication is not true of it. 'This is a tree,' I say, and understand what I'm saying, even if the situation is such that what I indicate by 'this' is not in fact a tree. And perhaps *no general* term I am disposed to predicate of the referent of 'this' is in fact true of it—perhaps the referent is merely a holographic image of a tree, which I mistake for a tree.[10] Yet this will not prevent me from understanding which item I refer to by means of 'this.' If this kind of thorough-going predicative error is possible, it is only because I have a way of understanding what I'm referring to that does not require I correctly characterize it in general terms. This

is a way of understanding demonstrative reference that doesn't involve correctly *conceptualizing* or *classifying* the referent as of a certain *kind*, but rather, *locating it in a particular space*.

However, neither this, nor any analogous way of identifying the referent of a demonstrative is available, where that referent is an experience referred to in a demonstrative report correlative to a normal, non-inferential first-person attribution of experience. For I just don't have in such cases a non-general way of thinking about the referent, the experience, without correctly classifying it as of some kind or other. I cannot understand what I'm referring to when I refer to an experience by relying just on my perception of particular spatial locations. And there seems to be nothing *like* our non-general perception of space—a nonconceptual "inner sensory" representation—whereby we can represent our experience to ourselves, and thereby "locate" the referents of our demonstrative experience reports.

But, one may ask, why should we think my reference to my experience *needs* something like the nonconceptual means of representation employed by sensory perception to identify a referent that one can completely mischaracterize in general terms? Even if I do not represent my experience as occupying a particular "location" in "inner space," why can I not simply direct my attention to a particular experience, and thereby understand myself to be thinking of *that* one, even though, as it happens, I am disposed to classify it correctly *in no way whatsover*?

Again consider: in the case of demonstrative reports about things in my surroundings, it seems that, if it is possible for me to be *so* mistaken in the way I conceptualize a given demonstrative referent that *no* way in which I am inclined to classify it in general terms is correct, that is only because I have some way of understanding what I am referring to that does not require an ability to classify or conceptualize it correctly in general terms. And this involves its appearing to me as if what I refer to is located in a place, or my thinking of it as located in certain place. Of course I can and do understand what I'm referring to by directing my attention to it. But attention does not somehow operate alone here to help me understand what I'm referring to, without working through some form of representation. It's not as if I can understand my demonstrative reference to some spatial object just by directing some "pure" act of attention upon it, independently of representing it *as somewhere*. If "pure attention" separated from any form of representation would not be enough to secure an understanding of one's reference to objects in one's surroundings, then it cannot do this in the case of reference to one's experience either, *unless* there is something special about the way attention works in the experience case, which enables it to perform this feat. Is there?

It's true that, on my view, one can attend to one's experience without representing it to oneself, truly or falsely, accurately or inaccurately, in any way. And attending to one's experience is essential to understanding demonstrative reference to it, in the usual case. But it doesn't follow that one can, by attend-

ing to one's experience, understand one's reference to it, even though one is not inclined to think anything true *of* it.

Perhaps it doesn't *follow*, but still we may ask: can it happen? To answer this question, I see nothing to do but to try to conceive of a situation in which, though one understands well enough which of one's experiences one demonstratively refers to, still no classification whatsoever that one is inclined to give of that experience is in fact true of it. If I can do this, I should be able to conceive of thinking that some demonstratively identified experience I take myself to have belongs only to kinds, to which in fact it doesn't belong. But when I consider examples, I find I simply cannot do this.

First: can I imagine thinking that some demonstratively identified experience that I take to be my *visual experience of a red square*, is (not that, but instead actually) a *sensation of nausea*, or an *experience of smelling ground coffee*, or *of feeling silk fabric by running my hand along it*, or…? I can no more do this than I can imagine thinking that a demonstratively identified object that I take to be my coffee cup is actually not that, but really: *an ocean liner*, or *a subatomic particle*, or *a session of Congress*. (I can of course, say, point at a Giant Sequoia and say *words* like, 'I imagine thinking that *this* is my coffee cup' (or whatever)—but that is not enough for me really to conceive of thinking the *thought*.) The point here is not that, in such a situation, 'This is my coffee cup' is meaningless or unintelligible. It is intelligible, and false. The point is that I cannot conceive of asserting sincerely what I mean by that utterance. For I cannot conceive of how one could do that, and still understand that utterance to mean what I do by it.

Now the crucial point about experience is this. Not only can I not conceive of thinking that some demonstratively identified experience I take myself to have, which in fact belongs to one kind (say a feeling of itching), belongs instead to some other kind (say a smell of frying butter) to which it does not belong. My claim is that I cannot take the terms I actually use to describe my experiences, and conceive of employing them in thinking of some demonstratively identified experience I take myself to have, while being disposed to apply to it *only* those terms from this lexicon that are *false* of it and *none* that are *true* of it—at least, I cannot do this, as long as I understand what I am saying about it. I cannot conceive of making a general predicative error this massive regarding an experience I can think of demonstratively and believe I have, while still understanding what I am saying. By contrast, arguably, I *can* conceive of making a general predicative error this extensive, regarding something I can think of demonstratively that is spatially located, and which is not an experience, while still understanding what I am asserting of it.

I believe this needs more looking into—more than I can offer in this summary exposition of my approach to self-knowledge. But let's suppose that my efforts to conceive of the relevant errors have been searching enough to give me more warrant for the view that such massive predicative errors regarding experience are not conceivable than that they are. Then I can reasonably con-

clude: neither through some nonconceptual form of representation, nor through a mere exercise of attention, could I understand my demonstrative reference to my experience in the face of a complete failure to classify the referent in any way correctly in general terms. Neither some similarity with the way I understand reference to objects in my surroundings nor some peculiarity distinguishing that from reference to experience will enable me to understand my successful demonstrative reference to my experience, when predicatively I am a total failure.

The conclusion I draw is: if my demonstrative experience report is intelligibly to *mis*classify its referent, I must also be disposed to classify it *correctly* somehow. Otherwise, I *literally won't understand what I'm talking about*. So, if I have warrant for thinking I understand what I'm saying when I make a given demonstrative experience report, then there is an experience I am speaking of, and at least *something* general I am inclined to judge about it is correct. Thus:

(5) It is possible for me to understand my reference, if I make such a demonstrative experience report, only if I am disposed to classify its referent correctly somehow.

What now can I conclude from this about the warrant I have for a given classification I am disposed to make in a demonstrative experience report? In a given case, I have warrant for thinking I understand my demonstrative experience report. Then I will also have warrant for holding what I know follows from the truth of this thought. And what follows is that some classification to which I am disposed, of the experience to which I refer, is a correct one. So I have warrant for thinking *some* demonstrative report that I'm inclined to give of that experience is true. Thus:

(6) If (a) I have warrant for believing I understand a given demonstrative experience report I am able to make, then (b) I have warrant for thinking that some way in which I am disposed to classify its referent is correct.
(7) Since (a) is often true, so is (b): often, I *do* have warrant for thinking that *some* way in which I am disposed to classify the referent of such a report is correct.

Now I may ask myself, regarding a given demonstrative experience report that I can make, is there any alternative such report I can make, that I can understand as referring to the same experience, for which I have *more* warrant? If the answer is no, then I have at least as much warrant for the demonstrative experience report in question, as I have for any such report about the same experience. So either: (i) I have warrant for that (initial) report ; or (ii) I have warrant for *no* reports I'm able to make about the same experience. But

(ii) would be so only if, in those situations in which I took myself to know what I was experiencing, I had at most true beliefs, without warrant. However, we ruled that unacceptable back in section 1, as leading to absurd conclusions. So then it follows that I do have warrant for the demonstrative experience report under consideration, again, provided that there is no other alternative report about the same experience for which I have *more* warrant. Thus we may reason as follows:

(8) If I have warrant for believing: (a) *some* way I am disposed to classify the demonstrative referent of a given experience report I am able to make is correct; and (b) there is no way of classifying its referent, other than that involved in this report, to which I am disposed, and for which I have *more* warrant, then I have warrant for this demonstrative experience report.
(9) Often I do have warrant for believing (a) and (b).
(10) Therefore, often I have warrant for believing a demonstrative report of experience that I am able to give, corresponding to a first-person attribution I am disposed to make, is true.

So much for the warrant I have for my demonstrative experience reports. How does this yield warrant for the first-person attributions of experience to which they correspond? First recall that earlier I limited our consideration to typical cases of demonstrative experience reports corresponding to first-person attributions—cases where, if the referent exists at all, it is phenomenally unified with the expressed thought. And phenomenal unity entails that the thinker of the thought has the experience reported. But then, if *I* am the thinker of that thought, I also have the experience that it is about. It follows then that the first-person attribution of experience corresponding to the demonstrative experience report to which I am disposed is true. Thus:

(11) In those cases, the corresponding first-person attribution of experience I am disposed to make is true.

Now, what about my possession of warrant for such attribution? Shall we say that while I have warrant for my true demonstrative experience report, and while the corresponding first-person attribution is true, nevertheless, I lack warrant for that first-person attribution? Only if there is, in such circumstances, something more I could want in the way of warrant for my true belief that *I* am the one who has the experience in question. And what could that be? Is there some kind of evidence or justification that it is appropriate to demand for my true belief that *I* have a particular experience, which I might lack in circumstances in which I do have warrant for believing that *this* experience occurs, and in which I do in fact have that experience? I cannot conceive of any. If I do feel pain, *and* think that I do, *and* my thought is phenomenally

unified with this feeling, *and* I have warrant for thinking (as I am able to think) that *this* is a feeling of pain, well then, there is, it seems to me, nothing *more* I could possibly need, to have warrant for thinking that *I* have this feeling. Under such conditions, there is no evidence or inference on which this belief could be justifiably based, and no further condition one could add which would alter a merely true belief that I feel pain, to a belief I have warrant for holding.[11] And again, it won't do here to suppose that perhaps I simply have true, but quite warrantless belief. We cannot suppose that one lacks warrant for first-person belief in such cases without inviting the absurd conclusion that one cannot knowingly deceive others about one's experience. Thus, finally, I can conclude:

(12) In these cases, I also have warrant for believing the corresponding first-person attribution of experience to which I am disposed is true.

4

I take the foregoing to show that you have warrant for a present-tense first-person judgment attributing experience if:

(i) You are disposed to make it *non-inferentially*.

(ii) There is a corresponding *demonstrative experience report* R you are able to make.

(iii) The experience referred to in R would be *phenomenally unified* with the thought expressed in it.

(iv) You would have warrant for believing you understood R (if you made it).

(v) There is no other report you are able to give upon serious consideration, which you would understand to refer to the *same experience* referred to in R, and for which you have *more* warrant than you have for R.

My claim is that since these conditions often obtain, quite commonly one has the sort of warrant constituted by these conditions for one's first-person present-tense judgments about experience. I don't say that in order to have warrant for first-person attributions you must be able and inclined to offer this argument in justification of them. I maintain that one has the warrant because the argument is available—if one made it, one would state what one's warrant is.

The question remains whether these conditions account for our possession of a *distinctively first-person* warrant for such judgments. I claim that they do. Notice: the route by which we derived this conclusion about one's warrant for first-person judgment will not yield a similar conclusion regarding corresponding *third-person* beliefs. For the conclusion is reached via the premise that the thought about experience is phenomenally unified with the experience it is a thought about. And while thought about one's own current experience is nor-

mally phenomenally unified with the experience it is about, thought about *someone else's* current experience is not. In fact, it is plausible to claim that third-person thought cannot possibly be phenomenally unified with the experience that it is about, for if it were possible, then two persons could have one and the same numerically identical experience. But it is plausible to claim that this is impossible. In any case, even if it were possible for you and I literally to share an experience, the point remains: it still wouldn't follow from the fact that I can demonstratively refer to an experience unified with my demonstrative thought, and correctly classify it, that an attribution I make of this very experience to *someone else* is correct. But it does follow that my attribution of this experience to myself is true.

So we have here an account that states sufficient conditions under which first-person judgments about current experiences enjoy warrant—the type of warrant is specified by these conditions: it is the warrant that follows from *these* conditions. And it is a distinctively first-person warrant, because the conditions suffice to warrant only first-person, not third-person judgments about experiences.

Now we need to see how this account applies to the cases we would like it to cover. How, for example, does it answer the question: how do I know I feel pain? This is the question, what kind of warrant do I have for my first-person judgment that I feel pain? And this is the warrant I have. I am able to think a corresponding demonstratively expressible thought: *This is a painful feeling*. Since, if I exercised this ability, I would believe I understood this thought's expression and would have no reason to think I didn't, I would have warrant for believing I did understand it. And if I would understand this thought's expression, and this is not a weird anomalous split brain case (as it is not), then there is an experience of which I would thus demonstratively think, unified with that thought, and I am able and disposed to classify that experience *somehow* correctly. If there is no *other* classification that I am disposed to apply to the very same experience, and for which I have *more* warrant, then I have warrant for regarding the classification "painful feeling" as correct. And it happens there is no such alternative, more warranted classification. So I have warrant for thinking that this is a painful feeling. And, since I am the thinker of this thought, and the phenomenal unity of my thought with this feeling guarantees that the thought's thinker is also one who has the feeling thought about, I have warrant for thinking *I* feel pain.[12]

Does this proposal also apply to the deceiver's self-knowledge? What it needs to do is explain why what gives one first-person warrant allows such warrant to persist, even in the face of third-person evidence that firmly supports the contrary of the first-person belief—as in cases where one knowingly deceives others about one's experience. And it can do this. For the phony evidence one skillfully provides for others does not override one's warrant for the judgment that one has the experience. The third-person evidence *would* override one's warrant for the first-person judgment to which a demonstrative

experience report R1 corresponds, only if there were an *alternative* report R2 the subject were disposed to give about the very same experience referred to in R1, and for which he had more warrant than R1, because it is supported by that evidence. But, in such cases of knowing deception, there just isn't one. Having thought, of my feeling of back pain, "This is a painful feeling," I just find myself unable to think, *of that very same experience*, "This is not a feeling of pain, but a feeling of thirst." There just is no alternative way I can conceptualize this experience, to which I am inclined, and which is supported by the evidence on offer to others. Now of course it may happen that, having lied that I was painless but thirsty, I might then come to realize—"Come to think of it, I do feel a little thirsty." But even so, I am simply not able to think of that very experience I took to be a *feeling of pain*, as instead a *feeling of thirst*. And since I am unable to do this, the deceitful evidence I provide to others does not give me warrant for thinking it was not deceitful after all, but an accurate reflection of my experience.

Can my account deal with the following objection? "The third-person evidence in such a case might be reasonably taken by the subject to indicate that the conditions alleged to imply first-person warrant do not themselves obtain. But then the deceiver has warrant for believing that he doesn't know what in fact he knows. And this result is unacceptable."

My response is first, the fact that you would express your judgment by a given utterance in a situation where you are aware of evidence that counts against its truth is surely not enough by itself to override the presumption that you *understand* your own utterance—so condition (iv) is not threatened. As for the other conditions I take to entail first-person warrant in a given instance, notice that they are such as enable one also to have warrant for the yet higher-order judgment that those conditions obtain. For, given my account of conditions sufficient to bestow first-person warrant, one can have such warrant for believing that one thinks a thought phenomenally unified with and about an experience (conditions (ii) and (iii)). The phenomenal unity of thought with the experience it is about is as first-person knowable as the fact that the thought and experience occur. Also, try as one might, one cannot find warrant of any sort to suppose one can think of *that very same experience* in terms supported by the deceitful evidence, nor can one find warrant for thinking one's disposition to judge as one does is anything other than non-inferential. And where no warrant is anywhere to be found for *asserting* one can think such a thought, or that one's disposition to judge as one does is explicitly inferentially based, the corresponding *denial* has more warrant. (So conditions (i) and (v) are unthreatened.) Thus, in the deceiver case, the third-person evidence doesn't give me reason to think I don't know what I do know about my experience, by giving me reason to think that the conditions necessary for this knowledge do not obtain. For I have more warrant for believing those very conditions do obtain than that they don't, and my deceptive behavior doesn't give me reason to say otherwise.

Finally, we might ask: does my account leave open the possibility for first-person error, and its correction? It does. For it allows that non-inferential first-person attributions of experience can be mistaken and corrigible. However, on my account such judgments can be justifiedly corrected, *only if* the correction is supported via a support for other re-descriptions of the same experience that one can accept.

The picture I am presenting is this. In order to get you to change your mind legitimately about the character of your own experience, I must respect and support at least some of the judgments you are inclined to make about it. To put it another way: you can perhaps legitimately get me to believe that I am misdescribing my own experience. But you do this only by getting me to accept *re-descriptions* such as I am capable of applying to the very same experience as that I misdescribed. There needs to be a moment, in which, as we might put it, *I recognize my experience in the re-description you offer*.

5

The account offered here of self-knowledge stands in ironic contrast to the more familiar introspectionist, perceptual model that has attracted philosophers over the years. According to that model, what gives knowledge of one's own experience its distinctive character is that one enjoys in one's own case (and *only* in one's own case) something rather similar to the sensory perception that allows one to identify particulars in public space (the "external" or "outer" world) and apply concepts to them—except this subjective analogue of perception "locates" *its* objects of judgment in a private, "inner" mental realm. Since each person "innerly" senses only his or her own experiences, you know what *you* experience as no one else does. On my story, by contrast, what gives experiential self-knowledge its distinctive character is to be found in a deep *dis*analogy with perception-based spatial identification of objects of judgment. There *isn't* anything relevantly similar to nongeneral forms of spatial perception on which to base demonstrative reference to one's own experiences. And it is ultimately because of this that we cannot understand demonstrative thoughts about our own experience as subject to the kinds of mistakes (reference failure, total predicative error) as we can suppose befall the judgments we make, when we attempt to refer to objects by representing their location. The absence of intelligible error of this sort, combined with the presumption that we do understand the expression of our thoughts, yields a warrant for demonstrative experience reports that correspond to first-person attributions of experience. And this, together with the phenomenal unity of experience, yields truth and warrant for these first-person attributions, but not for third-person attributions. It's not that I know my own experience in a way others don't, because I possess some special way of *perceiving* it—which, like sight, hearing, taste, touch and smell, allows me to judge of its objects—but *unlike* these, never informs other people of the same objects. Rather, I know my experience as others don't, because I can judge of it, but

without anything *like* these sensory means of access to it, and their attendant opportunities for intelligible error.

However, by contrast with others who reject perceptual models of self-knowledge, I am happy to tolerate and even embrace the notion of demonstrative reference to one's own experience. But then: how can there be such reference, without something like "inner sense"? What underlies this is rather an exercise of attention to experience, involving a subtle change in its phenomenal character, which, while it does not essentially consist in some form of representation of one's experience, does enable one to have thoughts phenomenally unified with the experience attended to, thoughts then which are about one's own (but only one's own) experience.[13] I suggest that one may be tempted to misconstrue this form of attention as itself a form of higher-order representation, rather than (what it really is)—the basis for developing and exercising the capacity for such reflection. And in succumbing to this temptation, some feel drawn to "higher-order thought" and "inner sense" theories of consciousness and self-knowledge.

The account I have offered here draws on several ideas that admittedly need much further elaboration and defense—more than I could reasonably be expected to give in a single paper—the notions I have employed of phenomenal unity, attention to one's experience, and demonstrative experience reports, among others. Also, a more nearly complete account faces rather large tasks—such as considering what range of first-person judgments we are entitled to think are warranted in the way proposed, and how far and in what manner this range might be expanded, by adding to the core account.[14]

But I believe I have done enough to articulate and argue in outline for a promising core theory of how we know our own experience, a theory worth further exploration as an alternative—both to those that appeal to perceptual models of consciousness,[15] and to those that give no central role to phenomenal consciousness.[16,17]

Notes

[1]Notice that, if I am assumed to have no warrant for first-person beliefs but what I make available to others, the claims I assert in this case warrant not only the belief that I feel thirst, not pain, but also the belief that *this* is what *I* believe. So, even if my first-person beliefs are presumed accurate, it seems I would still have more warrant for believing my own lie than for believing the truth in such circumstances. Secondly, my deceitful evidence need not take the form of assertions about what I do or don't feel. I can mislead others about my experience by engaging in thirsty, rather than "back pain feeling" behavior.

[2]See Siewert 1998, Chapter 3.

[3]I argue for this view in detail in Siewert 1998, Chapter 8.

[4]My rejection of inner sense is crucial to my case here, but a defense of this rejection would require too much detail to pursue here. I discuss "inner sense" and "higher-order representation" generally in Siewert 1998, chapters 4 and 6.

[5]Here and elsewhere, it has been pointed out to me that what I have to say is close to Kant's and Husserl's discussions of the unity of consciousness. I am sure that my conception of phenom-

enal unity is indebted to my having read these two, but since the interpretation of these authors is controversial, and requires painstaking exegesis, this is not the place to sort out the relation between their views and mine.

[6]Assuming I am normal (at least in this respect).

[7]The unity of consciousness is a difficult topic that certainly deserves consideration in its own right, and I do not claim to have done more here than secure a foothold. Some recent discussions to contrast with my own: Gertler (forthcoming); Lockwood 1989, p. 88ff; Marks 1981; Parfit 1985, pp. 245ff; Shoemaker 1996 pp. 178ff.

[8]See Shoemaker 1996, 218–20. Perhaps here I should note that my position can accommodate the Wittgensteinian view that one could not learn the meaning of sensation terms solely through "private ostention," and without applying them in connection with "public behavior." I can also accept the claim that *thinking demonstratively* about one's experience does not play a role at all in one's initial learning of experiential language. Though maybe *attending to* one's own experience (which is usually necessary, but never in itself sufficient for thinking about or referring to one's experience) *does* play an essential role in such initial learning.

[9]An example I owe to A.J. Kreider.

[10]It might seem that, if the general terms are "abstract" enough, it will be hard to see how one could be mistaken in applying them, even to a holographic image. For instance, one is not mistaken in believing, of the image, that *this is visible*, has *some* sort of *shape* and *color*, is an *entity*. But it seems doubtful to me that a person in this situation—"fooled" by the image—*must* have general concepts of visibility, shape, and entityhood, which he would apply to the referent of his 'this' (the image) *and* which *would in fact be true of it*. Maybe according to *his* concepts of visibility, shape and entityhood, neither a holographic image nor a hallucination could qualify as genuinely visible, shaped, colored entities. On his metaphysical view perhaps, they don't have what it takes to be properly construed as *subjects of predication* at all.

[11]The point I am making here bears some affinity with a central theme in Shoemaker's discussions of self-reference—the notion that when one refers to oneself as 'I' one is immune to certain errors of misidentification (see Shoemaker 1996, pp. 12, 19, 21, 196–7, 210–11).

[12]It may seem odd to give an account of one's warrant for first-person attributions of experience in terms of one's warrant for reports employing demonstrative reference to experience, because it may seem that surely the use of the first-person singular pronoun is primary and more basic than the use of demonstratives to refer to experiences. I can admit that in some sense this is correct. For instance, I can allow that one must first acquire the capacity to refer to items demonstratively (including oneself) that one locates in space, before one can learn to refer to one's own experiences demonstratively. What I would contest, however, is the idea that the way I understand such demonstrative reference to my own experience can be expressed entirely by relying on descriptions formed from the use of demonstratives by which I refer to myself, times, places, and objects identified by their location in space. I hold that the understanding of demonstrative reference to one's own experience cannot be reduced to the understanding involved in these other forms of reference.

[13]Perhaps this would be as good a place as any to make explicit: I do not believe that 'x attends to x's own experience' entails 'x forms some sort of representation of x's experience.' How can there be a sort of "attending to" that doesn't essentially involve a representation of what is attended to? (A question pressed on me by Bill Lycan.) Here I would first go back to the point that there is a sort of attending to one's experience that isn't merely a matter of judging that one's experience is of this or that kind. Consider the sort of attending that is involved when one's feeling of pain or nausea "occupies the center of one's attention," and the sort of attending to visual experience that is involved producing a perspectival drawing. I assume the former sort of attending can be done by animals and babies without their employing experiential *concepts*. And the latter sort of attending involves a change in the phenomenal character of one's experience, which cannot be identified with a change in what one *judges* about one's experience. Should we then assume that "attending to" requires some kind of representation of what is attended to, and infer that there

must be a *nonconceptual* form of higher order representation at work? Here I would say that the kind of attending to experience just illustrated seems to me plausibly regarded as falsifying that assumption, in the absence of independent support for this notion of a nonconceptual representation of experience.

[14]However, we shouldn't *underestimate* the range to which the account already provided here is applicable. It may be that two classifications I'm disposed to make of an experience are *equally warranted*—and I have warrant for both—though either one, even without the other, would give me some way of correctly conceptualizing the experience in question.

[15]Space does not permit me to critique the perceptual model in detail here; prominent recent exemplars include Armstrong 1980 and Lycan 1996.

[16]Again, space does not allow me to examine here such views in the kind detail they deserve, but recent examples of this kind of account can be found in Shoemaker 1996 and Bilgrami 1998.

[17]Many thanks to: Fred Altieri, David Anderson, Eivind Balsvik, Melissa Bergeron, Brie Gertler, Oliver Kaufman, A.J. Krieder, Noa Latham, Kirk Ludwig, Robert Lurz, William Lycan, David Pitt, Harvey Siegel, Tuula Tanska, Amie Thomasson, Corina Vaida, and Mike Veber.

References

Armstrong, D.M. (1980) "What Is Consciousness" in D.M. Armstrong, *The Nature of Mind and Other Essays*. Ithaca: Cornell University Press.

Bilgrami, A. (1998) "Self-Knowledge and Resentment," in C. MacDonald, B. Smith, C. Wright (eds.) *Knowing Our Own Minds*. Oxford: Oxford University Press.

Gertler, B. (forthcoming). "Introspecting Mental States," *Philosophy and Phenomenological Research*.

Lockwood, M. (1989) *Mind, Brain, and Quantum: The Compound 'I'*. Oxford: Basil Blackwell.

Lycan, W.G., (1996) *Consciousness and Experience*. Cambridge MA: MIT Press.

Marks, C.E., (1981) *Commissurotomy, Consciousness, and Unity of Mind*. Cambridge MA: MIT Press.

Parfit, D. (1980) *Reasons and Persons*. Oxford: Oxford University Press.

Shoemaker, S. (1996) *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press.

Siewert, C. (1998) *The Significance of Consciousness*. Princeton: Princeton University Press.