MOORE’S PARADOX AND CONSCIOUSNESS†

David M. Rosenthal
City University of New York Graduate School

I. Moore’s Paradox and Transparency

As G. E. Moore famously observed, sentences such as ‘It’s raining but I don’t think it is’, though they aren’t contradictory, cannot be used to make coherent assertions. The trouble with such sentences is not a matter of their truth conditions; such sentences can readily be true. Indeed, it happens often enough with each of us that we think, for example, that it isn’t raining even though it is. This shows that such sentences are not literally contradictory. But even though such sentences have unproblematic truth conditions, we cannot say the same about their conditions of assertibility. There are no circumstances in which one can use such sentences to perform coherent assertoric speech acts. Situations exist in which these sentences would be true, but none in which anybody could use them to say so.

This phenomenon is known, following Wittgenstein, as Moore’s paradox. As some authors have noted, the difficulty arises not only with assertions, but also with speech acts whose illocutionary force is not assertoric. Thus I cannot coherently say ‘Thank you but I feel no gratitude’ or ‘Rain is likely, but I don’t expect it’. If somebody were to produce such an utterance, we would automatically try to interpret the words nonliterally, or as having been used ironically or with some other oblique force. Only by doing so could we regard the speaker as having performed any speech act at all.

One reason to trace the absurdity of Moore’s paradox to the impossibility of performing any coherent speech act is that the absurdity may vanish when the very words of such sentences are embedded in a larger sentence. Consider, for example, ‘Suppose it’s raining but I don’t think it is’. If, instead, the trouble lay with the semantic content of those words, it would pursue the words even when they’re embedded in such larger contexts.

It has not generally been noted that there’s an important kinship between Moore’s paradox and Descartes’s cogito. The sentence ‘I don’t exist’ has unproblematic truth conditions. Not only is it possible for the sentence to be true; it once was true for each of us. Nonetheless, the sentence has no coherent conditions of assertibility; no circumstances exist in which one could coherently
perform a speech act by assertively producing that sentence. It’s arguably that which underwrites Descartes’s claim in Meditation II that “the statement ‘I am, I exist’ is necessarily true every time it is produced by me, or mentally conceived.”

Like Moore’s paradox, the cogito is a function not of the truth conditions of sentences, but rather of the performance conditions of speech acts and, possibly, the mental analogue of these performance conditions for the corresponding propositional attitudes. A useful test is to see whether a change of grammatical tense or person relieves the difficulty. There’s no problem about saying ‘I didn’t exist’ (or ‘won’t exist’), or ‘It’s raining but you don’t think it is’.

The sentence ‘I am not thinking’ resembles the sentence ‘I don’t exist’ in exactly these ways. Just as ‘I exist’ is true every time I think it or assert it, so also is the sentence ‘I am thinking’. This led Descartes to conclude that the “I” whose existence he had established is essentially a thinking thing. And this appears to point to a certain kind of self-knowledge. I cannot, when I’m thinking, doubt that I am thinking. So it’s tempting to suppose that my being engaged in thinking is enough for me to know that I’m thinking. And because this conclusion rests on just the kind of reasoning that underwrites Moore’s paradox, Moore’s paradox seems to be relevant at least to a certain sort of self-knowledge.

If being engaged in thinking is enough to know that one is thinking, then to that extent at least the mind is transparent to itself. As Descartes put it, “nothing can be in our mind of which we are not at that time conscious.” But this thesis of transparency is untenable. Not all of our thinking is, in fact, conscious thinking. This is clear in part from results in clinical and cognitive psychology, but it’s also obvious from everyday, commonsense considerations.

We often consciously puzzle over a problem, for example, only to have the solution occur to us later, apparently spontaneously, without the problem having in the meantime been in any way consciously before our mind. It’s hard to see how this could happen unless problem solving thinking sometimes occurs without being conscious. Such nonconscious problem-solving sometimes occurs, moreover, when we’re not even aware that any thinking is going on, for example, when we’re asleep. Many other commonsense considerations support the same conclusion. Simply being engaged in thinking plainly isn’t sufficient for that thinking to be conscious.

What, then, went wrong with the reasoning that seemed to show otherwise? It’s undeniable that, just as I cannot sensibly assert ‘It’s raining but I don’t think it is’, so I cannot assert that I’m not thinking, or perform the mental act of thinking that I’m not thinking. But this hardly shows that whenever I do think, I automatically think that I’m thinking. The question of whether I’m thinking may simply not come before my mind, consciously or otherwise. Indeed, though thinking does occur during sleep, we have no reason to suppose that, when it does, the question of whether one is thinking ever occurs to us as well.

That thinking can occur without one’s being aware of it is evident also from consideration of nonhuman cases. It’s overwhelmingly likely that some creatures
have the capacity to think despite their having no concept of thinking. Such creatures would therefore be unable ever to think that they think.

Descartes and, in a somewhat similar spirit, Donald Davidson both deny that such a thing is possible. But it’s hard to see any reason for that denial that doesn’t beg the question. Descartes insists that “in order to know what doubting is, and what thinking is, only doubting or thinking is needed,” but he gives no reason for that claim. Davidson, by contrast, argues for his view that one cannot believe things without having the concept of believing, maintaining that unless one understands the difference between true and erroneous believing one cannot have beliefs at all. But his argument shows at best that the ability to distinguish between truth and error is needed to have the concept of belief, not that it’s also needed simply to have beliefs. Davidson offers no argument to bridge that gap.

Humans, of course, do often engage in thinking of a kind that would be impossible without a concept of thinking. We sometimes think about the thinking that we or others are engaged in. Indeed, I’ve argued elsewhere that our being conscious of our conscious mental states is due to our having higher-order thoughts about those states. But often our thinking isn’t in any way conscious. And such consciousness apart, it’s relatively seldom that we think about anybody’s thoughts, our own or anybody else’s. So why couldn’t there be creatures whose thinking is always like our nonconscious thinking, and who moreover never think about anybody’s thoughts? Such creatures might well have no concept of thinking. Though their mental lives would plainly be far less rich than ours, if the concept of thinking applies in our case both to thinking that we think about and to thinking that we don’t, it should apply also to the more limited thinking that these other creatures engage in.

Sydney Shoemaker has developed a different argument for the view that we cannot straightforwardly apply our concepts for human mental phenomena to nonhuman animals. The reason, according to Shoemaker, is that many features of human mental phenomena are central to the very concepts of those phenomena. Minimal rationality, for example, is central to many such concepts. Since nonhuman animals apparently lack minimal rationality, Shoemaker concludes that applying these concepts to nonhuman animals is “problematic.”

But Shoemaker’s principle about when we can apply various mental concepts leads to untenable results. The concepts whose application to nonhuman creatures Shoemaker sees as problematic are human folk-psychological concepts. And it would be unsurprising if these concepts were tailored to human mental phenomena. But the distinctive features of human mentality that these concepts embody may show relatively little about mind generally. For one thing, there could be creatures whose mental states exhibited important features not shared by human mental states, for example, rationality superior to our own. Concepts for human mental phenomena should apply unproblematically to corresponding states of such superior creatures. But on Shoemaker’s view, such beings could not have a concept of mind, in our sense of the word.

There’s also reason to doubt Shoemaker’s claim that minimal rationality is
essential to human mentality. Human mental states don’t always manifest the ties with other states that subserve such rationality and hence, on Shoemaker’s view, self-knowledge. But our mental concepts plainly apply to human states, even when they lack those ties. So we have no reason not to apply these concepts to states of creatures, even when those states never exhibit the connections that subserve minimal rationality.

Whatever we say about nonhuman creatures, not all human thinking is conscious, nor can it always readily become conscious. So it’s a mistake to expect the analogue of Moore’s paradox that underlies the cogito to help establish any thesis about transparency. All it shows is that when creatures with the concept of thinking are actually thinking, they cannot sensibly deny, in speech or in thought, that they are thinking.

It’s worth taking note of another thesis about transparency that’s arguably related to Moore’s paradox. It’s often held that knowing something is sufficient for one to know that one knows. And if so, knowing is in effect transparent to itself. Consider the sentence ‘I know that p but I don’t know that I know it’. Though not contradictory, it seems that there cannot be coherent conditions for asserting the sentence.

Such sentences are not strictly cases of Moore’s paradox, since even if one didn’t know that one knew one might still believe that one did. But Moore’s paradox does help us understand what’s wrong with that sentence. By Moore’s paradox, saying I know that p makes it incoherent to deny that I believe that I know it. And saying I believe that I know that p would make it incoherent to deny that it’s true that I know it. As for justification, if I say I know that p, I take my assertion to be justified; so I can’t then coherently deny knowing that I know on the basis of my not being justified in thinking I know. Similarly, if I say I know that p, it would be incoherent to deny that I’ve tracked my knowing that p. Nonetheless, this cousin of Moore’s paradox no more warrants concluding that knowing actually implies knowing that one knows than the cogito entitles us to hold that all thinking is conscious. All it shows, rather, is that we cannot coherently assert that we know something and at the same time deny knowing that we know it.

II. Language and Consciousness

I’ll return in section IV to the question of whether Moore’s paradox does help establish some form of self-knowledge about one’s own mental states. But whatever the case about that, I want now to argue that Moore’s paradox does point to factors that help explain a special case in which our thinking is almost without exception conscious thinking.

Whenever we say anything sincerely, we express some intentional state that we’re in. If I sincerely say, for example, that it’s raining, I express my thought that it’s raining. Similarly with other sorts of speech act and the mental attitudes that correspond to them. If I say thank you, I express my gratitude; if I say it
Moore’s Paradox and Consciousness / 317

will probably rain I express my expectation that it will. And so forth. Every sincere speech act expresses an intentional state with the same content as the speech act and a mental attitude that corresponds to the illocutionary force of the speech act.

Moore’s paradox reflects this connection between our speech acts and the intentional states they express. I cannot use the sentence ‘It’s raining but I don’t think it is’ to make a coherent assertion precisely because the assertion that the first conjunct purports to make expresses the very intentional state that the second conjunct denies I’m in. To be used in performing a conjunctive speech act, the first conjunct of a Moore’s-paradox sentence must express a corresponding thought. But the second conjunct denies that any such thought exists, thereby denying that the whole sentence can be used to make any coherent assertion.17

The same holds for the sentences ‘Thank you but I’m not grateful’ and ‘It’ll probably rain but I don’t expect it to’; thanking somebody expresses one’s gratitude, and saying something will probably happen expresses one’s expectation. Similarly for versions of Moore’s paradox derived from other illocutionary forces and the mental attitudes that correspond to them.

Many of our intentional states are not in any way conscious states. But when we express our intentional states in speech, those states are always conscious, or almost always. Suppose I think it’s raining. My thought may or may not be conscious. But if I verbally express that thought by asserting that it’s raining, the thought is invariably conscious. Indeed, with an exception that I’ll mention in closing, any intentional state that I express with a speech act will be a conscious intentional state.

Intentional states are expressed not only by speech, but also by many forms of nonverbal behavior. Taking an umbrella may express my belief that it will rain, or my desire not to get wet, or both. Facial expressions and bodily movements of various kinds may express my delight in something or my dislike of it, my fear of something or my anticipation of some future event. In all these cases the intentional state my nonverbal behavior expresses may well be conscious; but it may also fail to be. I may take the umbrella absently, “without thinking,” as we might say—that is, without thinking consciously. And one’s facial expressions and bodily movements often betray delight, distaste, fear, and expectations of which one is, oneself, wholly unaware.

This difference between expressing intentional states verbally and expressing them by one’s nonverbal behavior is striking. When an intentional state is expressed in speech, it’s always conscious, but when it’s expressed nonverbally it needn’t be. Doubtless this contrast helped to persuade Descartes and others that language and consciousness are both essential to mentality. It’s tempting to think and talk about intentional states in terms of the speech acts that would express them, since doing so enables us to describe the content and mental attitude of intentional states with an accuracy and precision not otherwise available.18 And if we describe our own intentional states and those of others by reference to speech acts that would express those states, we may take the further step of
thinking about all intentional states as though they are expressed in speech, and hence conscious. But as we’ve seen, there is ample evidence that our commonsense, folk-psychological conceptions don’t require that all intentional states are conscious, and indeed ample evidence that many of them are not.

It might be argued that slips of the tongue, such as those said to occur in Freudian parapraxis, cause difficulty here. Such slips, it may seem, express thoughts of which we’re not conscious. If so, the thoughts our speech acts express sometimes aren’t conscious. But there would also be another problem. The latent content such slips supposedly express is typically distinct from the manifest content that matches the semantic meaning of the speech act. So the content of these speech acts would diverge from the content of the intentional states they express. Of course, the latent thought expressed by one’s slip is often conscious; witness the conscious embarrassment sometimes caused by realization of what one’s slip reveals. Even so, it’s arguable that the content of the speech acts involved in such slips corresponds not to the latent content, but to the manifest content.

But in fact things are more complicated. Suppose an unintended word intrudes into one’s performing of a speech act, thereby revealing some nonconscious thought that one has. The occurrence of that word is best construed not as an integral part of the speech act, but in effect as a piece of nonverbal behavior. It’s on a par with cases in which one’s tone, or other aspects of one’s utterances, unintentionally betray one’s nonconscious intentional states. The way we utter things often manifest intentional states without thereby expressing them verbally. The latent thought in these cases is simply a causal factor, somewhat like an external noise, which interferes with the correct expression of the conscious thought corresponding to the speech act. So the slips that occur in parapraxis are not counterexamples to the foregoing generalizations.

It’s important here to distinguish between verbally expressing an intentional state and reporting that state. Although verbally expressing our intentional states and reporting them are both ways of conveying to others what intentional states we’re in, there are important differences between these two ways. When I think it’s raining, I verbally express my thought by saying that it’s raining. My verbal expression has the same content as the intentional state it expresses and an illocutionary force that corresponds to the mental attitude of the intentional state. By contrast, I report my thought that it’s raining when I explicitly say that I have that thought, for example, when I say ‘I think it’s raining’. And the content of my thought differs from the speech act that reports it.

The contrast between reporting and expressing one’s intentional states emerges most decisively, however, with intentional states that have a nonassertoric mental attitude. If I wonder whether it’ll rain, I express my state of wondering by saying ‘Will it rain?’, whereas I report my wondering by saying ‘I wonder whether it’ll rain’. Here the illocutionary force of the speech act that verbally expresses my intentional state is that of a question, corresponding to the mental attitude of wondering. By contrast, the illocutionary force of my report
is assertoric, as it is with all reports of intentional states.

Again, Moore’s paradox is helpful. I cannot assertively produce the sentence ‘It’s raining but I don’t think it is’ because asserting the first conjunct would express an intentional state that the second conjunct denies I am in. Suppose, now, that there were no difference between reporting an intentional state and verbally expressing it. Then my denial that I am in the intentional state of thinking that it’s raining would be tantamount simply to expressing the thought that it’s not raining. Accordingly, the Moore’s-paradox sentence would be equivalent to ‘It’s raining and it’s not raining’, which is an actual contradiction. But Moore’s paradox is not literally contradictory. To avoid this result, therefore, we must distinguish reporting our intentional states from verbally expressing them.

Ordinary usage of verbs of propositional attitude may sometimes seem to run together reporting with expressing of our intentional states. For example, when one says ‘I doubt (or suppose) that $p$’, ‘I choose this one’, or ‘I sympathize with you’, it may seem that one expresses one’s doubt, supposition, choice, or sympathy, rather than reports those states. But this appearance results from focusing exclusively on performance conditions. Once we take into consideration the truth conditions of such sentences, it’s clear that these speech acts report the relevant attitudes. Indeed, as just noted, Moore’s paradox would otherwise be an outright contradiction.

Distinguishing between reporting our intentional states and expressing them allows us to formulate a second connection between consciousness and speech. Verbally expressing an intentional state is, with a certain type of exception, a sufficient condition for that state to be conscious. But also, when a creature has the requisite concepts and linguistic ability, a mental state’s being conscious is sufficient for the creature to be able to report being in that state.

This second connection between language and consciousness is to be expected. Conscious mental states satisfy two conditions: we’re conscious of being in them, and the way we’re conscious of them seems to us to be immediate. We needn’t, of course, be conscious of our conscious states in a way that’s at all attentive or focused; we’re only peripherally conscious of the vast majority of our conscious states. But when mental states occur of which we are not in any way conscious, those states are not conscious states. And, given the requisite concepts and linguistic ability, being conscious of something is sufficient for being able to report about it.

What about the other way around? Being able to report about a mental state is not sufficient for that state to be conscious, because a state’s being conscious requires not just that one is conscious of that state, but also that one is conscious of it in a way that’s from an intuitive point of view immediate. Being able to report some mental state one’s in solely because of behavioral evidence one has that one is in that state is not sufficient for the state to be a conscious state.

This intuitive immediacy may not amount to much. It’s enough for our consciousness of our own mental states to be intuitively immediate that we’re
conscious of them in a way that doesn’t rely on any inference, at least not on any inference of which we’re aware. So, although being able to report on a state doesn’t suffice for that state to be conscious, being able to report on it noninferentially does.

III. Verbally Expressed Thoughts

We are now in a position to explain why verbally expressed intentional states are invariably conscious. Moore’s paradox is absurd because the speech acts of asserting that $p$ and asserting that I think that $p$, though they differ in respect of their truth conditions, have roughly the same conditions of assertibility. Any circumstances in which I could say that $p$ are circumstances in which I could say I think that $p$. And with a qualification that won’t affect the argument here, the converse holds as well.

More important, this performance-conditional equivalence is second nature for us. We automatically take saying ‘$p$’ and saying ‘I think that $p$’ to amount to much the same thing insofar as conditions of assertibility are concerned. Indeed, we ourselves tend insensibly to slip between saying the one and saying the other; we may even have difficulty recalling on specific occasions which of the two forms we used. It’s a matter of well-entrenched linguistic habit that the two are, for practical purposes, interchangeable. It’s because this performance-conditional equivalence is second nature to us that Moore’s paradox is not just absurd, but intuitively jarring. We know automatically that no circumstances can exist in which somebody could sensibly say that $p$ but deny thinking that $p$.

The relevant performance-conditional equivalence emerges in the absurdity of the following conversation. Suppose I ask you whether it’s raining, you say ‘I think so’, and I remonstrate that I asked not about your intentional states but about the weather. That’s a bad joke in roughly the way Moore’s paradox is, though it’s harder here to imagine any charitable reinterpretation that would save things. And it’s jarring in just the way Moore’s paradox is, because in that context we automatically regard your saying ‘I think so’ as performance-conditionally equivalent to your saying that it’s raining. Indeed, it’s so automatic as to be second nature.

Suppose I think that it’s raining, and I express my thought by saying ‘It’s raining’. Because saying ‘It’s raining’ is performance-conditionally equivalent to saying ‘I think it’s raining’, I could equally well have said that I think it’s raining. And because that equivalence is second nature for us, I might as easily have said the other; in most circumstances it’s likely to be a matter of complete indifference to me which I say.

Now put performance conditions aside for a moment, and think instead of truth conditions. What makes the sentence ‘I think it’s raining’ true isn’t the rain, but my being in a certain intentional state. Its being true requires that I think it’s raining. So if I were to say ‘I think it’s raining’, however we may take that remark in respect of performance conditions, I am literally telling you about one
of my intentional states. I am reporting a certain thought.

So, when I express my thought that it’s raining by saying ‘It’s raining’, I could equally well have reported my thought that it’s raining. Moreover, because it’s second nature for us that saying ‘It’s raining’ is performance-conditionally equivalent to saying ‘I think it’s raining’, if I had instead said ‘I think it’s raining’, my report would have not have been based on any inference, at least not on any inference of which I was conscious. So, whenever I actually say ‘It’s raining’, I could equally well have noninferentially reported my thought that it’s raining.

But noninferential reportability is sufficient for a state to be a conscious state. So, given that it’s second nature for us that saying ‘It’s raining’ is performance-conditionally equivalent to saying ‘I think it’s raining’, whenever I verbally express any intentional state, that state will be conscious. And, since Moore’s paradox is a reflection of that performance-conditional equivalence, we have used the factors that underlie Moore’s paradox to explain why verbally expressed intentional states are always conscious.

Let me again briefly rehearse the argument. I’ve urged that Moore’s paradox is absurd because saying ‘\(p\)’ is performance-conditionally equivalent to saying ‘I think that \(p\)’. And it’s intuitively jarring because that equivalence is so automatically a part of how we use these words. Given the equivalence, whenever I say ‘\(p\)’, I could equally well have said ‘I think that \(p\)’. But to say ‘I think that \(p\)’ is, literally, to report one’s thought that \(p\). And because it’s second nature that saying ‘\(p\)’ is performance-conditionally equivalent to saying ‘I think that \(p\)’, saying ‘I think that \(p\)’ would be reporting one’s thought noninferentially. Being able to report an intentional state noninferentially, however, is sufficient for that state to be conscious. So the factors underlying Moore’s paradox also explain why all verbally expressed intentional states are conscious.

The explanation applies equally to intentional states whose mental attitude is not assertoric. If I ask, ‘Is it raining?’, for example, the performance-conditional equivalence revealed by Moore’s paradox shows that I could just as well have said ‘I wonder whether it’s raining’. If I say ‘Close the door’, I could instead have said in so many words that I want you to close it. Whenever I verbally express these intentional states, I’m able also to report those states noninferentially; similarly for states with other mental attitudes. A intentional state’s being verbally expressed is sufficient for its being conscious.

The connection between an intentional state’s being expressed and its being conscious holds only, as we’ve seen, when the state is expressed in speech. States expressed by nonverbal behavior often aren’t conscious. It will reinforce the foregoing explanation if we can use it to show why things are different in the two kinds of case.

Unlike speech acts, the pieces of nonverbal behavior that express our intentional states do not have established performance conditions. Taking my umbrella may in certain circumstances be odd or irrational or inappropriate, but
these things don’t define performance conditions for my action. Because taking my umbrella has no performance conditions, even if that action nonverbally expresses, say, my desire not to get wet, the action cannot have the same performance conditions as a speech act that reports the desire I nonverbally express. So taking the umbrella will not be interchangeable, as a matter of well-entrenched linguistic habit, with the making of such a report. I might well perform the action of taking the umbrella even when I could not readily report the desire my action expresses. Similarly with other cases of nonverbally expressing our intentional states. An intentional state’s being nonverbally expressed is not, therefore, sufficient for that state to be conscious.

We could, perhaps, imagine a piece of nonverbal behavior becoming so well-entrenched in our social practices that we came to see that behavior as having performance conditions. For example, taking an umbrella might come not simply to indicate the likelihood of rain, but also actually to mean that it’s going to rain, that is, to have that semantic content. Taking an umbrella would then be performance-conditionally equivalent to reporting one’s thought that it’s going to rain. It’s arguable that if that equivalence also became second nature, one would be unable to take an umbrella without one’s thought that it’s going to rain being conscious. Intuitions about this kind of case, however, are unlikely to be firm enough to test this idea, since that kind of thing is unlike anything that actually happens.

I’ve argued that Moore’s paradox is absurd because every speech act is roughly equivalent, in respect of performance conditions, to a report of the intentional state that the speech act expresses. But one might wonder whether the trouble with Moore’s paradox is simpler than this. Perhaps all that’s wrong is that one cannot say ‘It’s raining but I don’t think it is’ without betraying one’s insincerity in saying it. In effect, Moore’s paradox would then be simply self-defeating. Indeed, Moore himself at one point suggests this kind of diagnosis.

On such an account, we need to explain why it is that saying ‘It’s raining but I don’t think it is’ automatically betrays one’s insincerity. One possibility is suggested by Paul Grice’s idea that my meaning something involves intending that my hearer believe that I believe what I say. On this proposal, I couldn’t mean that p if in the same breath I said I don’t believe it. So I couldn’t say it sincerely. But this is implausible as an explanation of what’s wrong with Moore’s paradox. Moore’s paradox is absurd independent of any context of communication. It’s absurd even in soliloquy, where no betrayal of insincerity is relevant; one cannot say even to oneself ‘It’s raining but I don’t think it is’. The best explanation of this independence from any context of communication will arguably appeal to the performance-conditional equivalence I’ve been relying on.

I suggested earlier that this equivalence holds because of a connection between intentional states and the speech acts that express them. All speech acts express corresponding intentional states; we cannot coherently assert ‘It’s raining but I don’t think it is’ because what the first conjunct purports to assert expresses
the intentional state that the second conjunct denies I am in.

Earlier I described the regularity concerning speech acts’ expressing intentional states solely in terms of sincere speech. But how about insincere speech? When my utterance of ‘It’s raining’ is a lie, none of my assertoric intentional states has the content it’s raining; so my words can’t express any such actual state. Still, don’t I say it’s raining? Only in a qualified way. ‘Say’ and related verbs of illocutionary act are sometimes used in a weak sense, as roughly equivalent to ‘utter’; saying in this sense involves no illocutionary act. For example, when we recite lines in a poem or play, we seldom if ever produce them with illocutionary force, though typically we pretend to; we say things then in only the weak sense. When an actor utters ‘My heart is in the coffin there with Caesar’, nobody would take him to perform an actual illocutionary act, despite his pretense to do so. As Frege remarked, “stage assertion is only sham assertion.”

Insincere speech also operates this way. When we speak insincerely we pretend to be in intentional states that don’t exist, and thereby pretend to perform the relevant illocutionary acts. Lying and play acting differ, of course, in the motives we have for pretending; so in lying my pretense isn’t candid, as it is in play acting. Moreover, in lying the character I play is a fictional version of myself, one who actually believes the things I pretend to assert. So when I speak insincerely, it’s only in the reciting sense that I say anything.

One might reject this account of insincere speech on the ground that it treats sincere and insincere speech differently. But a uniform treatment has significant disadvantages. Since no corresponding intentional state causes insincere utterances, a uniform treatment must require that the connection between thought and speech generally is oblique. And that oblique connection is almost certain to appeal to a context of communication, even though speech, sincere and insincere alike, often occurs independently of any such context. We can conclude that insincere speech notwithstanding, all illocutionary acts express corresponding intentional states.

What, then, is the status of the regularity about speech acts’ expressing corresponding intentional states? It cannot be purely a conceptual matter, a matter of nothing more than the meanings of the relevant words. If it were part of the meaning of ‘assert’ that assertions express corresponding beliefs, Moore’s paradox would not simply have problematic performance conditions, but would be an outright contradiction. Nonetheless, it’s somehow a part of the way we automatically think about asserting and other speech acts. The only alternative is that the connection between speech acts and corresponding intentional states is a particularly well-entrenched part of our folk-psychological conceptions.

Those who diagnose Moore’s paradox as being conceptually defective sometimes urge that the only alternative to that view is an account on which Moore’s paradox “merely...depict[s] situations which we take to be extremely unlikely.” But here as elsewhere, the dichotomy between conceptual and merely empirical misleads. Much that’s intuitively impossible runs counter not to
established semantic connections, but only deeply entrenched background beliefs about the way things are.

For a speech act to express an intentional state is, in part, for that state to be among the causal factors leading to the speech act. When I say that it’s raining, many things causally contribute to my saying it. But if my speech act verbally expresses my thought that it’s raining, my having that thought is one of the causal factors that lead to my saying it. I’ve argued that this connection between speech and thinking is, for us, so strongly second nature that simply saying that p inevitably makes it obvious to one that one thinks that p. And the very same connection is responsible for there being no coherent performance conditions for asserting Moore’s paradox.

IV. Moore’s Paradox and Self-knowledge

It’s impossible, even speaking to oneself, to assert ‘It’s raining but I don’t think it is’. Moreover, it’s no less impossible to think that thing assertorically. Just as there are no coherent conditions for asserting ‘It’s raining but I don’t think it is’, so there’s no coherent mental analogue of assertibility conditions for thinking ‘It’s raining but I don’t think it is’.

But the tie between speech acts and the intentional states they express can’t by itself explain why it’s impossible to think Moore’s paradox. Nor does that tie have any suitable mental analogue. Because asserting expresses a corresponding belief, it’s impossible to assert anything without believing it. But it’s plainly possible to think something without thinking that we think that thing. Indeed, it’s arguable that that typically happens when our thoughts aren’t conscious. We can’t explain the impossibility of thinking Moore’s paradox by appeal to the same factors that explain the impossibility of saying it.

It’s sometimes held that a satisfactory explanation of Moore’s paradox must explain in a uniform way why we cannot say it and why we cannot think it. This requirement would preclude explaining the linguistic version of Moore’s paradox by reference to performance conditions. Proponents of a uniform treatment typically also insist that semantic contradictoriness somehow underlies Moore’s paradox. Speech acts and intentional states literally share their content; so, if we must give an uniform explanation, it will have to be in semantic terms. But we can use the uniformity requirement to undermine the performance-conditional explanation only if we have independent reason to adopt that requirement. And it’s unlikely that any reason exists apart from the claim that Moore’s paradox demands a semantic explanation, which in this context is question begging.

How, then, can we explain the impossibility of thinking Moore’s paradox? Here, it seems, we must appeal to a certain rationality that governs many of our thoughts. We can perfectly well have the thought that p without thinking that we have it. But it’s irrational both to think that p and to think that one doesn’t think it. So, if the question arises about whether one thinks that p and one does actually think that p, it would be irrational to conclude that one doesn’t think it.
Rationality does not of course dictate that this question will arise. And when it doesn’t, it’s generally possible to think that $p$ without thinking that one thinks that $p$. In that kind of case, the first-order thought arguably will not be a conscious thought. But insofar as we are rational in this particular way, if one has some particular thought and it’s possible for the question to arise as to whether one has it, it must also be possible for the relevant higher-order thought to occur. As Kant insisted, in the case of rational beings “the representation ‘I think’...must be capable of accompanying all other representations...”\(^\text{33}\)

Sydney Shoemaker has argued for an even stronger connection between rationality and our having thoughts about our thoughts. Adopting the functionalist view that believing and desiring are defined in terms of certain connections with actions, sensory input, and other mental states, Shoemaker argues that these connections actually constitute a certain minimal rationality. Moreover, he urges that beliefs and desires do more than simply cause the behavior with which they’re rationally connected; they also “rationalize” that behavior. And to rationalize a course of action, beliefs and desires must refer to it.

But rationality applies not merely to overt actions, but also to the very mental activities of believing and desiring, themselves. Just as having certain beliefs and desires may make a certain course of action rational, so too it may be rational to adopt certain beliefs and desires given that we have certain others. But as with overt courses of action, we do not count as rational a person’s adopting or changing certain beliefs or desires unless others of the person’s beliefs and desires rationalize that adoption or change of beliefs and desires. And the beliefs and desires that do this rationalizing will have to refer to those which are rationally adopted. The rational fixation of beliefs and desires requires one to have higher-order beliefs about the first-order beliefs and desires one rationally fixes. Accordingly, such rationality cannot occur without our being conscious of our first-order beliefs and desires.\(^\text{34}\)

As Shoemaker notes, one could deny that rationality in the adopting of beliefs and desires is due to explicit rationalizing by higher-order beliefs and desires. Perhaps rationality here requires only that the causal relations that hold among our first-order beliefs and desires conform to rational standards.\(^\text{35}\) It’s far more plausible that beliefs and desires are states whose causal ties tend to conform to patterns we count as rational than that they are definitionally states whose rational fixation requires the causal influence of higher-order beliefs and desires. Perhaps, then, rationality doesn’t require us to have higher-order beliefs.

But even if rationalization of the sort Shoemaker describes does occur, there’s another difficulty with his argument. Beliefs and desires are sometimes occurrent, but often they’re dispositional. In their dispositional versions, belief and desire are simply dispositions to be in the relevant occurrent intentional states. And the beliefs and desires that rationalize actions will often be dispositional in just this way. But simply being disposed to have an occurrent belief or desire about something doesn’t make us conscious of that thing. So, when the higher-order thoughts and desires that rationalize our adoption of...
certain first-order beliefs and desires are merely dispositional, they won’t make us conscious of the first-order states. So they won’t make those first-order states conscious states. Moreover, there could be creatures for which that’s how it always happens. In the human case, the higher-order beliefs and desires that rationalize other intentional states are often occurrent. But rationality would not be compromised if the higher-order states were always dispositional. Intentional states can be rational without being conscious; indeed a creature could exist for which all rationality among its intentional states was that way.\footnote{36}

Connections between rationality and consciousness to one side, Shoemaker has developed an ingenious and probing argument that links Moore’s paradox to one’s having first-person access to one’s mental states.\footnote{37} He defines as self-blind a being that has our concepts of mental phenomena and can therefore ascribe to itself mental states, but which has no first-person access to its mental states. A self-blind individual could come to believe that it’s in some particular mental state, but only in a characteristically third-person way. Such an individual, moreover, might be presented with highly compelling third-person evidence both that it’s raining and also that it doesn’t believe that it’s raining. Since the self-blind creature lacks any first-person access, this third-person evidence would give it good reason to assert ‘It’s raining but I don’t think it is’.

Shoemaker argues, however, that Moore’s paradox would be absurd even for a self-blind creature, so defined. He urges that such an individual, simply by being conceptually unimpaired, would recognize the absurdity of Moore’s paradox; on my account, that would mean recognizing the performance-conditional connections between saying that \( p \) and saying ‘\( I \) think that \( p \)’. But then, Shoemaker urges, our ostensibly self-blind individual would be indistinguishable from somebody that does have ordinary first-person access. And if this is so, self-blindness is not a possibility, after all; any creature with our mental concepts and the ability to ascribe mental states to itself in a third-person way would also be able to do so in a first-person way as well. First-person access to one’s own mental states would follow from that conceptual capacity alone.\footnote{38}

For creatures relevantly like us, all verbally expressed intentional states are conscious. So any such creatures will have first-person access to all their verbally expressed intentional states. The factors that explain Moore’s paradox, I’ve argued, also underlie this connection between verbal expression and consciousness. But very few of our intentional states are actually expressed in speech; indeed, few come close to being so expressed. Moore’s paradox, moreover, arguably pertains in the first instance to whether certain speech acts can be performed. So even if a self-blind creature does have first-person access to its verbally expressed intentional states, it’s unclear why merely recognizing the absurdity of Moore’s paradox would extend such first-person access to intentional states that aren’t expressed in speech.

Moreover, even a self-blind individual that recognized the absurdity of Moore’s paradox might not be like us in the ways required for it to have first-person access even to its verbally expressed intentional states. By the definition
of self-blindness, such an individual would have our mental concepts, and be able to believe of itself that it's in various mental states. And, as Shoemaker's argument makes clear, it will have the ability to express its intentional states in speech. But on the argument I put forth earlier, that's not sufficient for all the self-blind individual's verbally expressed thoughts to be conscious. It's also necessary that the performance-conditional equivalence between saying 'p' and saying 'I think that p' is second nature to that individual.

This is important because, if that connection is not second nature, the access an individual has to its verbally expressed thoughts will be based on some conscious inference. And access that's thus consciously mediated would be characteristically third-person access, even if it's access to the individual's own states. It would therefore not make one's verbally expressed thoughts conscious. And since genuine self-knowledge requires conscious first-person access, it would not result in self-knowledge, either.

Shoemaker often writes as though sensitivity to the absurdity of Moore's paradox would be a direct result of having unimpaired conceptual capacities. If so, the relevant performance-conditional equivalences would presumably be conceptual truths; they would thus be second nature for us. I've argued against this idea. It's difficult to see how to explain Moore's paradox by appeal to the meanings of words like 'assert' and 'think' without construing it as an outright contradiction. It's more plausible, I think, to see the connection between asserting and believing as part of our folk-psychological knowledge about these things—part of a folk theory so well-entrenched as to constitute commonly shared background knowledge. The relevant conceptual competence is necessary, but not sufficient, for the performance-conditional equivalence. If this is right, our self-blind creature's access to its own mental states would be mediated by a conscious inference that relies on the relevant folk-theoretic connection.

But even if the relevant tie between thinking and asserting were an exclusively conceptual matter and not part of folk theory, our self-blind individual might still have to rely on some conscious inference for access to its own mental states. Conceptual truths function in effect as null premises in inference. So any inference that relies solely on such truths would be automatic. But the self-blind individual's inference would rely on more than the conceptual connection between speech and thinking. That inference would go from that conceptual connection plus the fact that the individual asserts that p to that individual's thinking that p. Indeed, Shoemaker sometimes seems to envisage the need for some conscious inference. "[I]t ought to possible," he writes at one point, "to get [our self-blind creature] to recognize that the assertive utterance of Moore-paradoxical sentences involves some sort of logical impropriety."

Perhaps, however, the relevant inference would after a time come to be second nature, and would therefore no longer be conscious. The resulting access that the self-blind individual would have to at least those of its thoughts which are verbally expressed would then not be based on any inference of which it was aware. Such access would therefore be indistinguishable from the first-person
access we have to our conscious states. But the concept of self-blindness does not ensure that the required inference would ever stop being conscious. And if it didn’t, the self-blind individual would continue to be limited to having only third-person access to its mental states.

Shoemaker argues that any individual with mental concepts “will be aware of having [mental states] when it does, or at least will become aware of this under certain conditions (e.g., if it reflects on the matter).” Shoemaker may be assuming here that “reflect[ing] on the matter” is enough to raise the question whether one is in a certain mental state. And that, along with the particular kind of rationality discussed earlier (pp. 324-5) would be sufficient for one to be aware that one is in that state.

But reflection will not, in general, raise the question of whether one is in a particular mental state unless, independently of reflection, one already has access to the state. If one’s access to the state is characteristically third-person, reflection cannot transform that into first-person access, since that would mean making one unaware of the relevant third-person considerations. And it begs the question at hand to invoke reflection if one must already have first-person access.

According to Wittgenstein, Moore’s paradox is absurd because “the statement ‘I believe it’s going to rain’ has a meaning like, that is to say a use like, ‘It’s going to rain’.” In effect, this is to construe the speech act ‘I believe it’s raining’ as expressing my belief that it’s raining, rather than as reporting that belief. This diagnosis of Moore’s paradox can be seen as having consequences about self-knowledge. If I make no claim about my mental states when I say ‘I believe it’s raining’, my remark cannot be challenged on that score. And that would help us understand the air of incorrigibility such remarks have, even though they would not then express knowledge about one’s beliefs.

This view notoriously faces serious difficulties. On it, for example, I cannot literally deny another person’s claim that I don’t believe that it’s raining. Shoemaker’s argument, by contrast, appeals to no such diagnosis, and he countenances both genuine self-knowledge and the semantic difference between “I believe it’s raining” and “I think that p”. Still, it’s not clear exactly what the semantic difference between these two amounts to if conceptual competence alone is enough to take one automatically from one to the other.

I’ve argued that Moore’s paradox sheds light on our first-person access by pointing to those factors which explain why all verbally expressed intentional states are conscious. But as I also mentioned, there’s an exception to that regularity. Suppose I assert ‘I think it’s raining’. Speech acts express intentional states with the same content; so here my speech act expresses my thought that I think it’s raining. It expresses, that is, a higher-order thought to the effect that I have the thought that it’s raining. But that’s not the thought I’m typically aware of having when I say I think it’s raining. Rather, when I say I think it’s raining, the thought I’m ordinarily conscious of is my first-order thought that it’s raining. When I perform a speech act whose content is that I’m in some intentional state, the state that’s conscious is not the state my speech act expresses, but the state
it reports.

This is an important exception. It’s doubtless this kind of case that has encouraged some to assimilate the reporting of mental states to the verbal expression of those states. As I remarked earlier, Moore’s paradox helps resist that tendency; if reports and verbal expressions weren’t different, Moore’s paradox would be an actual contradiction, which it isn’t. Still the exception demands explanation. That, however, is a task for another occasion.

Notes

†This is an expanded version of a paper read at the American Philosophical Association, Pacific Division, Symposium on Moore’s Paradox and Self-Knowledge, April 1994, in Los Angeles. The shorter version appears, slightly revised, in Philosophical Studies (77, 2-3 (March 1995): 195-209), under the title “Self-Knowledge and Moore’s Paradox,” along with the contributions of the other symposiasts, Sydney Shoemaker, “Moore’s Paradox and Self-Knowledge,” 211-228, and Rogers Albritton, “Comments on ‘Moore’s Paradox and Self-Knowledge,’” 229-239.


Rogers Albritton has urged that what Wittgenstein regarded as paradoxical was not that such sentences as ‘It’s raining but I don’t think it is’ are troubled, but that they are troubled even though counterparts not in the first-person present tense are not. (“Comments,” p. 229. This view recalls Remarks on the Philosophy of Psychology, vol. I, §490.) But even if that’s so, we could understand this paradoxical contrast only if we first understand what’s defective about the first-person, present-tense case, and the correct account of those might then make the contrast no longer seem paradoxical.


4. Roy A. Sorensen has argued that we can assert certain indirect versions of Moore’s paradox, e.g., ‘The atheism of my mother’s nieceless brother’s only nephew angers God’, which implies ‘God exists but I believe that God does not exist’. (See his
Blindspots [Oxford: Oxford University Press, 1988], p. 28; I am grateful to Sorensen for calling this argument to my attention.) But one can assert Sorensen’s sentence only if one doesn’t fully understand its content; it’s as though one asserts a certain sentence to be true, without really asserting its content.

5. As Wittgenstein in effect noted and, more explicitly, Albritton.


8. Fourth Replies, Adam and Tannery, VII, p. 232. Descartes is aware that his reasoning establishes transparency only with respect to what mental attitudes our intentional states exhibit, and not also with respect to the content of those states, and thus explicitly qualifies his claim in the quoted passage (see also VII, pp. 246-246).


13. Shoemaker evidently also thinks that the only alternative to using our concepts of human folk psychology would be to use concepts of subpersonal states “describable in the terminology of neurophysiology” (“Rationality and Self-Consciousness,” p. 147). But that’s far from obvious. We can avoid discredited aspects of our commonsense conception of mentality without restricting ourselves to subpersonal description; we could simply strip the unwanted features away. Similarly, if human folk-psychological concepts embody parochial features special to human mentality, we may well be able to subtract them, and use the resulting concepts to talk equally well about both our mental states and those of other creatures.

14. Nor, pace John Koethe, are sentences of the form ‘p but I don’t know that p’, of which the target sentence is a substitution instance, since I may unequivocally believe that p even if I don’t know it. (See Koethe, “A Note On Moore’s Paradox,” Philosophical Studies, 34 (1978): 303-310, p. 303.)


16. Knowing can be either active or latent; active knowing is conscious, and latent knowing is, roughly, being disposed to have active knowledge. It’s typically not
Moore's Paradox and Consciousness / 331

noted that only when the relevant knowing is uniformly conscious is there intuitive
appeal to the claim that knowing implies knowing that one knows; when knowing
is latent, that intuitive appeal vanishes. This reinforces the suggestion that knowing
appears to imply second-order knowing only because one cannot coherently
represent oneself as knowing while at the same time denying that one know that one
knows.

17. So on my view the trouble does not result, as Albritton suggests, from one’s being
conscious of the belief the first conjunct purports to express.

18. Indeed, it’s likely that when we describe intentional states in terms of the content
and illocutionary force of speech acts that might express those states, we draw more
fine-grained distinctions among those states than their nature warrants. We project
distinctions among our words back onto the intentional states those words express,
even though nothing about the intentional states themselves would allow us to
distinguish them so finely.

19. There’s no circularity in this last qualification, since we’re explaining what it is for
a state to be conscious in terms of what it is we’re conscious of. A state is conscious
just in case one is conscious of being in that state, and conscious of that in a way
that’s independent of any inference of which one is, in turn, conscious.

20. I think that p
\textsuperscript{1}
 can, of course, be used to qualify the firmness of one’s conviction
that p is the case, in ways that simply asserting \(p\) doesn’t.

21. I take that qualification for granted in what follows.

22. As with Paul Grice’s “natural” meaning; see his “Meaning,” The Philosophical
Review LXVI, 3 (July 1957), 377-88.


24. Many speech acts telegraph their insincerity; what’s allegedly different about
Moore’s paradox is that it does this solely as a result of its semantic properties, and
not, e.g., because of the way it’s uttered.

25. Paul Grice, “Utterer’s Meaning and Intentions,” The Philosophical Review,
LXXVIII, 2 (April 1969): 147-177. Mellor endorses this explanation. See
(1977-78): 87-101, pp. 96-7, and “Consciousness and Degrees of Belief,” in his
38. See also M.F. Burnyeat, “Belief in Speech,” Proceedings of the Aristotelian

26. For more shortcomings of the appeal to sincerity, and of a Gricean explanation, see
Jay David Atlas, “What is Paradoxical about Moore’s Paradox?: Rationality,
Sincerity, Implicature, and the Self-Limitations of Belief,” Claremont, California:
Pomona College, MS.

27. Gottlob Frege, “Thoughts,” in Logical Investigations, translated by P.T. Geach and

28. Cf. J.L. Austin’s claim that if I insincerely say ‘I promise’ I don’t strictly speaking
promise, but only say I do (“Other Minds,” in Philosophical Papers, second edition
Similarily, Black notes that “[a] man who lies is trying to deceive his hearers by
behaving like somebody who makes an honest assertion.” Thus “the making of an
utterance [in a certain tone of voice] in the absence of the corresponding knowledge
or belief is properly treated as a violation of the language” (“Saying and
Disbelieving,” pp. 116-7). I argue this at some length in “Intentionality,” §V, and
the postscript to it in *Rerepresentation: Readings in the Philosophy of Mental Representation*, ed. Stuart Silvers (Dordrecht: Kluwer, 1989), pp. 341-344 ($\S V$ of the article was radically truncated there).

29. One might argue that, even if ‘assert’ had such a meaning, the sentence isn’t an outright contradiction because it doesn’t actually use the word ‘assert’. Compare one’s at once running and saying ‘I’m not moving fast’; there’s no contradiction there, even though ‘run’ in part means to move fast. Similarly with ‘It’s raining but I’m not asserting it’, which purports to combine the act of asserting with the denial that I’m performing that act.

But these considerations are not decisive. If ‘assert’ meant in part that one thinks the relevant content, then the sentence ‘I assert that it’s raining but I don’t think it is’, which uses both ‘assert’ and ‘think’, would be a contradiction. That it’s instead an instance of Moore’s paradox argues against the meaning hypothesis.

30. It’s worth noting that we speak not only about speech acts’ expressing intentional states, but also about their expressing abstract thoughts, i.e., propositions. And we sometimes describe both in the same terms. Thus to say an assertion expresses a particular belief may mean either that it expresses an intentional state or that it has a certain content and an assertoric illocutionary force. Because Moore’s paradox is cast in terms of the speaker’s intentional states, it’s only the expressing of those states that’s relevant here.


32. E.g., Heal, who bases her claim that a common explanation is required on her observation that “the idea that someone realises the sentences to be true of him or herself” is just as strange as the idea of somebody’s asserting such a sentence (p. 6). But this misrepresents the purely mental version of Moore’s paradox; thinking that $p$ is not the same as thinking that the sentence ‘$p$’ is true of something. Thinking a sentence is true is tantamount to considering asserting that sentence; one can’t think the sentence ‘$p$ but I don’t believe it’ is true of oneself, precisely because one can’t consider asserting it. By contrast, the mental version of Moore’s paradox has to do with my thinking the content that $p$ but I don’t believe it, independently of any overt sentences. And it may be possible to do that even if it’s impossible to think that the corresponding sentence is true.

Shoemaker argues that explaining what’s wrong with thinking Moore’s paradox would automatically explain what’s wrong with saying it (“Moore’s Paradox and Self-Knowledge,” §1). But this ignores the possibility, argued for here, that believing Moore’s paradox is less defective than asserting it; asserting something may well require more than just believing it.


In section I, I argued that the factors that explain Moore’s paradox also underlie the *cogito*. But Descartes’s demon can cause doubt by undermining rationality. So, if the impossibility of thinking Moore’s paradox rests on rationality and the *cogito* relies on the same factors, then there is some question about whether the *cogito* can, after all, effectively resist the Demon.

34. “On Knowing One’s Own Mind,” pp. 188-191; see also “Rationality and Self-Consciousness,” p. 126, and “First-Person Access,” *Philosophical Perspectives*, 4,
Moore’s Paradox and Consciousness

36. If self-knowledge could be latent (see fn. 16), then rationalizing dispositional states might yield latent self-knowledge. It’s likely, however, that only active, conscious knowing can count as self-knowledge. Similarly, nothing counts as first-person access unless we’re conscious of having the access.
37. The most extensive development of this provocative argument occurs in “On Knowing One’s Own Mind,” esp. pp. 193-198.
38. This conclusion recalls a remark of Wittgenstein’s, though he and Shoemaker rely on very different arguments. Wittgenstein claimed that “it is possible to think out circumstances in which...someone [could] say ‘It is raining and I don’t believe it’.” Such a situation, he thought, would be one in which one could say: “Judging from what I say, this is what I believe.” But I could say this, he insisted, only if my “behavior indicat[ed] that two people were speaking through my mouth” (Philosophical Investigations, II, p. 192). In effect, he held, one cannot be a unified individual and also be in a position to say such a thing, and thus to judge from the outside what beliefs one has.
40. “On Knowing One’s Own Mind,” p. 194. Elsewhere he describes a plainly conscious inference that would take the putatively self-blind individual from believing that $p$ to having a motive for saying, indifferently, ‘$p$’ or ‘I believe that $p$’. Arguably, Shoemaker begs the question there, since the individual’s conscious reasoning from its having the belief that $p$ presumably presupposes that the belief is conscious.
42. Philosophical Investigations, II, p. 190. Indeed, Wittgenstein understands Moore’s paradox exclusively in terms of use:

Moore’s paradox can be put like this: the expression “I believe that this is the case” is used like the assertion “This is the case”; and yet the hypothesis that I believe this is the case is not used like the hypothesis that this is the case (p. 190).

43. Cf. Wittgenstein’s better known claim that saying ‘I’m in pain’, like crying ‘ouch’, expresses rather than reports my pain (Philosophical Investigations, §§244, 256). And if no speech act reports our states of pain, we cannot be said to have knowledge of them (§246). Here the sense of incorrigibility is stronger, since one can deny ‘It’s raining’, but not ‘ouch’.
44. Or that I’m not in pain. Moreover, as noted above, collapsing the distinction between reporting and expressing leads to construing Moore’s paradox as an outright contradiction.
45. I address this question in “Why Are Verbally Expressed Thoughts Conscious?”, Report No. 32/1990, Center for Interdisciplinary Research (ZiF) University of Bielefeld.