

## *Thinking that One Thinks*

---

DAVID M. ROSENTHAL

There are two distinct kinds of thing we describe as being conscious or not conscious, and when we describe the two kinds of thing as being conscious we attribute two distinct properties. The term 'conscious' thus conceals a certain ambiguity.

In one use, we speak of mental states as being conscious or not conscious. Mental states, such as thoughts, desires, emotions, and sensations, are conscious if we are aware of them in some intuitively immediate way. But we also apply the term 'conscious' to the creatures that are in those mental states. A creature's being conscious consists, roughly, of its being awake and being mentally responsive. Plainly, this property of being conscious is distinct from the property a mental state may have.

It is the notion of a mental state's being conscious that occasions such difficulty in understanding what consciousness amounts to. It is the consciousness of mental states, as Thomas Nagel (1974) points out, that makes understanding the nature of those states seem difficult, or even impossible. If we bracket or ignore the consciousness of mental states, the problem of how to explain their nature will no longer seem intractable. Our explanation will then proceed simply in terms of the intentional or sensory content mental states have, without reference to their being conscious. Because the notion of consciousness that applies to mental states is the more difficult to explain, it is that notion which I shall be concerned with in what follows.

By contrast, no special problems impede our understanding what it is for a creature to be a conscious creature. A creature's being conscious means that it is awake and mentally responsive. Being awake is presumably an unproblematic biological notion. And being mentally responsive amounts simply to being in some mental states or other. That will occasion no special difficulty unless those states are themselves conscious states, and if they are we can trace the difficulty to the notion of a mental state's being conscious, rather than a creature's being conscious.

It is possible, however, to dispel our sense that special difficulties face any explanation of what it is for mental states to be conscious. The sense that such consciousness is somehow intractable derives at bottom from the tacit, and unnecessary, assumption that all mental states are conscious states. If being a conscious state did coincide with being a mental state, we would then be unable to rely on any prior account of mentality in trying to explain what it is for mental states to be conscious. For if the concept of mind we started from had consciousness already built in, the resulting explanation would be circular, and if it did not, our explanation would rest upon a conception of mentality that, by hypothesis, is defective. There is no third way; we plainly can explain consciousness only in terms of what is mental. So if mental states are all conscious, no informative, non-trivial explanation of such consciousness is possible.

This result perfectly suits Cartesian theorists. If we can give no informative explanation of consciousness, the gulf that intuitively separates mind and consciousness from the rest of reality will seem impossible to bridge. Our explanations will thus be limited to tracing the conceptual connections holding among such terms as 'mind', 'consciousness', 'subjectivity' and 'self'. Cartesians such as Nagel (1974) and Colin McGinn (1991) encourage us to embrace this limitation by evoking that sense we have of ourselves on which consciousness is the most central feature of our existence. If consciousness were the most basic aspect of our nature, why should we expect to be able to explain it in terms of anything more basic?

The sense that consciousness is thus basic is closely tied to the idea that mental states are all conscious states. And if mental states are not all conscious, the foregoing difficulty dissolves. We can then seek first to explain the nature of those mental states which are not conscious, and build on that understanding of non-conscious mental states to arrive at an informative account of what it is for a mental state to be conscious. In particular, if consciousness is not essential to a state's being a mental state, we can reasonably identify a state's being mental with its having either intentional or sensory character. And this account does not presuppose that such states are conscious. A mental state's being conscious, moreover, is our being aware of that state in a suitably immediate way. So we can then go on to argue that a mental state's being conscious is its being accompanied by a roughly simultaneous higher-order thought about that very mental state. On this account, not all mental states are conscious, and we can explain how the conscious ones differ from those which are not. It is this hypothesis which I shall defend here.

On this account, we explain the property of a mental state's being conscious in terms of our being conscious of that state. In general, being conscious of something means having a thought about it or a sensation of it. One may be conscious of a chair, for example, by thinking something about the chair or by having some sensation of it. Sensations will not help with the present concern. Although discussions of consciousness often make metaphorical appeal to so-called inner sense, no such sense actually exists. We may

conclude that mental states are conscious by virtue of our having suitable thoughts about them.

It might be supposed that higher-order thoughts can help explain introspective or reflective consciousness, but not the so-called simple consciousness our mental states have in virtually every moment of our waking lives. Indeed, the connection between higher-order thoughts and introspection has sometimes been drawn. Writers such as D. M. Armstrong (1968, pp. 94-107, 323-38; 1980, esp. pp. 59-63), Daniel C. Dennett (1978b, esp. pp. 216-22), David Lewis (1966, p. 21; 1972, p. 258) and Wilfrid Sellars (1963, pp. 188-9, 194-5) have urged that being introspectively aware of a mental state means having a roughly simultaneous thought about that state. And Dennett (1976, esp. pp. 281-6), in a probing discussion of higher-order thoughts, uses that notion to explicate the concept of a person.<sup>1</sup>

If higher-order thoughts could explain only introspective consciousness, that would not do much to dispel the apparent difficulties in the notion of what it is for a mental state to be conscious. Introspective consciousness occurs when we pay deliberate, reflective attention to some mental state. That is relatively rare, and is a lot more elaborate than the non-reflective, phenomenologically immediate awareness we have of mental states in everyday life.

I have argued elsewhere (1986, 1990a), however, that we can in fact explain the ordinary, non-introspective consciousness of mental states in terms of higher-order thoughts.<sup>2</sup> On my account, a mental state is conscious – non-introspectively conscious – just in case one has a roughly contemporaneous thought to the effect that one is in that very mental state. Since not all mental states are conscious, it is open for not all of those higher-order thoughts to be conscious thoughts, though having such a thought will always mean that the lower-order thought it is about is conscious.

An account of introspective consciousness follows naturally. Introspection is the attentive, deliberately focused consciousness of one's mental states. So introspecting a mental state is not just being aware of it, but being actually conscious that one is thus aware. Since a state's being conscious is its being accompanied by a suitable higher-order thought, introspective consciousness occurs when a mental state is accompanied both by such a second-order thought, and also by a yet higher-order thought that one has that second-order thought. A mental state is conscious, but not introspectively conscious, when the higher-order thought it is accompanied by is itself not conscious. Introspective consciousness is the special case in which that second-order thought is also conscious. It is only if we assume that higher-order thoughts themselves must all be conscious that higher-order thoughts will seem useful in explaining introspective consciousness, but not ordinary non-introspective consciousness.

In previous work (1986, 1990a) I have mainly argued that an account of consciousness in terms of higher-order thoughts can save the phenomenological appearances and explain the data of introspection even more successfully than the traditional Cartesian view. Here I develop a wholly

different kind of argument – one that more directly and decisively supports an account in terms of higher-order thoughts. The next section sets out the background and premises of this argument, and the section after that puts forth the actual argument. The last three sections, then, defend the argument against various objections.

### *Expressing and Reporting*

Saying something and thinking it are intimately connected (see Sellars, 1964; Vendler, 1972, chs 1–3; Searle, 1983, ch. 1; Rosenthal, 1985). If one says something meaningfully and sincerely, one thereby expresses some thought that one has, and the thought and speech act will have the same propositional content. The speech act and thought will also in cases of sincere speech have the same force; both, that is, will be a matter of affirming, suspecting, wondering, denying, doubting, and the like. We usually speak of people expressing their thoughts; by an innocent metonymy, it is natural to talk also of a person's speech act as itself expressing, or giving expression to, the person's thought.<sup>3</sup>

But it is also possible to communicate what we think in another way, by saying something that does not literally express the thought we have. Instead of expressing our thoughts, we can describe them. If I think that the door is open, for example, I can convey this thought to you simply by saying 'The door is open'; that speech act will express my thought. But I could equally well convey the very same thought by saying, instead, 'I think the door is open'. Similarly, I can communicate my suspicion that the door is open either by expressing my suspicion or by explicitly telling you about it. I express the suspicion simply by saying, for example, that the door may well be open, whereas I would be explicitly telling you that I have that suspicion if I said that I suspect that it is open.

In every case, the speech act that expresses my thought has the same force and the same propositional content as the thought itself. But, if I say that I think the door is open, the propositional content of my speech act is not that the door is open; it is that I think it is open. And if I say I suspect the door is open, my speech act tells of a suspicion. But my speech act then has the force, not of a suspicion, but of an assertion. In saying I suspect something, I report, rather than express, my suspicion.

In general terms, then, I can convey my thought that *p* either just by saying that *p*, or by saying that I think that *p*.<sup>4</sup> These two distinct ways of conveying our thoughts to others are plainly distinct; still, it is easy to conflate them. This is because the conditions in which I could assert that *p* are the same as the conditions in which I could tell you that I think that *p*. Any conditions that warranted my saying that *p* would equally warrant my saying that I think that *p*, and conversely. Things are the same for other speech acts, and the mental states they express, even when their force is not that of asserting something. The conditions for appropriately expressing doubt, suspicion or wonder are the same as those for explicitly reporting that one is in those

mental states, at least when such social considerations as tact and discretion are not at issue.<sup>5</sup>

But the truth conditions for saying that *p* and saying that one thinks that *p* are, of course, dramatically different. That these truth conditions differ, even though the corresponding performance conditions are the same, is vividly captured by G. E. Moore's observation that the sentence 'I think that *p*, but I don't think that *p*', though not literally a contradiction, is still plainly absurd.<sup>6</sup> Such a sentence cannot have coherent conditions of assertibility, since the thought I seem to express by saying that *p* is precisely the thought I deny I have by going on to say that I don't think that *p*. Parallel remarks hold for speech acts other than assertions; I cannot, for example, coherently say 'Thank you, but I am not grateful.'<sup>7</sup> And, though I can perhaps actually have both thoughts simultaneously, I could not coherently convey both at once. Nor could I think them in, so to speak, the same mental breath. If the truth conditions for reporting thoughts and expressing them were not distinct, Moore's example would be not merely absurd, but an actual contradiction. We can infer to the distinction between expressing and reporting propositional states as the best explanation of why Moore's paradox is not an actual contradiction.<sup>8</sup>

Moore's paradox also helps with an earlier point, that all sincere speech acts express mental states with the same force and propositional content. If reporting and expressing were the same, Moore's paradox would be contradictory. If, on the other hand, I could say that *p* without thereby expressing the thought I have that *p*, Moore's paradox would not even be problematic, in whatever way. There would be no difficulty about saying that *p* and going on to deny that I have any such thought. We can thus infer to the claim that sincere speech acts express corresponding thoughts as the best explanation of Moore's-paradox sentences' being in some way absurd.

In ordinary conversation, however, we typically focus more on conditions of assertibility and other conditions for correct performance than on truth conditions. And as just noted, this may lead us to elide the difference between expressing our mental states and reporting them. This point will be crucial to dealing with certain objections in the last three sections. For now, an example will do. People untrained in logic generally regard literal contradictions as meaningless. It is wrong, however, to suppose, as many do, that this betrays some confusion of meaning with truth. It is simply that, until we are taught otherwise, we tend to rely on conditions of assertibility, rather than conditions of truth. Since a contradictory sentence lacks coherent conditions of assertibility, we can perform no meaningful speech act with it. The sentence itself is false, and so must have semantic meaning. But nobody could mean anything by asserting it.

The distinction between expressing and reporting is pivotal to the argument I want to advance for the theory that a mental state's being conscious consists in its being accompanied by a suitable higher-order thought. So it is important to see whether that distinction applies not only to our thoughts, doubts, suspicions and the like, but to all our conscious mental states.

It turns out that it does. One way to convey one's desires and emotions is to express them (see, for example, Hampshire, 1971). One does this both by the things one says and by one's facial expressions, gestures, choice of words and tones of voice. But one can also communicate these states by explicitly reporting that one is in the state in question. In the case of one's thoughts, the thought and its expression are about the same things, and the two have the same propositional content and the same truth conditions – they are true under the same circumstances. This holds also for emotions and desires; to the extent to which one's desire for food or fear of a tiger are about the food and the tiger, one's expressions of these states will be as well.<sup>9</sup> The same goes for whatever propositional content these states may have. But one's report of being in such a mental state is never about the very thing that the mental state itself is about. Rather, any such report must be about the mental state, and its propositional content is that one is now, oneself, in that very mental state.<sup>10</sup>

Sensations are a special case. Sensations have no propositional content, and are not therefore about things. Still, there is one kind of sensation that we plainly express non-verbally, namely, our bodily sensations such as pain. And these sensations may even be expressible in speech. We use various interjections, for example, to express pains, and perhaps this counts as speech. If so, such an expression of a bodily sensation would not diverge from the sensation expressed in respect of propositional content, since neither the sensation nor its expression has any propositional content. And reports of bodily sensations, such as 'It hurts', are about those sensations, and have propositional content in a way exactly parallel to reports of other mental states. So it is not surprising that the sentence 'Ouch, but nothing hurts' is like a standard Moore's-paradox sentence in being absurd, but not contradictory.<sup>11</sup>

Whereas bodily sensations are plainly expressible non-verbally and possibly verbally as well, neither sort of expressing is possible in the case of perceptual sensations. No speech act or other form of behaviour can express, for example, a sense impression of red. At best, a sense impression may occasion the comment that some observable object is red or, more rarely, that one has a red after-image or hallucination. But such remarks at best report red objects or red sense impressions; they will not express any sensation at all, but rather a thought about a red object or red sense impression. Perhaps it seems that one expresses a perceptual sensation when a startled cry is provoked by one's sensing a sharp or rapidly moving object. But this is hardly a clear case of one's expressing a perceptual sensation, as opposed to expressing, for example, one's feeling of fear. Or it may seem that saying 'Ah' as one savours a wine or settles into one's bath should count as a verbal expression of the relevant perceptual sensations.<sup>12</sup> But it is perhaps more reasonable to regard such borderline cases as expressing a bodily sensation of pleasure that accompanies the perceptual sensations in question.

The problem is that perceptual sensations seldom have any effect on our behaviour except when they are part of our perceiving something. But when one perceives something, one's behaviour, both verbal and non-verbal,

expresses the propositional content of the perception, and not its sensory quality. One can always isolate the sensory content for special attention. But even then, what one says and does will express one's thought about the sensory quality, and not the quality itself.

Sense impressions enter our mental lives, therefore, in a kind of truncated way, compared with other sorts of mental states. We express every other kind of mental state in fairly standard ways, sometimes even when we are not conscious of that state. Perhaps it is this odd feature of perceptual sensations that has made some follow Descartes in doubting whether sensations are mental states at all.

But our concern is with the distinction between reporting and expressing. In particular, we want to know whether expressions of mental states invariably have the same content as the states themselves, whereas reports of mental states always diverge in content from the states they are about. And the foregoing considerations show that sense impressions are not counter-examples to this generalization. Even though we speak of perceptual sensations as being 'of' various sorts of perceptible objects (see Sellars, 1963, pp. 154-5), such sensations are not actually about anything. And reports of sense impressions are, again, about those impressions, and their propositional content conforms to the pattern described earlier.

There is a view often associated with Wittgenstein that might be thought to cast doubt on the distinction between expressing and reporting one's mental states. In *Philosophical Investigations* Wittgenstein (1953) seems to have held, roughly, that although one can report that some other person is, for example, in pain, in one's own case one can only express the pain, and not report it as well.<sup>13</sup> If so, sentences like 'I am in pain', which ostensibly report bodily sensations, actually just express them.

But however suggestive this idea may be, it is plainly possible to report explicitly that we are in such states. And it is indisputable that others sometimes assert of us that we are, or are not, in particular mental states, and we sometimes explicitly contradict what they say. It is not just that we undermine what they say, as I might by saying 'ouch' when you say I am not in pain. Rather, we literally deny what others say about us. If we were unable to report on our own states of mind, but could only express them, this direct denial of the ascriptions others make about us would be impossible. If you deny that I am in pain and I simply say 'ouch', we have not thus far contradicted each other.<sup>14</sup>

We may thus conclude that, for creatures with the requisite linguistic capability, reporting mental states is possible, and that such a report differs in content from a verbal expression of that state. Not all types of mental state can be verbally expressed; perceptual sensations, for example, cannot be, and it is not clear what to say about bodily sensations. But creatures with the requisite linguistic ability can verbally express all other types of mental state, and in every case the verbal expression has the same propositional content as the state being expressed, and an illocutionary force corresponding to the state's mental attitude.

*The Argument*

Distinguishing clearly between expressing one's mental states and reporting them has important consequences for consciousness. Whenever one says something meaningfully and sincerely, one's speech act expresses some thought that one has. Speech acts that do not express one's thoughts either are parrotingly produced, as in something recited by rote, or else are cases of intent to deceive or dissimulate. So, whenever one meaningfully and sincerely reports being in some mental state, one's very report invariably expresses some thought that one has.

Moreover, speech acts that are meaningful and sincere express thoughts that have the same propositional content as the speech acts. So whenever one meaningfully and sincerely reports being in some particular mental state, one thereby expresses one's thought that one is, oneself, in that mental state. Unless one's words expressed that higher-order thought, the ostensible report would fail to be an actual speech act, rather than a piece of parroting behaviour. The ability to report being in particular mental states requires the ability to express higher-order thoughts that one is in those states. Clarity about the distinction between expressing and reporting points towards those very higher-order thoughts needed for the theory of consciousness I am defending.

There is an even more intimate tie, however, between the question of what it is for a mental state to be conscious and the distinction between expressing and reporting mental states. The ability to report being in a mental state of course presupposes moderately sophisticated capacities to communicate. But, given that a creature has suitable communicative ability, it will be able to report being in a particular mental state just in case that state is, intuitively, a conscious mental state. If the state is not a conscious state, it will be unavailable to one as the topic of a sincere report about the current contents of one's mind. And if the mental state is conscious, one will be aware of it and hence able to report that one is in it. The ability to report being in a particular mental state therefore corresponds to what we intuitively think of as that state's being in our stream of consciousness.<sup>15</sup>

But the ability to report being in a particular mental state is the same as the ability to express one's thought that one is in that mental state. So a mental state's being conscious will be the same as one's having the ability to express one's higher-order thought that one is in that mental state. It is unclear how one could have the ability to express some particular thought without actually having that thought. The best explanation of our ability to express the higher-order thought in question is plainly that one actually has that thought.

The converse holds as well. When a mental state is not conscious, we cannot report on it, and thus we cannot express higher-order thoughts about it. The best explanation of our inability to express higher-order thoughts about non-conscious mental states is that when the states are not conscious

no such higher-order thought exists. And, if conscious mental states are invariably accompanied by suitable higher-order thoughts, but non-conscious mental states never are, we have every reason to conclude that a mental state's being conscious consists simply in its being accompanied by such a higher-order thought.<sup>16</sup>

If a mental state is conscious, one can both express that state and report that one is in it. But when a mental state is not in one's stream of consciousness, even though one cannot then report being in that state, one can often still express it, at least non-verbally. One's non-verbal behaviour often betrays non-conscious mental states, by giving unwitting expression to them. A person may sometimes even explicitly deny being in a particular mental state whose presence is made overwhelmingly obvious by some non-verbal expression of it; we have all had occasion to remark, with the Queen in *Hamlet*, that somebody 'doth protest too much' (III, ii, 240).<sup>17</sup> This kind of occurrence shows that the abilities to express and report one's mental states need not coincide, any more than the mere ability to express a mental state non-verbally implies that that state is conscious.

The kind of consciousness we are focusing on is that special awareness we all have of our own mental states which is intuitively immediate, and seems to require no particular act of attention. Common sense puts few constraints on what positive account we should give of this intuitive immediacy. But part of our sense of immediacy plainly results from the awareness's being independent of both inference and observation, at least as these are ordinarily conceived.<sup>18</sup> And we must exclude both proprioceptive and visceral observation, as well as observation by way of the five exteroceptive senses.

Because our behaviour can express mental states we are unaware of being in, others can learn about such states by observation and inference. So occasionally others can point out to us that we are in some mental state we had not previously noticed – say, that we are irritated or pleased. But the feeling of pleasure or irritation is not a conscious mental state if one's awareness of it relies solely on ordinary observation and inference, as the other person's knowledge does. The feeling would become conscious only if we also came to know, non-observationally and non-inferentially, that the other person's comment is correct. For my mental state to be conscious, my higher-order thought about it must not be based on inference, at least not on any inference of which I am aware.<sup>19</sup>

Typically one's higher-order thoughts are not themselves conscious thoughts. Indeed, our feeling that the consciousness of mental states is somehow immediate is most vivid in just those cases in which the higher-order thought is not conscious. This is because conscious higher-order thoughts normally distract us from the mental states they are about,<sup>20</sup> so that those states no longer occupy centre stage in our stream of consciousness. But when we are unaware of having any higher-order thought, we lack any sense of how we came to know about the conscious mental state. It is this very feeling of mystery about how we come to be aware of conscious mental

states which encourages us to regard such consciousness as phenomenologically immediate.<sup>21</sup>

Is it necessary for a creature to have something as elaborate as human linguistic ability to be able to report its mental states? It is sometimes urged that we would not regard a creature's signals as making assertions unless those signals were embedded in something like human language. If so, non-human terrestrial animals would be unable to report their mental states.

But the ability to make assertions may well not require such elaborate resources. In particular, the syntactic complexity and semantic compositionality that permit the prodigious expressive powers characteristic of human language may not be necessary to performing simpler speech acts. It is far from obvious that a creature must be able to express the seemingly unlimited range of things that humans can to be capable of performing any speech acts at all. One factor that is more important is whether the creature can differentially express distinct mental attitudes. This ability seems more important to the core idea of what is involved in performing a speech act than the range and complexity of thoughts a creature can express. Another factor is whether there is some measure of conventionality about what various signals convey and a suitable degree of deliberate, voluntary choice about when the signal is used.<sup>22</sup>

It will take more, however, to report one's mental states than to be able to perform other sorts of speech acts. Unless a creature's signals exhibited a fairly sophisticated compositional structure, perhaps nothing would justify us in concluding that it was reporting its mental states, rather than just expressing them. Norman Malcolm's well-known distinction between thinking and having thoughts (1977, §II) seems to capture the distinction. Malcolm contends that non-linguistic animals can think, but cannot have thoughts. As he describes the difference, thinking seems to correspond to intentional states one can express but not report, whereas having thoughts corresponds to those one can report as well.<sup>23</sup> So what Malcolm calls thinking will be non-conscious thinking, and the having of thoughts will be conscious thinking.<sup>24</sup>

In any case, even if reporting one's mental states did require a communicative system with the full resources of human language, that would not show that creatures that cannot report their mental states have no conscious mental states. Being able to report a mental state means being able to express a higher-order thought about that state. Most creatures presumably have far more thoughts than they can express; the inability to express a thought hardly means that no such thought occurs. So if a creature were unable to express any of its higher-order thoughts, that would not imply that it had none.

The ability to report mental states is important here only because we understand what it is for a mental state to be conscious by appeal to creatures who can say what mental states they are in. We fix the extensions of terms for the various types of mental state by way of the conscious cases. We understand what it is for a mental state to be of this type or that by reference to conscious examples of that type of mental state, both our own and those of others. But we often fix the extensions of terms by way of a range of

phenomena narrower than those to which the terms apply. So using the conscious cases to fix the extensions of our terms for mental states does not show that all such states are conscious.

Similarly, we fix the reference of the term 'conscious' itself, as it applies to mental states, by the special case of creatures like ourselves that can report being in such states. But this does not show that creatures that cannot make such reports cannot be in conscious mental states. The connection between a state's being conscious and our being able to report that state reflects the fact that conscious states are accompanied by thoughts about those states and we can express those thoughts. In the human case, we may describe the ability to have higher-order thoughts in terms of the language system having access to certain mental states (see, for example, Dennett, 1978b, §§3-4). But what matters to a state's being conscious is the higher-order thought, not the resulting ability to report.<sup>25</sup>

### *A Dispositional Alternative*

On the foregoing argument, conscious mental states are those mental states we are able to report, and any such report must express a higher-order thought about the conscious state in question. It may seem, however, that these considerations do not support the conclusion that higher-order thoughts actually accompany all conscious states. Rather, they may support only the weaker conclusion that a higher-order thought must be able to accompany every conscious state. If so, the foregoing argument would show, instead, that a mental state's being conscious consists only in a disposition to have such a higher-order thought, and not in its actual occurrence.

This conclusion seems to receive support from an independent line of argument. Conscious mental states are mental states we can readily introspect, pretty much at will. So it is reasonable to think of a mental state's being conscious as a matter of our being able to become introspectively aware of it; conscious states are normally<sup>26</sup> introspectible states. Moreover, it is natural to think of being introspectible as a dispositional property. So it may seem but a short step to the conclusion that a mental state's being conscious is, itself, a dispositional property: a disposition to have higher-order thoughts about one's mental states. Similarly, conscious states are those we can report, and it is natural also to think of being reportable as a dispositional property.<sup>27</sup>

These considerations recall Kant's (1781, B131-8; cf. B157-9, A122-3, B406) well-known claim that the representation 'I think' must be able to accompany all other representations. Kant insists that the representation 'I think' be a non-empirical (B132) or transcendental representation (B401, A343); the possibility of its accompanying all other representations is a condition for those representations all being mine, united in one centre of consciousness (B132-5, esp. B134; on mental states' belonging to a subject, see note 21). But this qualification is irrelevant for present purposes, since the reflexive

representation Kant has in mind is presumably like other, more mundane thoughts in that a sincere, meaningful speech act could express it.

Kant does not say in so many words that a representation's being conscious is due to its being able to be accompanied by the representation 'I think'. But there is reason to think he holds this. The representation 'I think' must accompany all other representations because we could not otherwise explain what it is for my representations to be mine. And he seems to hold that a mental state's being mine coincides with its being conscious.<sup>28</sup>

Kant's dictum therefore suggests an account of a mental state's being conscious in terms of higher-order thoughts. But Kant does not say that the representation 'I think' actually accompanies all other representations, but only that it must be able to do so. His view is therefore a version of the dispositional account just sketched: A mental state's being conscious is not its being actually accompanied by a suitable higher-order thought, but its being able to be thus accompanied.

This dispositional view, however, does not readily square with our intuitive idea of what it is for a mental state to be conscious. A mental state's being conscious is our being conscious of being in that state in a suitably immediate way. And being conscious of things generally is occurrent, not dispositional. On the present theory, we are conscious of being in mental states when they are conscious states because we have higher-order thoughts about those states. Merely being disposed to have such thoughts would not make us conscious of the states in question; we must have actual, occurrent higher-order thoughts. Having a disposition to have a thought about a chair could not make one conscious of the chair; how could having a disposition to have a higher-order thought about a mental state make one conscious of that state? This conclusion accords well with our common-sense intuitions, on which, whatever being conscious may amount to, it seems to be a clear case of a non-dispositional, occurrent property of mental states.

Moreover, the fact that conscious states are all introspectible and reportable does not show that a state's being conscious is solely a dispositional matter. Being conscious can perfectly well be a non-dispositional, occurrent property of mental states and yet involve dispositions. One and the same property can often be described in both dispositional and occurrent terms. Something's being red plainly involves various dispositions, such as causing bulls to charge. And perhaps something's being flammable or soluble consists in something's having a certain physical make-up, though we pick out those properties by way of a disposition to burn or dissolve. Similarly, a mental state's being conscious is an occurrent property, even though it involves such dispositions as being introspectible and reportable.<sup>29</sup>

Positing occurrent higher-order thoughts as accompanying all conscious states also readily explains why those states are introspectible and reportable. It is my ability to express my thoughts verbally that enables me to report on the mental states my higher-order thoughts are about. And it is because my higher-order thoughts can become conscious that I can come to introspect

the mental states those thoughts are about. This second point will figure below toward the end of this section.

On the present theory, a mental state's being conscious is a relational property – the property of being accompanied by a higher-order thought. This accords poorly with common sense, which seems to represent being conscious as non-relational. Why should it matter, then, that common sense represents a mental state's being conscious as occurrent, and not dispositional?

Common-sense considerations are hardly decisive, and may well be overruled by theory. But we should try to do some sort of justice to those intuitions. There are different ways to do this. The present theory preserves the intuition that consciousness is occurrent. And, though it does not preserve the idea that consciousness is non-relational, the theory does explain why that idea is so appealing. Consciousness seems to be non-relational because we are generally unaware of the higher-order thought that makes a mental state conscious, and thus unaware of the relation by virtue of which that consciousness is conferred.

An apparent advantage of the dispositional view stems from the difficulty in accepting the existence of so many higher-order thoughts. At most waking moments we are in a multitude of conscious states; it seems extravagant to posit a distinct higher-order thought for each of those conscious states. When a mental state is conscious, we plainly have the ability to think about it, but it seems equally plain that we do not actually think about all our conscious states. A dispositional account circumvents this difficulty by requiring only that we be disposed to have a higher-order thought about each conscious state, and not that we actually have all those thoughts.

But this line of reasoning rests on a mistake. Thinking about something is not just having a train of thoughts about it, but having a conscious train of thoughts about it. We seldom think, in that way, about any of our conscious states. But the higher-order thoughts the theory posits are typically not conscious thoughts. The intuitive difficulty about how many higher-order thoughts we could have arises only on the assumption that those thoughts must be conscious; higher-order thoughts of which we are unaware will pose no problem. The worry about positing too many higher-order thoughts comes from thinking that these thoughts would fill up our conscious capacity, and then some; we would have no room in consciousness for anything else. But this is a real worry only on the assumption that all thoughts are automatically conscious thoughts.<sup>30</sup>

A mental state's being conscious consists in one's being conscious of being in that state. Being conscious of being in a mental-state type will not do; I must be conscious of being in the relevant token of that type. A dispositional account faces a difficulty here, since it is far from clear how a disposition to have a higher-order thought can refer to one mental-state token rather than another. Perhaps a dispositional account will require not that the disposition refers to a mental-state token, but that it is a disposition to have a higher-order thought that refers to it.<sup>31</sup> Still, if such a disposition is responsible for

a mental state's being conscious, the disposition must somehow connect with the right mental state, and it is unclear how that can take place.

This problem becomes especially intractable in the case of sensations. No higher-order thoughts could capture all the subtle variations of sensory quality we consciously experience. So higher-order thoughts must refer to sensory states demonstratively, perhaps as occupying this or that position in the relevant sensory field. It is especially unclear how mere dispositions to have higher-order thoughts could accomplish this.

A headache or other bodily sensation may last an entire day, even though one is only intermittently conscious of it. The point is not merely that one introspects or pays attention to headaches only intermittently; a day-long headache is unlikely to be constantly in one's stream of consciousness in any way at all. And, in general, sensory states need not be conscious states. The distinctive sensory qualities of such states are simply those properties in virtue of which we distinguish among sensations as having distinct sensory content. There is no reason to hold that these differences can obtain only when the sensation is conscious. The distinctive sensory properties of non-conscious sensations resemble and differ in just the ways that those of conscious sensations resemble and differ, differing only in that the one group is conscious, whereas the other is not.<sup>32</sup>

Explaining the intermittent consciousness of such a headache is easy if we appeal to occurrent higher-order thoughts; occurrent thoughts come and go. So one and the same mental state, such as a headache, could persist, sometimes accompanied by a higher-order thought, sometimes not. Mental states would accordingly enter and leave our stream of consciousness. Dispositions seem less well-suited to this task. Because they are more long lasting, dispositions seem intuitively less likely to come and go with the desired frequency, as occurrent higher-order thoughts might.

The argument of the previous section also tells against a dispositional account. A mental state's being conscious is manifested by reports that one is in that state, and to be meaningful these reports must express corresponding thoughts about those mental states. And speech acts plainly do not express mere dispositions to have thoughts.<sup>33</sup>

Conscious states can normally be introspected. And one might argue that even though this does not imply a dispositional account, a dispositional account is necessary to explain why it is so. But that is a mistake. A state is introspectible if it can become an object of introspection. And introspecting a state consists in being aware of that state, and also being conscious that one is thus aware. So introspecting is having a conscious thought about a mental state. On the present theory, a state's being conscious is its being accompanied by a suitable higher-order thought. Those higher-order thoughts are typically not conscious; but once one has such a higher-order thought, it can itself become conscious. And its being conscious results in the mental state it is about being introspectively conscious. Conscious states are introspectible because higher-order thoughts can themselves become conscious thoughts.

If all mental states were conscious states, however, this explanation would

be unavailable. There would then be no difference between having a conscious thought about one's mental state and having a thought about it, *tout court*. So one could not explain why a state's being conscious coincides with its being introspectible by saying that the accompanying higher-order thought is not conscious but can become so. Moreover, introspecting would simply be having such a thought, and a mental state would be introspectible just in case one were disposed to have such a thought about it. A state's being non-introspectively conscious could not then consist in having a higher-order thought, on pain of collapsing the distinction between being introspectively and non-introspectively conscious. We would thus have to say that a state's being non-introspectively conscious consists in one's being disposed to have such a thought.

The idea that a state's being a conscious state consists in a disposition to have a suitable higher-order thought does not explicitly presuppose that mental states are always conscious. But the foregoing considerations suggest that this Cartesian picture may underlie much of the appeal a dispositional theory has. Only if we tacitly assume all mental states are conscious will the dispositional account be needed to explain why conscious states are introspectible. But the assumption that all mental states are conscious is plainly question-begging in the context of evaluating the present theory.<sup>34</sup>

### *Are Higher-order Thoughts Possible?*

The objection just considered sought to show that conscious mental states need not be accompanied by occurrent higher-order thoughts. But there are other considerations that seem actually to cast doubt on whether such higher-order thoughts are possible at all. When a conscious mental state is a thought, the mental analogue of performance conditions will be the same for that thought as for the higher-order thought about it. No circumstances exist in which I can appropriately think that *p*, but not appropriately think that I think that *p*. Perhaps, moreover, the right way to individuate mental states is by reference to these mental analogues of performance conditions. If so, the ostensibly higher-order thought would be indistinguishable from the thought it purports to be about; the conditions for having a thought about a thought would be the same as those for just having that thought. The very idea of distinct higher-order thoughts about other thoughts would accordingly be incoherent.

Brentano (1973, p. 127) actually advances just such an argument, applying it even to the case of perceiving.<sup>35</sup> Thus he maintains that my hearing a sound and my thought that I hear it are one and the same mental act. And he goes on to insist that the very content of that perception must be contained in the content of any higher-order thought about it, thus reasoning from performance conditions to mental content. Accordingly, he concludes, every mental state is, in part, about itself; in his words, all mental acts 'apprehend [themselves], albeit indirectly' (1973, p. 128). Every mental state,

in addition to having its standard nature, will also function as a higher-order thought about itself.

This idea is not uncommon. Locke seems to express it when he writes that 'thinking consists in being conscious that one thinks' (1700, II, i, 19; cf. Locke, 1700, I, ii, 5; II, i, 10-12; II, x, 2; and IV, vii, 4; on Locke's reflexive model of consciousness see also: 1700, II, xx, 1 and II, xxvii, 9). Some such idea seems also to underlie Descartes' and Hobbes's insistence that it is absurd to suppose that one thought could ever be about another.<sup>36</sup> If higher-order thoughts must be a part or aspect of the mental states they are about, an account of consciousness in terms of distinct higher-order thoughts will be unintelligible.<sup>37</sup>

It is useful to see Brentano's argument as the mental analogue of Wittgenstein's (1953) idea that meaning is use.<sup>38</sup> Understanding how sentences are correctly used is knowing their performance conditions. So focusing on use will suggest typing speech acts by reference to performance conditions, rather than by such semantic features as truth conditions or propositional content. This fits well with Wittgenstein's suggestion, noted at the end of the earlier section entitled 'Expressing and Reporting', that one cannot report, but can only express, one's own pains and other bodily sensations. 'I am in pain' is indistinguishable from 'ouch' in respect of performance conditions; so if one focuses solely on performance conditions, it is natural to type the two together.

This has consequences for how we think about consciousness. 'It hurts' and 'I am in pain' plainly have propositional content; they are about one's pain, and thus express one's thought that one is in pain; that is, they express one's awareness of being in pain. And saying 'ouch' plainly expresses one's pain. So if 'I am in pain' were on a par with 'ouch', it too would express one's pain, as well as expressing one's awareness of being in pain. Accordingly, sentences such as 'I am in pain' and 'It hurts' intuitively seem to yoke together the pain and one's awareness of it, suggesting that the two actually cannot occur separately.<sup>39</sup>

But there is more to a mental state than the mental analogue of its performance conditions. When we individuate mental states, we must also take into account their truth conditions, if any, and their propositional content. Two states are the same only if their semantic properties and performance conditions are the same. Brentano's argument hinges on the idea that it is sufficient for two states to be the same that they have the same mental analogue of performance conditions. That argument cannot therefore be sustained, and there is thus no incoherence in the idea of higher-order thoughts distinct from the thoughts they are about. Indeed, mental attitude is by itself sufficient to undermine Brentano's argument. Suppose the higher-order thought is about a suspicion or doubt; that state will perforce have a mental attitude distinct from any higher-order thought, since higher-order thoughts will invariably have the mental attitude corresponding to an assertion.

There is in any case a somewhat idle air to Brentano's claim that higher-order thoughts are part of the mental states they refer to. How could we ever

show, in a non-question-begging way, that a higher-order thought is part of the mental state it is about, rather than that the two are just distinct, concurrent states? It would be more tempting to hold this if all mental states were conscious. If we trace a state's being conscious to the presence of a higher-order thought and every mental state is conscious, there will be a higher-order thought for every mental state. Since no mental state would then occur without its higher-order thought, it might seem inviting to hold that higher-order thought to be part of the state itself. But, if higher-order thoughts are distinct mental states, we can explain why we are generally unaware of them only by saying that such thoughts are usually not conscious thoughts. And this explanation would be unavailable if all mental states were conscious. It begs the question against the present theory to suppose all mental states are conscious, and in any case we have excellent reason to hold that mental states exist that are not conscious states.

There is an even more dramatic way to see how the view suggested by Brentano and Locke goes wrong. If every mental state is conscious and every conscious mental state is, in part, about itself, every mental state without exception will, in part, be about itself. Those who endorse this reflexive model of consciousness presumably find this consequence acceptable. But there is a further implication that has generally not been noted. To say anything meaningfully and sincerely, one's speech act must express some thought that has the same force and the same propositional content. So, if every mental state is, in part, about itself, it will be impossible to say anything at all that is not, in part, literally about one's own mental states.

Locke actually seems to endorse something like this when he claims that the primary use of words is to refer to the ideas in one's mind.<sup>40</sup> Indeed, in advancing this view, Locke seems deliberately to assimilate expressing one's ideas to reporting them. As he puts it, 'the Words [a person] speak[s] (with any meaning) . . . stand for the Ideas he has, and which he would express by them' (1700, III, ii, 3).<sup>41</sup> Words, on this account, are about the very ideas they express.

This assimilation of reporting and expressing recalls Locke's reflexive model of consciousness, though his doctrine about words and ideas derives not from that model but from his views about meaning. Locke's semantic theory thus fits well with his views about consciousness. Because words apply primarily to the speaker's ideas, all speech acts will be about the speaker's mental states. It is interesting to note in this connection that J. R. Ross (1970) has argued, on grammatical grounds, that the deep structure of every declarative sentence is dominated by the pronoun 'I' plus some verb of linguistic performance, as though every such sentence implicitly reported its own illocutionary force and meaning.

But just because particular grammatical or semantic theories fit neatly with the reflexive model of consciousness does not mean that those theories provide any support for that model. The conclusion that the reflexive picture of consciousness forces on us is truly extravagant: Every speech act, to be meaningful and sincere, must literally refer to one of the speaker's own

mental states. It is hard to see how any grammatical or semantic theory could render this claim acceptable.

### *Is Reporting Distinct from Expressing?*

Stressing performance conditions over propositional content raises doubts about whether a higher-order thought can really be distinct from the thought it is about. But there is another source for such doubts. We sometimes use verbs of mental attitude in ways that may appear to undercut the sharp distinction between expressing and reporting, on which the argument three sections back (in the section entitled 'The Argument') relied. And if reporting and expressing a mental state are the same, a speech act such as saying 'I think it's raining', will indifferently express both the ostensible higher-order thought that I think it's raining and the thought that it's raining. Those thoughts will then arguably be the same, since an unambiguous speech act presumably expresses a single thought.

The problem is this. I can express my doubt about something by saying, for example, that it may not be so. But, even when I say 'I doubt it' or, more explicitly, 'I doubt that  $p$ ', it seems natural to take me to be expressing my doubt, and not just reporting what mental state I am in. Similarly, if I say that I suppose or choose something, or sympathize with somebody, it is natural again to see this as actually expressing my supposition, choice or sympathy, and not just telling you about the contents of my mind (cf., for example, Alston, 1965, esp. p. 16). A parody will illustrate the point especially vividly. If you ask me whether it is raining and I say 'I think so', it would be bizarre to take me to be talking about my mental state, rather than the weather (see note 13).

This challenge is important. Higher-order thoughts entered our account of consciousness because conscious mental states are those we can report non-inferentially. And those putative reports will not be actual speech acts unless they express thoughts about the conscious mental states in question. But if such ostensibly higher-order remarks are not really second-order reports about our mental states at all, but only express those states, those remarks will not express any higher-order thoughts.

This conclusion would thus vindicate the Cartesian claim that consciousness is intrinsic to mental states. The second-order character of such remarks would be a surface illusion, and would not imply the existence of any distinct second-order thoughts. More important, these ostensibly second-order remarks would presumably also report the very mental states they express. The speech act 'I doubt it' would both express and report one's doubt, so that the doubt itself would have both the content of the doubt and the content that one had that doubt. Every conscious mental state would be, in part, about itself; consciousness would be a reflexive feature of mental states, in Ryle's apt words, a matter of their being 'self-intimating' or 'self-luminous' (1949, p. 159). We would then have no choice but to swallow the strikingly

unintuitive consequence noted earlier about sincere speech acts' invariably referring to the speaker's own mental states.

It is worth noting that not all ostensible reports follow the pattern illustrated above. If I say that I gather, deduce, covet or recognize something, it is not all that tempting to hold that I thereby express, rather than report, my mental state. And if I say that I expect, want, understand or suspect something, it is plain that I am then explicitly talking about my mental states, and not merely, as we say, 'speaking my mind' – not just expressing those states.

But even in cases such as saying that I think, doubt, suppose or choose something, the tendency to take my remarks to express, rather than report, my mental states misleads. As already noted, that temptation stems from focusing on the performance conditions of such sentences at the expense of their distinctively semantic characteristics – their truth conditions and propositional content. The sentences 'I doubt it' and 'I think so' may superficially seem to express one's doubts and thoughts, and to be about whatever those doubts and thoughts are about. But this is because the circumstances in which one can appropriately say that something may not be so are the same as those for saying that one doubts it; similarly for saying that something is so and saying that one thinks it is.<sup>42</sup>

Once again, meanings and truths conditions tell a different story. If I am asked whether it is raining and I say I think so, my remark is not semantically equivalent to saying 'Yes, it's raining'.<sup>43</sup> The sentence 'It's raining but I don't think so' is absurd, but not contradictory; the sentence 'I think it's raining but I don't think so', by contrast, is an actual contradiction.<sup>44</sup> Moreover, if you say that I believe, doubt, suppose or choose something, I can deny what you say, and in so doing I would contradict what I would have said if instead I agreed with you. And agreeing with you would naturally take the form of saying that I believe, doubt, suppose or choose that thing.

The distinction between reporting and verbally expressing one's mental states is crucial to the argument of this chapter. Given the failure of the most plausible attempts to undermine that distinction, we may conclude that the argument successfully supports a theory of consciousness in terms of higher-order thoughts.

### *Acknowledgements*

Much of the work on this chapter, resulting in the penultimate draft, was done while I was a fellow at the Center for Interdisciplinary Research (ZiF), University of Bielefeld, Germany, in 1989–90. I am indebted to the Center for generous support and exceptionally congenial and stimulating surroundings. Previous drafts were read at Tufts, Columbia, The University of Wisconsin at Milwaukee, and the August 1989 Joint Conference of the Sociedad Filosófica Ibero-Americana and the Sociedad Argentina de Análisis Filosófico, in Buenos Aires. I am grateful to friends and colleagues in those audiences

for helpful reactions, and to Martin Davies for useful comments on the penultimate draft. Special thanks to Daniel Dennett.

### Notes

- 1 D. H. Mellor appeals to higher-order believing to explain not merely our introspective consciousness of beliefs, but the ordinary, non-introspective consciousness beliefs often have (Mellor, 1977–8, 1980). But he holds that this view, for which he argues forcefully, applies only to the case of conscious believing, and not to other mental states as well.
- 2 Dennett (1978b) develops a related strategy for avoiding the difficulty about consciousness being an intrinsic feature of mental representations.
- 3 There are differences. If the expressing is deliberate, it may be more natural to say that the person expresses the mental state; otherwise one may say instead that the person's behaviour expresses that state. On these two ways of speaking, see Alston (1965, esp. pp. 17–8, 23–6). (I use 'thought' throughout as a generic term covering all types of propositional mental states, and 'say' as a generic verb to cover all kinds of speech act, whatever the illocutionary force.)
- 4 Cf. Dennett's related distinction between reporting and expressing in (1969, esp. §13), though Dennett's concern there is with introspective, rather than simple, consciousness.
- 5 Strictly speaking, it may be that the performance conditions for saying that one thinks that *p* are not identical with, but rather include, the performance conditions for saying that *p*. This refinement is irrelevant in what follows.
- 6 Moore (1942, p. 543; 1944, p. 204). Moore uses believing, rather than thinking, as his example, but the point is the same. On Moore's paradox (so called by Wittgenstein, 1953, II, x), see Black (1954), Burnyeat (1967–8), and Wittgenstein (1953, II, x).
- 7 Pace Mellor, who claims that Moore's paradox 'has no analogue for the other attitudes' (1984, p. 38). Mellor also restricts his account of consciousness to the case of believing (see note 1 above). As will emerge in the next section, the generalization of Moore's paradox shows that this restriction is unwarranted.
- 8 It may seem possible to explain this absurdity without appeal to any such distinction. On a Gricean view, my sincerely saying something involves my intending that my hearer believe that I believe what I say. Such a view thus implies that I cannot at once sincerely say that *p* and say that I do not believe it (see Mellor, 1977–8, pp. 96–7; cf. 1980, p. 148). But Moore's paradox strikes us as absurd independent of any context of communication; it is absurd because it lacks coherent conditions of assertibility. Insincere speech on the Gricean picture involves my intending you to believe that I believe something that in fact I do not (see Mellor, 1980, p. 148). But it is more reasonable to regard insincere speech

acts as in fact a degenerate kind of speech, similar to reciting lines in a play. Like play acting, insincere speech is a kind of pretence; in both cases, one in effect pretends to perform normal speech acts. Thus Austin (1946, pp. 69–71 in 1961 reprinting, pp. 101–3 in 1970 edition) describes a sense of 'promise' and related words in which if I speak insincerely I do not promise. Compare Austin (1962, pp. 48–50, 135–6; pp. 48–50, 136–7 in 1975 edition) in which he urges that an insincere speech act, though it succeeds, is defective (1962, pp. 15–6 and Lecture IV, esp. pp. 39–45 in 1975 edition). Also see Rosenthal (1985), esp. §§2, 3 and 5, and the 1989 postscript. There is also nothing automatically problematic about one's speaking insincerely in ways that betray that insincerity. So one cannot express the absurdity, as Moore once proposed (1942, pp. 542–3), as due to one's betraying one's insincerity if one says something with the form of Moore's paradox.

- 9 This is plainly true of verbal expressions, and it is plausible for non-verbal expressions as well.
- 10 It is not sufficient that the report be about somebody who happens to be oneself. Rather, the report must be about oneself, as such; that is, it must be a report that the being that is in the mental state is oneself. See Castaneda (1968), Anscombe (1975), Lewis (1979), Perry (1979), Boër and Lycan (1980) and Chisholm (1981) for discussions of the special sort of reference involved.
- 11 Care is necessary here. Sentences such as 'It hurts' may at first glance seem to express pains, rather than report them. But since 'Ouch' if anything expresses one's pain rather than reporting it, it is reasonable to explain why 'Ouch, but nothing hurts' is not contradictory by taking 'It hurts' as reporting one's pain, and not expressing it.
- 12 This was suggested by Daniel Dennett.
- 13 Wittgenstein (1953, I, §§244, 256, 310, 377); cf. Malcolm (1963a, pp. 138–40; 1963b, pp. 105–17). In an illuminating discussion, Rosenberg (1977) argues that it is characteristic of speech acts such as 'I am in pain' that we have no criteria for their being true independent of our criteria for their being performed truthfully, i.e. sincerely. He concludes that such avowals are 'report[s] judged as ... expression[s]' (1977, p. 159). This suggestion goes far in capturing Wittgenstein's idea, while still recognizing the reporting status of the relevant speech acts. Intuitively, it seems out of place to evaluate expressions of feelings with respect to their cognitive credentials and success, as one might evaluate reports. Accordingly, Wittgenstein's view seems to capture whatever sense we have that speech acts such as 'I am in pain' have some special epistemic privilege. Wittgenstein actually seems to extend the expressive theory beyond the case of sensations. See Wittgenstein (1953, II, §x, p. 190): '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".' Sentences that ostensibly report one's own beliefs, like those which ostensibly report one's own sensations, would then really just express those beliefs and sensations.

- 14 Parallel remarks hold for Wittgenstein's claim about 'I believe it's going to rain'. See also the final section of the present chapter. It may seem that the expressive theory works better with 'I am in pain' and its kindred than with 'I am not in pain'. Perhaps one can contradict such negative remarks, but not the affirmative counterparts. Moreover, 'I am not in pain' must express my thought that I am not in pain, since if I am not in pain, there is nothing else for it to express. But in both cases, whatever inappropriateness exists in my contesting your word about your own mental states disappears if we imagine that I speak first. There is nothing intuitively amiss if I say you are not in pain and you insist you are, or I say you are and you deny that. Moreover, you would then be contradicting what you would say if, instead, you agreed with me that you are not in pain. Even if the expressive theory applies only to affirmative sentences, it cannot accommodate these facts. Although 'I am in pain' and 'I think that  $p$ ' express thoughts about my pain and my thought that  $p$ , they may still have the same force as 'ouch' and ' $p$ ', respectively. My saying that I am in pain or that I think that  $p$  will then be appropriate when, and only when, I am in pain or think that  $p$ . These considerations would explain our sense that these speech acts are somehow privileged, and even capture the kernel of truth in Wittgenstein's stronger claim that we never report, but only express, our bodily sensations.
- 15 Cf. Dennett (1969, ch. 6; 1978b, esp. §§3-4). Robert Van Gulick has urged that we detach self-consciousness from the ability to report (1988a, p. 160). But Van Gulick identifies self-consciousness in terms of the subpersonal possession of 'reflexive meta-psychological information' (1988a, p. 160ff.); self-consciousness occurs whenever a mental state has informational content that involves some other mental state. This notion covers far more than the intuitive notion of a conscious mental state, which is under present consideration. So even if Van Gulick's defined notion is independent of any abilities to report, nothing follows about our intuitive notion of a conscious state.
- 16 In an earlier publication (1986, p. 339) I assumed that a mental state's being conscious required not only that it be accompanied by such a higher-order thought, but also that it cause that higher-order thought. This causal requirement may seem natural enough; after all, what else would cause that higher-order thought? But the requirement is unmotivated. Being a conscious state requires only that one be conscious of being in that state; the causal antecedents of being thus conscious do not matter to whether the state one is conscious of being in is, intuitively, a conscious state. The causal requirement may seem tempting as a simulation of the essential connection Cartesians see between mental states and their being conscious. But this weaker connection is still problematic. If mental states cause accompanying higher-order thoughts, why do many mental states occur without them? We might posit causal factors that block the causal connection, but that wrongly

makes being conscious the normal condition for mental states. It is more natural to suppose that higher-order thoughts are caused by a coincidence of mental factors, many of which are causally independent of the state in question.

- 17 Sometimes even one's speech acts will give unwitting expression to mental states one is not conscious of; Freudian slips are the obvious example. But here things are more complicated. When a speech act unwittingly expresses a non-conscious mental state, one is aware of the content of one's speech act. So the content of the speech act is distinct from the content of the non-conscious state it betrays. Indeed, it is probable that, with systematic exceptions, whenever one expresses a thought verbally, that thought will be conscious. (On the explanation of this generalization, and on why it does not threaten the argument of this section, see Rosenthal, 1990b, 1991b.) It is therefore natural to understand these cases on the model of non-verbal expressing. One's speech act reveals one's non-conscious state not by functioning linguistically, but by being a piece of non-verbal behaviour that gives non-verbal expression to that state.
- 18 Perhaps, as Gilbert Harman (1973) convincingly urges, much of our knowledge derives from non-conscious inference. If so, such non-conscious inference may well underlie the presence of the higher-order thoughts that make mental states conscious. Such non-conscious inferences are not precluded here, since they would not interfere with the intuitive immediacy of such consciousness.
- 19 Dennett has remarked (personal communication) that there will be penumbral cases in which one simply cannot tell whether or not one's higher-order thought is based on inference, so understood. This is no problem; in such cases one will plausibly also be unsure whether or not to count one's mental state as a conscious state.
- 20 As Ryle (1949, p. 165; cf. p. 197) in effect observed, though he talks simply of higher-order mental activities, and omits the qualification that they be conscious. It is important for a theory of consciousness in terms of higher-order thoughts that non-conscious higher-order thoughts do not distract one from the mental states they are about.
- 21 There is another way in which the consciousness of conscious mental states seems intuitively immediate. In the case of non-conscious mental states, it is arguable that their belonging to a particular subject is no more problematic than the bond between physical objects and their properties. But when a mental state is conscious, it is tempting to insist that there is more to say about how it belongs to a particular subject. The present theory explains what is phenomenologically special about the way in which conscious mental states belong to their subjects. Propositional states, both conscious and non-conscious, intuitively seem bound to their subjects by their mental attitudes; the attitude is a kind of relation joining a thinking subject to its thoughts. And sensory mental states, whether conscious or not, seem similarly tied to their subjects by occurring within a field of experience, which connects these mental states

to others of the same and different sensory modalities. But conscious states seem tied to their subjects in some way above and beyond the bond they have in common with non-conscious states. We can explain this additional tie that such consciousness seems to add as due to the content of the accompanying higher-order thought. That higher-order thought is a thought to the effect that one is, oneself, in the mental state in question. Because such higher-order thoughts are both about oneself and the mental state one is in, they carry with them the sense that the tie between one's mental state and oneself is stronger when the state is conscious than when it is not.

- 22 When the means of expressing thoughts is not all that systematic, as with all the non-human terrestrial animals we know about, we will want to see more conventionality and deliberateness to be convinced that speech acts are occurring.
- 23 Malcolm would not put it this way, since his (1977) inclination towards an expressive theory of first-person ascriptions of mental states leads him, in effect, to assimilate reporting and expressing. And this, together with a view of expressive ability modelled on human language, lead him in turn to a rather restrictive view of the mentality of non-human animals (1977, §iii).
- 24 As noted above (note 17), verbally expressed thoughts are typically conscious. But when a creature lacks the ability to report its mental states, verbally expressing them may well imply nothing about whether the state expressed is conscious. This point is exploited in Rosenthal (1990c).
- 25 Indeed, Dennett in (1978b, e.g. p. 217) also puts his point in terms of the having of thoughts – what he calls 'thinkings'. These considerations show that, *pace* Van Gulick (1988a, p. 162), the fact that many mental states of non-linguistic creatures are conscious provides no reason to deny the connection between a state's being conscious and the ability for creatures with suitable linguistic endowment to report on that state.
- 26 I.e. normally for creatures like us that have the capacity for being introspectively conscious of their mental states.
- 27 I am especially indebted to Daniel Dennett for pressing on me the virtues of some form of a dispositional view, and also for many other helpful reactions to drafts of this paper.
- 28 Kant explicitly allows that I need not be conscious of my representations as being mine (B132). So, if its being mine is its being non-introspectively conscious, my being aware of it as mine would be my introspecting it.
- 29 There are additional sources of confusion. Something can plausibly be dispositional from the point of view of common sense and occurrent from the point of view of a scientific treatment, or conversely. Similarly, something can count as an occurrent property from the vantage point of science and common sense but figure dispositionally within a functional or computational description, or again conversely. Moreover, it is sometimes difficult to draw any useful distinction between short-term dispositions and occurrent properties. In claiming that a mental state's

being conscious is its being actually accompanied by a suitable higher-order thought, the present theory is operating with our ordinary folk-psychological categories and concepts. Still, such higher-order thoughts might correspond to something dispositional when we move to scientific theory.

- 30 One might hold that the distinction between occurrent thoughts and dispositions to have them coincides with the distinction between conscious and non-conscious thoughts. If so, occurrent thoughts could never be non-conscious. The only way, then, to put the present theory would be in terms of dispositions to have higher-order thoughts, since these thoughts are typically non-conscious. (I am grateful to Ernest Sosa for pressing this point.) But it is not easy to see how one might independently substantiate this Cartesian denial of non-conscious occurrent mental states.
- 31 I am grateful to Martin Davies for this point.
- 32 For more on sensory quality see Rosenthal (1986, §iii; 1990a, §ii; and 1990b). Norton Nelkin, if I understand him, holds that sensory states, though they can occur non-consciously, always have what he calls phenomenologicality, and thus are invariably felt (Nelkin, 1989b, p. 139). It is unclear in what sense non-conscious states might be felt. Apart from these issues, Nelkin presents a view very similar to that defended here, though he advances different arguments for it.
- 33 The argument that speech acts express corresponding thoughts relied on Moore's paradox; 'It's raining but I don't think so' lacks coherent performance conditions because one cannot meaningfully say it's raining and not think it is. So one could argue that meaningful speech acts must express dispositions to have the relevant thought, since 'It's raining but I'm not disposed to think so' also lacks coherent performance conditions. But this shows at best that meaningful speech acts express both corresponding thoughts and dispositions to have such thoughts, since whenever the dispositional version of Moore's paradox works, the non-dispositional version will as well. Indeed, the dispositional version works presumably only because the corresponding non-dispositional version does. Since 'I'm not disposed to think so' is stronger than 'I don't think so', if 'It's raining but I don't think so' is problematic 'It's raining but I'm not disposed to think so' must be as well. So the reason meaningful speech acts are accompanied by dispositions to have higher-order thoughts is that they are accompanied by the actual thoughts themselves.
- 34 Introspecting a mental state means having a conscious higher-order thought, and non-conscious mental states are presumably unaccompanied by higher-order thoughts, whether conscious or not. Moreover, a state's being non-introspectively conscious involves less than its being introspectively conscious, but more than its not being conscious at all. So it may seem that only a disposition to have a higher-order thought could fit in between a conscious higher-order thought and its absence. But the Cartesian assumption that all mental states are conscious again figures

- in this reasoning. Such a disposition is the natural intermediate between a conscious higher-order thought and none at all only if we tacitly rule out the possibility of a higher-order thought that is not conscious.
- 35 As Brentano puts it, we must choose whether to individuate propositional mental states (presentations) in terms of their (propositional) object or the mental act of the presentation. Brentano credits Aristotle with the idea. Aristotle's actual argument (1907, III 2, 425b13-4), which Brentano adapts, is that if the sense by which we see that we see is not sight, then the sense of sight and the other sense would both have colour as their proper object, and distinct senses cannot share the same proper object.
- 36 For both, see *Third Replies*, Descartes (1964-76, vol. VII, p. 175; 1984-91, vol. II, p. 124); though see also *Seventh Replies* (1964-76, vol. VII, p. 559; 1984-91, vol. II, p. 382).
- 37 This conclusion echoes a certain interpretation of the thesis that knowing implies knowing one knows. On that interpretation, there is nothing to such second-order knowing above and beyond first-order knowing itself. Here, too, stressing performance conditions over propositional content seems to be at issue: the force of saying or thinking that I know is equivalent to that of saying or thinking that I know that I know, even if their propositional contents differ. Historically, however, this view has generally encouraged claims about the transparency of mind to itself, rather than vice versa. And if knowing can be tacit, the idea that knowing implies knowing one knows is independent of such claims of transparency.
- 38 Applied, however, to sentence-sized, rather than word-sized, mental units: 'The meaning of a word is its use in the language' (1953, I, §43). See also note 13. Wittgenstein's denial that there is anything to meaning above and beyond use amounts, in effect, to denying that anything other than performance conditions figures in these issues. In that respect, Brentano's argument is less clean, since he allows an independent role for propositional content. Just as Moore's paradox helps show that reporting and expressing are distinct (see p. 201), so it helps show that content is distinct from force, or performance conditions. It seems impossible to explain the difference between 'It's raining but I don't believe it' and 'It's raining and it's not raining' without invoking some distinction between force and content. Rosenberg's idea that our criteria for the truth of avowals are the same as our criteria for the truthfulness of their performance (note 13) also involves a move from semantics to performance conditions.
- 39 Another consequence of the Wittgensteinian focus on performance conditions, and the consequent idea that ostensible reports actually express one's sensations has to do with J. J. C. Smart's well-known topic-neutral translations of sentences that refer to sensations. According to Smart, the statement 'I have a yellowish-orange afterimage' is roughly equivalent to 'Something is going on in me like what goes on in me when I am visually stimulated by an orange' (Smart, 1959, p. 167). Smart's critics

have rightly stressed that these sentences differ in truth conditions, but perhaps they are, after all, equivalent in respect of performance conditions. Smart sometimes describes the relevant sentences as reports (1959, p. 168; cf. pp. 170–1), but he also concedes finding congenial the “expressive” account of sensation statements’ often attributed to Wittgenstein (1959, p. 162; see the next paragraph in the text). Seeing these statements as expressions rather than reports of perceptual sensations may explain Smart’s persistence in casting his topic-neutral accounts in the first-person singular, a feature of Smart’s treatment that other advocates of the topic-neutral approach have not followed, and that critics have not noted. Still, since statements such as ‘I have a yellowish-orange afterimage’ report, rather than express our sensations, this reconstruction cannot justify Smart’s topic-neutral programme.

- 40 I owe this observation to Margaret Atherton. See Locke (1700, III, ii, 2–3); words all ‘apply to’ (III, ii, 2), ‘are signs of’, ‘signify’, only one’s own ideas. Note, however, that Locke’s claim is that terms refer to the ideas they express, and is not directly about complete sentences.
- 41 ‘Nor can any one apply them . . . immediately to any thing else, but the *Ideas*, that he himself hath’ (Locke, 1700, III, ii, 2).
- 42 Cf. Vendler’s argument: ‘the utterance *I say that I order you to go home* (if it is acceptable at all) amounts to the same speech-act as the utterance *I order you to go home*. Similarly, to think that *p* and to think that one thinks that *p* (if we can speak of such a thing) are the same thought’ (1972, pp. 193–4; see also pp. 50–1). Vendler is plainly relying here on performance conditions and their mental analogues. The same holds for a related matter. If one held that ‘I think that *p*’ expresses, rather than reports my thought that *p*, one might urge that ‘I don’t think that *p*’ expresses my thought that it’s not the case that *p*. Perhaps this would explain why we use ‘I don’t think . . .’ and its kindred not to deny we think that *p*, but to say we think *p* isn’t so. But again, performance conditions explain this more successfully: It is normally appropriate to deny we think that *p* only when we think it isn’t so.
- 43 As Austin points out, ‘I think he did it’ can be a statement about myself, in contrast to ‘I state he did it’, which cannot (1962, p. 134; p. 135 in 1975 edition).
- 44 Similarly, as noted above (p. 202), ‘Ouch, but I’m not in pain’, unlike ‘I’m in pain, but I’m not in pain’, is like a standard Moore’s-paradox sentence: absurd but not contradictory. See also note 11 and pp. 203, 212.

# Consciousness

## *Psychological and Philosophical Essays*

---

*Edited by*

Martin Davies and  
Glyn W. Humphreys



**BLACKWELL**  
*Oxford UK & Cambridge USA*

Copyright © Basil Blackwell Ltd 1993

First published 1993

Blackwell Publishers  
108 Cowley Road  
Oxford OX4 1JF  
UK

238 Main Street, Suite 501  
Cambridge, Massachusetts 02142  
USA

All rights reserved. Except for the quotation of short passages for the purposes of criticism and review, no part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher.

Except in the United States of America, this book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

*British Library Cataloguing in Publication Data*

A CIP catalogue record for this book is available from the British Library.

*Library of Congress Cataloging-in-Publication Data*

Consciousness : psychological and philosophical essays / edited by  
Martin Davies and Glyn W. Humphreys.

p. cm. - (Readings in mind and language; 2)

Includes bibliographical references and index.

ISBN 0-631-18563-1 (alk. paper). - ISBN 0-631-18564-X (pbk. :  
alk. paper)

1. Consciousness. I. Davies, Martin, 1950- . II. Humphreys,  
Glyn W. III. Series.

BF311.C656 1993

126 - dc20

92-25223  
CIP

Typeset in 9<sup>1</sup>/<sub>2</sub> on 11 pt Palatino  
by Graphicraft Typesetters Ltd, Hong Kong  
Printed in Great Britain by TJ Press Ltd., Padstow, Cornwall

This book is printed on acid-free paper