

## Why Are Verbally Expressed Thoughts Conscious?

### I. THE PROBLEM

Suppose I think it's going to rain, and leaving my home I take my umbrella. My act of taking my umbrella expresses my thought that it's going to rain.

One way this can all happen is that I consciously think it's going to rain and deliberately get my umbrella to take with me. But it need not happen this way. I may be late and preoccupied with what I must do that day. So I may reach for my umbrella and take it with me automatically—without thinking, as we might say. In this case I may not even be aware of my thought that it's going to rain; it may be a nonconscious thought. Nonconscious thoughts often cause appropriate actions, and when that happens the action typically expresses the thought that causes it.<sup>1</sup>

Now suppose again that I think it's going to rain. But this time, instead of performing some action, I simply tell you that it's going to rain. Here my thought that it's going to rain is expressed not by any action, but by my speech act. I may even tell you this in an offhand, automatic way; still, if I express my thought by performing some speech act, the thought I thereby express is always a conscious thought.<sup>2</sup> In this respect, verbally expressing a thought differs strikingly from expressing that thought nonverbally. However offhandedly I may tell you that it's going to rain, the thought I thereby express is always a conscious thought.

There seems, however, to be an exception to this generalization. Let's alter the case so that I tell you not that it's going to rain, but rather that I think it's going to. It is natural to take this speech act to express a certain sort of higher-order thought, a thought whose content is about another thought. Here, my speech act expresses not the thought that it will rain, but rather the thought that I think that it will rain.

<sup>1</sup> Whenever a piece of nonverbal behavior expresses a thought, it's appropriate to engage in the nonverbal behavior, given the thought and the thought's causing the behavior. But these two conditions may well not be sufficient for a piece of behavior to express a thought; perhaps not every appropriate action caused by a thought expresses that thought. It won't matter to the following argument what additional conditions might be needed for a nonverbal action to express a thought.

<sup>2</sup> This commonsense generalization has largely gone unnoticed; D. H. Mellor does, however, observe that this connection holds for the special case of sincere speech acts. See D. H. Mellor, "Conscious Belief," *Proceedings of the Aristotelian Society*, New Series, LXXXVIII (1977–8): 87–101, p. 96; and "Consciousness and Degrees of Belief," in Mellor, ed., *Prospects for Pragmatism: Essays in Memory of F. P. Ramsey*, Cambridge: Cambridge University Press, 1980, 139–173, p. 148.

One commonsense reaction to these observations may be that they are due simply to the way one's speech directs attention to one's own thoughts. When I say something, my very act of saying it calls attention to the fact that I think that thing. Even if my thought that it's going to rain had not been a conscious thought, my saying it's going to rain would have made it conscious by directing my attention to what it is I think. On this account, that is all there is to the connection between my saying something and the consciousness of the thought I thereby express.

This reply suggests answers to the other two questions as well. The reason my thoughts need not be conscious even when I express them nonverbally is that nonverbal behavior does not always direct my attention to the thought that lies behind the behavior. And the reason my higher-order thoughts may fail to be conscious even when I express them verbally is that my saying that I think it's going to rain normally calls attention not to my higher-order thought, but to the thought that higher-order thought is about. Typically, my saying that I think it's going to rain calls attention to my first-order thought that it's going to rain, rather than to my higher-order thought that I think it's going to rain.

## II. THE NAIVE ANSWER

In section II, I consider a naive answer to these questions and argue that it is inadequate. I conclude that the facts just noted demand an explanation that relies essentially on the nature of consciousness and the verbal expression of thoughts. In section III, then, I advance a theory of what it is for a mental state to be a conscious state, and an argument for that theory that reveals important connections between such consciousness and verbal expression. In section IV, I show how this theory, and the argument that supports it, allow us to explain all three observations that connect the expression of thought with consciousness. I conclude in section V by placing that solution in the context of some general observations about consciousness, thought, and speech.

It might of course be that my higher-order thought—my thought, that is, that I think it's going to rain—is a conscious thought. But it seems clear that in this kind of case it need not be. We often say that we think it's going to rain even when we are wholly unaware of having any such higher-order thought. Verbally expressing such higher-order thoughts seems thus to be an exception to the generalization that whenever I express my thoughts in speech, those thoughts are conscious thoughts. These observations apply not only to assertoric thoughts, such as the thought that it's going to rain, but to any intentional state, whatever its mental attitude. Thus the observations apply, for example, to doubting, desiring, expecting, anticipating, wondering, suspecting, and the like.

Why must verbally expressed intentional states be conscious? And why does this generalization hold for the verbal expression of intentional states, but not when those states are expressed solely by nonverbal means? Why, finally, does the generalization break down in the case of higher-order thoughts of the sort just described? It is these questions I address in this paper.

This last idea suggests, moreover, that we may have falsely posed the problem about the expressing of higher-order thoughts. Saying it's going to rain normally has the same force as saying that I think it's going to, setting aside whatever slight additional hesitation the second formulation may convey. Bracketing such hesitation, which won't matter in what follows, the two assertions have what we can call the same performance conditions—the same conditions in which it would be appropriate to make the assertions. So perhaps the two remarks express the very same thought: the thought that it's going to rain. If so, then when I say 'I think it's going to rain', I do not, after all, express my thought that I think it's going to rain, but instead express only my thought that it's going to rain. There is then no problem about why I can say I think it's going to rain even when I have no relevant conscious higher-order thought.

These naive answers are inviting, and have the merit of proceeding independently of any special theories about consciousness, verbal expression, and their connection. And as a general rule, the more we can explain while remaining neutral about potentially controversial theories, the better.

Nonetheless, the naive answers will not do. After all, why does speaking call attention to the thought expressed, while nonverbally expressing the thought does not? Unless we can explain that, the naive approach does little more than postpone giving an informative solution to our problem. Indeed, in the absence of such further explanation, it may well be that we have simply redescribed in different terms the observations with which we began. After all, what is there, in the current context, to something's calling attention to a thought other than something's making that thought conscious? The naive answer might help if recasting things in terms of directing attention to the thought itself suggested a substantive explanation. But it's notorious that what is involved in shifting attention is at best no more tractable than consciousness.

As with nonverbal actions, we perform speech acts in ways that range from relatively automatically to fully deliberately, and the verbal expressing of intentional states is no exception. If one expresses an intentional state in a relatively deliberate way, that may indeed go hand in hand with one's shifting one's attention to that intentional state. But most speech acts that verbally express our intentional states are far from deliberate. Rather, they occur without any attentive focus on the thought we are expressing; ordinarily if we are concentrating on anything at all, it's the subject matter at hand, and perhaps the goal of informing, convincing, or otherwise affecting another person. So there's no reason to expect that such nondeliberate, automatic speech acts will call attention to the thoughts they express. But verbally expressed thoughts are conscious whether the verbal expressing is deliberate or instead automatic and unthinking.

An advocate of the naive answer might reply that speech acts are never as automatic as nonverbal behavior often is. We do sometimes perform nonverbal actions without being aware of doing so, but when we say something we are invariably conscious of saying that thing. So we are conscious, at least to some degree, of the thought our speech act expresses. But this again simply puts off the real explanation; why are we always conscious of our speech acts? Speech acts are, after all, kinds of actions; they

are actions of asserting, asking, demanding, predicting, retracting, and the like. If we can perform nonverbal actions, such as taking umbrellas, without being conscious of doing so, why isn't that also possible for verbal actions?

These considerations suggest that no naive answer of the sort I just proposed will do. We need an account that goes deeply enough into the nature of expressing to reflect the relevant difference between verbal and nonverbal expressions. And we need an account that goes far enough into the connection between expressing and consciousness to show how the difference between the two kinds of expressing bears on what we must in each case be conscious of.

The naive answer fails also to explain why the higher-order case, in which one says that one thinks that it's going to rain, rather than simply that it's going to rain, is an exception. We cannot explain why verbally expressed thoughts are conscious but not when expressed nonverbally by appeal to whether or not the expression calls attention to the thought. So we cannot then go on to invoke the more specific explanation that my thinking it's going to rain calls attention to my thought that it will rain, rather than to my higher-order thought that I think it will.

But in this case there is a related naive answer sketched earlier. Perhaps saying 'I think it's going to rain' does not, after all, express the higher-order thought that I think it's going to rain, but rather expresses only the first-order thought that it's going to rain. Can this view be sustained?

As already noted, the remarks 'It's going to rain' and 'I think it's going to rain' have the same performance conditions; any circumstances in which it would be appropriate to say one thing would also be circumstances in which it would be appropriate to say the other. But the two do not at all have the same semantic properties. They mean different things, and have distinct truth conditions.

Similar observations hold of the thought that it's going to rain and the thought that I think it is. The circumstances in which it would be appropriate to think these two thoughts are the same; in effect, the mental analogue of performance conditions for speech acts is the same for the two thoughts. But, as with the corresponding speech acts, the two thoughts have distinct truth conditions. And on any commonsense account their content is distinct.<sup>3</sup> Moreover, the distinct contents of this pair of thoughts presumably matches the semantic properties of the corresponding pair of speech acts. This gives us reason to insist that the remark 'I think that it's going to rain' expresses not the thought that it's going to rain, but the thought that I think it will.

Since we are evaluating the naive response to our problem, it is theoretically uninformed folk psychology that matters here.

Brian Loar has argued that the concepts in such a first-order thought will occur in the corresponding higher-order thought, and concludes that the conceptual role of the lower-order thought is included in that of the higher-order thought ("Subjective Intentionality," *Philosophical Topics*, XV, 1 [Spring 1987]: 89–124, p. 103). But a higher-order thought will include more concepts than the first-order thought; it will, in addition to those they share, have concepts that pertain to one's having a thought. So the conceptual role of the higher-order thought will be distinct from that of the first-order thought it is about.

There is another way to reach the same result. When a speech act expresses an intentional state, not only are the contents of both the state and the speech act the same; the speech act and thought also have the same force. Both, that is, will involve suspecting, denying, wondering, affirming, doubting, and the like. Whenever a speech act expresses an intentional state, the illocutionary force of the speech act corresponds to the mental attitude of that intentional state.

When the thought in question is assertoric, considerations of force do not matter; the thought that it's going to rain and the thought that I think it is have the same force. But nonassertoric intentional states are expressed by speech acts that have the corresponding, nonassertoric illocutionary force. Suppose I doubt it's going to rain. If I then say 'I doubt it will rain', my speech act is assertoric; it asserts that I doubt that thing. My speech act must therefore express the assertoric thought that I doubt it will rain; it cannot express the nonassertoric intentional state of my doubting. So we cannot explain why verbally expressed higher-order thoughts need not be conscious by insisting that ostensibly higher-order speech acts do not actually express higher-order thoughts, but only lower-level thoughts.

Every speech act expresses an intentional state that has the same content and a mental attitude that corresponds to that speech act's illocutionary force. Indeed, it's typical to understand the content and mental attitude of intentional states by reference to the content and illocutionary force of the speech act that would best express those states. The correspondence of the content and illocutionary force of speech acts with the content and mental attitude of intentional states is built into the very way we ascribe content to intentional states.<sup>4</sup>

### III. CONSCIOUSNESS AND HIGHER-ORDER THOUGHTS

Since the naive answer fails, we must try to find a more probing explanation of both the generalization about the consciousness and verbal expression of thoughts and the exception about higher-order thoughts. The best hope for this is an explanation that addresses what it is for a thought to be conscious. Only then can we expect to understand the connection between verbally expressing a thought and that thought's being conscious.

Conscious states are mental states we are conscious of. There are two models for what it is to be conscious of things; we are conscious of something when we sense that thing and when we think about it as being present. The sensory model creates

<sup>4</sup> For more on the way the properties of speech acts correspond to those of the intentional states they express, see "Intentionality," ch. 3 in this volume. For more on the way we interpret the mental properties of intentional states by reference to the speech acts that express them, see "Content, Interpretation, and Consciousness," ch. 12 in this volume.

Zeno Vendler has an extensive and highly illuminating treatment of the parallel between intentional states and the speech acts that express them in *Res Cogitans*, Ithaca, New York: Cornell University Press, 1972, ch. 3. The idea that we understand the intentional character of mental states in terms of the semantic character of the speech acts that would express those states is arguably the best explanation of the striking parallels that Vendler details.

unnecessary difficulties in this context; sensing always involves some mental quality, and being conscious of a mental state involves no qualities above and beyond those of the mental state we are conscious of. The only alternative is that a mental state's being conscious consists in its being accompanied by a higher-order thought (HOT) to the effect that one is in that very mental state. These HOTs must be assertoric, since nonassertoric intentional states do not, by themselves, make one conscious of things. And they must also be occurrent, since having a dispositional thought is being disposed to have an occurrent thought, and merely being thus disposed again doesn't make one conscious of anything.<sup>5</sup>

A compelling argument is available that supports this conclusion. This argument relies on independently defensible premises about what it is to express and to report a thought.<sup>6</sup> So the argument points the way to building a bridge between consciousness and the expressing of thoughts. I develop the argument here in three steps, each of which connects with the issues raised above about expressing one's intentional states.

### Expressing One's Mind

I begin with the connection between speech acts and the intentional states they express. Sincere, meaningful speech acts invariably express intentional states. Whenever one says something sincerely and meaningfully, one is in some intentional state that has the very same content as one's speech act (or at least very similar; see chapter 12). Moreover, as noted in section II, the illocutionary force of every sincere speech act corresponds to the mental attitude of the intentional state expressed by that speech act. But it is sameness of content that matters for our purposes here.

The well-known phenomenon of Moore's paradox provides evidence for the conclusion that no sincere speech act can occur unless the speaker is in an intentional state whose content coincides with that of the speech act. As Moore noted, sentences such as 'It's raining but I don't think so', though not contradictory, are nonetheless absurd.<sup>7</sup> They are absurd in that no such sentence can be used to perform a coherent speech act. Such a sentence has no coherent performance conditions that would allow

<sup>5</sup> In "Two Concepts of Consciousness" (ch. 1 in this volume) I argue that this kind of account can save out folk-psychological intuitions at least as successfully as any alternative theory; in "A Theory of Consciousness" (in *The Nature of Consciousness: Philosophical Debates*, ed. Ned Block, Owen Flanagan, and Güven Güzeldere, Cambridge, Massachusetts: MIT Press/Bradford Books, 1997, pp. 729–753) and "Explaining Consciousness" (in *Philosophy of Mind: Classical and Contemporary Readings*, ed. David J. Chalmers, New York: Oxford University Press, 2002, pp. 406–421), I show that independently well-motivated constraints on a satisfactory theory require us to adopt this one. In "Varieties of Higher-Order Theory," in *Higher-Order Theories of Consciousness*, ed. Rocco J. Gennaro, Amsterdam and Philadelphia: John Benjamin Publishers, 2004, pp. 17–44, I argue that Mellor puts forth a somewhat similar account ("Conscious Belief" and "Consciousness and Degrees of Belief"), though he argues that it applies only to conscious believing, and not to mental states generally.

<sup>6</sup> As above, I use 'thought' generically to cover all intentional states, regardless of their mental attitude.

<sup>7</sup> G. E. Moore, "A Reply to My Critics," in *The Philosophy of G. E. Moore*, ed. Paul Arthur Schilpp, New York: Tudor, 1942, 2nd edn, 1952, pp. 533–677, at p. 543; and "Russell's Theory

for the performing of any speech act. We can best explain this lack of performance conditions on the assumption that all sincere speech acts express intentional states that have the same force and intentional content. The sentence 'It's raining but I don't think so' is absurd because the corresponding speech act would purport to express an intentional state that it's raining and also purport to deny that one is in that state.<sup>8</sup>

It is even arguable that, unless a speech act expresses a thought that the speaker actually has, it is a degenerate kind of speech. Insincere speech, like lines uttered in play acting, is basically performed by rote; in both kinds of case one pretends to perform normal speech acts. When I play the part of Hamlet, the lines I utter express no actual thoughts; rather, the audience and I both pretend that my lines express thoughts of my fictional character. Similarly, when I speak insincerely, I pretend that my utterances express the thoughts of the fictional person I am pretending to be—roughly, a person just like me except for believing the things I pretend to say.<sup>9</sup> This is why, in insincere speech and play acting alike, one needs to depend more on one's exact words to remember what one has said, since there is no thought behind

of Descriptions,' in *The Philosophy of Bertrand Russell*, ed. Paul Arthur Schilpp, New York: Tudor Publishing Company, 1944, pp. 175–226, at p. 204.

See also "Moore's Paradox and Consciousness," ch. 8 in this volume, my "Moore's Paradox and Crimmins's Case," *Analysis*, 62, 2 (April 2002): 167–171, and "The *Mediations* and Moore's Paradox," MS.

<sup>8</sup> Parallel remarks hold for speech acts that have nonassertoric illocutionary force; I cannot, e.g., coherently say 'Thank you, but I am not grateful'. (Pace Mellor, who claims that Moore's paradox "has no analogue for the other attitudes" [D. H. Mellor, "What Is Computational Psychology?," *Proceedings of the Aristotelian Society*, Supplementary Volume LVIII (1984): 37–53, p. 38]. It is notable that Mellor similarly restricts his account of consciousness to the case of believing; see n. 5, above.)

One might instead seek to explain the absurdity of these sentences by appeal to Gricean considerations. On a Gricean view, my sincerely saying something involves my intending that my hearer believe that I believe what I say. (See, e.g., Paul Grice, "Utterer's Meaning and Intentions," *The Philosophical Review*, LXXVIII, 2 [April 1969]: 147–177.) So I cannot at once sincerely say both that *p* and that I do not believe it. (See Mellor, "Conscious Belief," pp. 96–7; cf. "Consciousness and Degrees of Belief," p. 148.) But Moore's paradox is absurd independent of any context of communication; it is absurd because it lacks coherent conditions of assertibility. (For more on the Gricean view, see the Postscript to "Intentionality.")

If I say it's raining and go on to say I don't believe it, I betray my insincerity. One might object that this is all that goes wrong in Moore's paradox. Moore himself offered such a diagnosis at one point ("A Reply to My Critics," pp. 542–3). But that cannot be correct, since there is in general nothing problematic about one's speaking insincerely in ways that betray one's insincerity. One can even do so knowingly. This underscores the inadequacy of a Gricean explanation, since one could not on that account knowingly betray the insincerity of one's speech.

<sup>9</sup> Such pretend speech is, in Wilfrid Sellars' useful phrase, produced parrotingly ("Notes on Intentionality," *The Journal of Philosophy*, LXI, 21 [November 12, 1964], 655–665, p. 657).

On these issues, see "Intentionality," esp. §§II, III, V, and the Postscript.

Bruce Bridgeman has urged (personal communication) that there may be types of speech production intermediate between genuine speech acts and pretend speech, as when one reads a newspaper article aloud to somebody else. But the sense that such cases are intermediate very likely results simply from their occurring in both ways; though reading a newspaper aloud to somebody is sometimes merely parroting, it also sometimes occurs with genuine illocutionary force.

those words to rely on. But this point is unnecessary for the present argument. No more is needed here than that, when speech is sincere, it expresses intentional states whose content matches that of our speech acts.

Slips of the tongue and Freudian parapraxis may seem, however, to be exceptions to this rule. Such slips sometimes reveal thoughts markedly different from what one consciously meant to say, and perhaps when that happens the content of the spoken slip differs from that of the thought it expresses. But such slips are not counterexamples to the generalization that speech acts express thoughts with the same content. When such a slip occurs, the consciously intended content that matches the semantic meaning of the uncorrupted speech act differs from the content of the thought the slip reveals. But a remark can reveal a thought without thereby verbally expressing it, as when one's tone of voice in saying something reveals a thought distinct from, or even unconnected to, the thought that remark expresses. Slips of the tongue are like that; they reveal thoughts beyond those they express. A slip may sometimes reveal a thought the speaker consciously intended to conceal, and in some cases of that sort the remark may verbally express no intentional state at all, but be a straightforward case of insincere speech.<sup>10</sup>

### Expressing and Reporting

The second step in the argument invokes the HOTs described earlier. Expressing is the most common and straightforward way to convey our intentional states to others. If I say "The door is open," my speech act conveys my thought that the door is open by expressing that very state. But this is not the only way to convey my thought. I could convey it equally well by saying instead "I think the door is open." The same goes for intentional states with mental attitudes other than that of belief. I can, for example, communicate my suspicion that the door is open either by expressing my suspicion or by explicitly telling you about it. Saying that the door may well be open would express the suspicion, whereas saying that I suspect the door is open would explicitly report that I have that suspicion.<sup>11</sup>

It is easy to conflate these two distinct ways of conveying our intentional states to others. As noted in section II, the performance conditions for asserting that *p* are the same as those for telling you that I think that *p*. Bracketing whatever hesitation one form of words may sometimes suggest, any conditions in which it's appropriate to

<sup>10</sup> For more on parapraxis, see "Consciousness and Its Expression," ch. 11 in this volume, §V. On this distinction, see, e.g., Sellars, "Empiricism and the Philosophy of Mind," in his *Science, Perception and Reality*, London: Routledge & Kegan Paul, 1963, pp. 127–196, §§viii, x, and xv, and Daniel C. Dennett, *Content and Consciousness*, New York: Humanities Press, 1969, §13.

<sup>11</sup> When subjects in psychological experiments are instructed to respond with particular nonverbal signals if they perceive something, it's not always obvious whether that nonverbal response constitutes a report of such perceiving or merely expresses it. Consider Anthony J. Marcel's report of hemianopic subjects who responded most sensitively to stimuli with an eye blink, less so with a button press, and least sensitively when they responded verbally ("Slippage in the Unity of Consciousness," in Gregory R. Bock and Joan Marsh, eds., *Experimental and Theoretical Studies of Consciousness*, Ciba Foundation Symposium No. 174, Chichester: John Wiley & Sons, 1993, pp. 168–180). It's tempting to explain this divergence by supposing that the eye blinks and button presses functioned as nonverbal expressions of subjects' perceptions, whereas the verbal responses were explicit reports.



say one thing would equally be conditions in which it's appropriate to say the other. Once again, the point generalizes beyond the case of assertoric intentional states. The conditions for appropriately expressing doubt, suspicion, wonder, and other mental attitudes are the same as those for explicitly reporting that one is in those intentional states.

Unreflective, colloquial usage may seem to suggest that when I say 'I think that  $p$ ', my speech act simply expresses my thought that  $p$ , and not my thought that I think that  $p$ . If I say 'I think it's raining', we normally take my speech act to be about the weather, not just about my mind. And if I say 'I doubt it's raining', doesn't my speech act express that doubt, rather than a HOT that I have the doubt? Similarly, it may seem that saying 'I choose this one' and 'I sympathize with you' express one's choice and sympathy, rather than merely reporting those states.

But these intuitions result from focusing solely on performance conditions, to the exclusion of the semantic properties of sentences.<sup>12</sup> As noted above, the speech acts one performs with the sentences 'I think that it's raining' and 'I doubt that it's raining' are performance conditionally equivalent to those one performs with the sentences 'It's raining' and 'It's probably not raining'. But these pairs plainly have distinct truth conditions, and so are semantically distinct. Similarly with the pairs 'I choose this one' and 'I'll take this one', and 'I sympathize with you' and 'That's too bad'.<sup>13</sup>

Once again Moore's paradox helps. Sentences such as 'It's raining but I don't think so' and 'It's surely raining but I doubt it' reveal the relevant differences in truth conditions. More generally, if the sentence 'I  $p$ , but I don't think that  $p$ ' could be used to perform a coherent speech act, that speech act would first express one's thought that  $p$  and then immediately go on to report that one had no such thought. So, if reporting and expressing were the same, then to say 'I  $p$ , but I don't think that  $p$ ' would in effect be both to report the thought and also to report its absence. Moore's paradox would then be not merely absurd, but an actual contradiction. Only if expressing thoughts is distinct from reporting them can we explain why Moore's paradox is not an actual contradiction.

### Reporting and Consciousness

The distinction between expressing and reporting one's intentional states has important implications for the question of what it is for mental states to be conscious. It's generally recognized that one can report a mental state one is in just in case that state is conscious. This connection has firm grounding in pretheoretic intuition, and is used

<sup>12</sup> Beginning students often take statements such as 'It's raining and it's not raining' to be meaningless, rather than meaningful but false. This too is best understood as a result of focusing on performance conditions, rather than the semantic meaning of sentences. It's tempting to see such sentences as meaningless because uttering such sentences would lack speaker's meaning.

<sup>13</sup> Saying that one sympathizes, doubts, or chooses has a performative aspect; as with 'I do', said in wedding ceremony, to say 'I choose' is to choose, and similarly with these other verbs. But these remarks also all succeed in making statements: that one chooses, doubts, or sympathizes. And the truth conditions of each statement again underscores that it reports a corresponding intentional state.

in experimental work as a reliable indicator of whether a mental state is conscious.<sup>14</sup> If a state is not conscious, it will be unavailable to one as the topic of a sincere report about the current contents of one's mind. And if it is conscious, one will be aware of it and hence able to report on it. The ability to report a particular mental state coincides with what we intuitively think of as that state's being in our stream of consciousness.

The connection between consciousness and reportability holds for creatures with the relevant linguistic abilities, such as ourselves. But that needn't limit our application of 'conscious state'. We can use the connection with reportability to fix the extension of the term 'conscious state' by reference to our own case. But what fixes the extension of 'conscious state' need not coincide with what is essential to such states.<sup>15</sup> So we can then discover what is essential to conscious states, and go on to apply the term also to creatures that lack the ability to report their mental states.

It is important to stress that the relevant reporting of one's mental states is intuitively unmediated. This is crucial to ensure the intuitive immediacy of consciousness. Reports of mental states reflect our consciousness of those states only if those reports do not appear to rely on inference or observation, only, that is, only if we are not conscious of the reports as relying on inference or observation. But excluding conscious inference is enough, since conscious observational mediation itself always relies on some inference of which one is aware.

But the ability noninferentially to report a particular mental state is the same as the ability to express verbally one's noninferential thought that one is in that state. So, for creatures with the relevant linguistic ability, a state's being conscious will coincide with one's having the ability verbally to express a noninferential HOT that one is in that state. The best explanation of one's having that ability is that the HOT one is able to express occurs, even if it is unexpressed. Given that explanation, a mental state's being conscious actually implies the presence of the very noninferential HOT that the current theory posits.

When a mental state is not conscious, we cannot report being in it. But that is equivalent to our being unable to express any HOT about the state in question. Bar-ring exceptional factors, we can perform a speech act when, and only when, we have a thought that has the relevant content and mental attitude, a thought that the speech act would express.

So the best explanation of our inability to report mental states when they are not conscious is that in those cases no relevant HOT occurs. Conscious states must be accompanied by suitable HOTs, and nonconscious mental states are not so

<sup>14</sup> See, e.g., Marcel, "Conscious and Unconscious Perception: Experiments on Visual Masking and Word Recognition," *Cognitive Psychology*, 15 (1983): 197-237; Lawrence, *Consciousness Lost and Found: A Neuropsychological Exploration*, Oxford: Oxford University Press, 1997; and Philip M. Merikle, Daniel Smilek, and John D. Eastwood, "Perception without Awareness: Perspectives from Cognitive Psychology," *Cognition*, 79, 1-2 (April 2001): 115-134, esp. p. 132.

<sup>15</sup> For more on the way that fixes the extension of 'conscious state' may depart from what's essential to such states, see "Two Concepts of Consciousness," §IV and "A Theory of Consciousness," §II.

accompanied. We can conclude that a mental state's being conscious consists in its being accompanied by a suitable HOT.<sup>16</sup>

Not all mental states are conscious. And the HOTs the theory posits will themselves seldom be conscious. When these HOTs are conscious, introspective consciousness results; we are introspectively conscious of a mental state when we are aware of being conscious of that state. But it's rare that we are aware of our HOTs. And that is to be expected; for a second-order thought to be conscious, we must have a yet higher-order thought about that second-order thought, and it's reasonable to assume that this seldom happens.<sup>17</sup>

It is in any case no objection that we are typically unaware of having any such HOTs. Indeed, the theory predicts that this would be so. Since those thoughts are rarely conscious, we are seldom aware of their presence: we come to be aware of them only in those rare cases in which we introspect.

The foregoing argument reinforces the conclusion of section II that, if I say 'I think it's raining', my speech act expresses my HOT that I think it's raining, and not simply my thought that it's raining. This is important for the present theory. There can be no doubt that saying that I think it's raining reports the thought that it's raining. If, in addition, that speech act also expressed that thought, the very distinction between expressing and reporting would collapse. We then could not infer from an intentional state's being conscious to the occurrence of a HOT.

Collapsing the distinction between reporting and expressing would thus undermine the foregoing argument for an account of consciousness in terms of HOTs. Indeed, it would undermine the account itself, since it would encourage a reflexive conception of consciousness on which consciousness is an intrinsic feature of every

<sup>16</sup> A HOT's being suitable requires that it have the content that one is in the state in question and that it not depend on any conscious inference.

Suppose that one has a HOT that one is not conscious of as depending on any inference, and one then comes to be conscious of an inference that leads to that HOT. That would not result in the target state's no longer being conscious so long as one continued to have the HOT independently of that conscious inference, i.e., so long as one would still have the HOT without that conscious inference.

<sup>17</sup> Third-order thoughts do occur when we introspect; can fourth-order thoughts also occur? There is reason to think so. Sometimes we are actually conscious of our introspecting, and that means having a fourth-order thought about the third-order thought in which such introspecting consists. And we may occasionally even consciously think that we're introspecting, which would take it a step higher.

Still, it seems implausible that thoughts with explicitly fourth-order content ever occur; a thought whose explicit content is that one has a thought that one has a thought that one is in some particular state would at best be difficult to process. But the concept of introspection helpfully short circuits the explicit hierarchy. The thought in virtue of which my introspecting is conscious very likely has the explicit content simply that I'm introspecting, since it's unlikely that such a thought explicitly represents the relevant first- or second-order states. Similarly, if such a thought is itself conscious, the explicit content of the HOT in virtue of which it is conscious is very likely simply that I have a thought that I am introspecting. We needn't countenance contents with the iterative character of explicitly fourth- or fifth-order thoughts.

On difficulties with iterative content, see n. 28, below.

I am grateful to Josef Perner (personal communication) for pressing me on this issue.

conscious state, rather than a distinct state of being aware of one's conscious states. Saying that one thinks that  $p$  expresses one's consciousness of one's thought that  $p$ .<sup>18</sup> So, if saying that  $p$  amounted to the same thing as saying that one thinks that  $p$ , expressing a state would be the same as expressing one's consciousness of that state, which would encourage identifying mental states with one's consciousness of them.

There is another way we can put the connection between consciousness and collapsing the distinction between reporting and expressing. It's tempting to think that we can express any intentional state we are in, and a mental state is conscious just in case we can noninferentially report it. And we can assume that if expressing and reporting were the same, reporting would be noninferential, and expressing would be verbal. So, if expressing and reporting were the same, every mental state, by being verbally expressible, would be noninferentially reportable, and hence conscious. Nor, then, would there be any problem about why a verbally expressed intentional state is always conscious, since in verbally expressing a state one would thereby also express one's consciousness of that state.

The reflexive conception of consciousness, on which the property of a state's being conscious is intrinsic to that state, was championed by Descartes and Brentano among others.<sup>19</sup> But it is also evident in Wittgenstein, who regards the idea of mental states that aren't conscious as involving a novel use of terms for such states.<sup>20</sup> Wittgenstein's reliance on performance conditions to the exclusion of such semantic considerations as truth conditions also leads him to hold that ostensible reports of mental states actually just express those states.<sup>21</sup> So it's tempting to think that his exclusive reliance on use and the consequent assimilation of reporting to expressing led to his rejection of mental states that aren't conscious.

But the exclusive reliance on performance conditions misleads. Considerations of mental attitude and illocutionary force, which undermined the claim that ostensibly higher-order remarks express first-order thoughts, also tell against the view that consciousness is intrinsic to mental states. Since the illocutionary force of 'I doubt it's

<sup>18</sup> At least when saying that one thinks that  $p$  does not rely on any conscious inference, which is the usual case.

<sup>19</sup> René Descartes, *Fourth Replies, The Philosophical Writings of Descartes*, tr. John Cottingham, Robert Stoothoff, and Dugald Murdoch (vol. III with Anthony Kenny), Cambridge: Cambridge University Press, 1984–91, II, 171; and *Second Replies*, II, 112.

Descartes also evidently collapses the distinction between expressing and reporting, as when he argues that nonlinguistic creatures have no thoughts because the lack of linguistic ability shows that they could not in any way even express thoughts; Letter to More, February 5, 1659, III, 366; the French is *exprimer*. See also *Discourse* (I, 140–1), and Letters to Marquess of Newcastle, III, 302–304, and to More, III, 364–367.)

And Franz Brentano, *Psychology from an Empirical Standpoint*, ed. Oskar Kraus, English edn., ed. Linda L. McAlister, tr. Antos C. Rancurello, D. B. Terrell, and Linda L. McAlister, London: Routledge & Kegan Paul, 1973 (original 1874), pp. 121–138.

<sup>20</sup> Ludwig Wittgenstein, *The Blue and Brown Books*, Oxford: Basil Blackwell, 1958, 2nd edn. 1969; pp. 22–3 in respect of sensations and pp. 57–8 in connection with thoughts.

<sup>21</sup> *Philosophical Investigations*, ed. G. E. M. Anscombe and R. Rhees, tr. G. E. M. Anscombe, Oxford: Basil Blackwell, 1953, §§244, 256, and 310 on sensations, and Part II, §x, p. 190 on thoughts.

going to rain' is assertoric, it cannot express the nonassertoric doubt that it's going to rain. Similarly, since no intentional state has two distinct mental attitudes, the assertoric HOT that I doubt that it's going to rain cannot be part of the nonassertoric doubt itself.<sup>22</sup>

#### IV. THE SOLUTION

The connections between consciousness and expressing just described provide the resources to answer the questions posed at the outset. Let me begin with the problem of why intentional states are conscious whenever we express them verbally.

As noted above, when I say that *p*, and thereby verbally express my thought that *p*, my speech act has the same performance conditions as those which would govern a speech act of saying that I think that *p*. Whenever it is appropriate to say that it's raining, it is also appropriate to say I think it is, and conversely.<sup>23</sup>

It is crucial for present purposes that this performance-conditional equivalence is not something we need to think through or figure out. Indeed, the ability we have to report our mental states noninferentially makes this equivalence wholly automatic for us. Whenever one says that it's raining, one could as easily have said that one thinks that it's raining, and conversely. In general, whenever we verbally express an intentional state, we might just as easily have reported that state, and whenever we noninferentially report an intentional state, we could just as easily have verbally expressed it. And, because reporting a state is no less automatic for us than verbally expressing it, any report of a state we make when we might equally have verbally expressed that state will be noninferential. We insensibly slip from saying one thing to saying the other, and indeed we often recall incorrectly which of the two we did say. The performance-conditional equivalence of reporting and verbally expressing is a matter of well-entrenched linguistic habit.

Whenever one says anything at all, one expresses an intentional state with the content of one's speech act. But, so far as performance conditions alone are concerned, saying that *p* amounts to the same thing as saying that one thinks that *p*, and this

<sup>22</sup> The foregoing section briefly summarizes the argument of "Thinking that One Thinks," ch. 2 in this volume.

For more on the idea that reporting intentional states is the same as expressing them, see §IV of that chapter. For reasons to think it's unlikely that an intrinsic theory allows an informative explanation of consciousness, see "A Theory of Consciousness," §III.

Chris D. Frith and Uta Frith cite brain-imaging studies in which subjects' being asked to report their mental states results in cortical activation in medial frontal cortex. The states reported are of many different sorts, and the cortical areas that subserve those states are all distinct from medial frontal cortex. Since reporting a mental state expresses a HOT in virtue of which one is conscious of that state, these findings suggest that such consciousness, and the higher-order content that figures in it, are extrinsic to the states themselves. See "Interacting Minds—A Biological Basis," *Science*, 286, i5445 (November 26, 1999): 1692–1695, p. 1693.

<sup>23</sup> If saying or thinking that one thinks it's raining does indicate a measure of hesitation not present in simply saying or thinking that it's raining, saying that *p* performance conditionally implies saying that one thinks that *p*, but not conversely. As noted earlier, this would not affect the present argument.