

Multiple Drafts and the Facts of the Matter

Folk psychology conceives of consciousness as a unified phenomenon, which bears some essential connection to a single, unitary self. This encourages a picture on which consciousness occurs at a single place, defined by the interface of certain causes and effects. But the causes and effects of a phenomenon do not always define a unique interface. There is no single place, for example, at which the causes and effects of an enclosed volume of gas all converge. The supposition that in the case of consciousness causes and effects do define a unique interface is a substantive assumption, which might well be mistaken. It is that assumption which underlies what Dennett calls the Cartesian Theater model of mind.

I agree with Dennett that the tacit operation of this model, which is seldom explicitly acknowledged, is responsible for much that's mistaken in current thinking about mind and consciousness. For example, the most promising strategy for explaining consciousness is to build up from phenomena that are mental but not conscious. Tacit adherence to the Cartesian Theater model, however, has led many to see this strategy as unworkable. That's because the afferent pathways leading to mental states are wholly nonmental; so if we locate conscious states at the interface of afferent and efferent, it may well seem that those states literally have no mental antecedents.

For these and many other reasons, Dennett's sustained, effective effort to expose and demolish the myth of the Cartesian Theater constitutes a highly important contribution to our understanding of consciousness. Moreover,

This is a much expanded version of my 'Multiple drafts and higher-order thoughts', *Philosophy and Phenomenological Research*, LIII (December 1993), 911-18. An earlier draft was read at the Society for Philosophy and Psychology Symposium on Daniel C. Dennett's *Consciousness Explained*, June 1993, in Vancouver.

any correct explanation of consciousness will, as I'll argue, share much with Dennett's alternative theory, the Multiple Drafts model (MDM). Many of Dennett's rich and penetrating discussions of conscious phenomena, both mundane and esoteric, provide the essentials of how we must understand those phenomena.

Despite my extensive agreement with Dennett's treatment, however, I want to urge that we can get the substantial explanatory benefits of Dennett's MDM with a theory that, in one respect, is somewhat weaker. In what follows I shall argue that this is so, indicating along the way the many respects in which Dennett's theory is exactly on target.

I. Multiple Drafts and the Revisability of Consciousness

Perhaps the most important virtue of the MDM is its ability to accommodate the temporal anomalies Dennett discusses. In colour phi, for example, alternating red and green flashes occur, respectively, on the left and right sides of the subject's visual field, but subjects seem to see a single spot that moves and changes colour. In the so-called cutaneous rabbit, three bunches of physical taps, at the wrist, elbow, and upper arm, result in the subject's feeling a sequence of single taps along the arm separated by small distances. Why is it that one doesn't first consciously sense the initial, stationary, red flash, and the initial bunch of taps at the wrist? Dennett believes the MDM has the answer.

According to that model, consciousness is continuously revised, much as a text changes through successive drafts. Some features of a text undergoing revision will persist through many drafts; others may be so transitory as to escape notice altogether. Similarly with consciousness. When we try to determine what conscious experience a subject has, eliciting reactions from the subject at different times may well result in different reports. At successive moments we may well, as Dennett puts it, 'precipitat[e] different narratives . . . [different] versions of a portion of "the stream of consciousness"'.¹ Nor is there any privileged moment at which a report would reveal the true nature of the subject's conscious experience; even first-person, introspective impressions of our own experiences sometimes vary through time. Reporting our experiences may seem to be the last word on this matter, much like publishing a text. But even publication fixes a text only relative to a social context; post-publication revision can and does occur. Similarly with reports of conscious experiences.

These points help explain colour phi and the cutaneous rabbit. If the initial stationary flash enters the subject's consciousness at all, its presence is so transitory that it escapes notice altogether. Perhaps at some early moment we could elicit a report of the initial red flash; but if we don't, it is for the subject as though no conscious experience of that flash ever occurred.

¹ Dennett 1991: 135. Unless otherwise noted, page references throughout are to this work.

This much is fairly straightforward. In many cases we simply won't know, from either a first- or a third-person point of view, whether or not a particular stimulus makes it to consciousness. But Dennett goes one step farther, and

makes 'writing it down' in memory criterial for consciousness . . . There is no reality of consciousness independent of the effects of various vehicles of content on subsequent action (and hence, of course, on memory) (132).

Accordingly, Dennett rejects both the 'Stalinesque' view, which says that the initial flash or taps are edited out prior to consciousness, and the opposing 'Orwellian' claim that the flash reaches consciousness but is immediately forgotten. Rather, he maintains, when no early reaction is elicited, there is simply no fact of the matter about whether the initial stimulus ever becomes conscious. '[T]here are no fixed facts about the stream of consciousness independent of particular probes' (138; cf. 275).

II. Temporal Anomalies and Higher-Order Thoughts

Central to the MDM is the idea that it's not fixed, from one moment to the next, what our conscious experiences are. Any satisfactory explanation of consciousness plainly must accommodate such revisability. But it's less clear that we must also adopt Dennett's '*first-person operationalism*' (132): his denial that there's any fact of the matter about whether certain stimuli reach consciousness, and about what our conscious experiences are at any particular moment.

We can explore this question by seeing whether the explanatory virtues of the MDM remain when we subtract first-person operationalism, that is, when we adopt a theory that accommodates the revisability just described but denies the operationalism. One such theory is the higher-order-thought explanation of consciousness that I've put forth elsewhere, and which Dennett discusses in chapter 10.²

I use 'thought' here as a term of art for any episodic intentional state with an assertoric mental attitude. When a mental state is conscious, one is conscious of being in that state. Occasionally one is conscious of the state introspectively, but typically one is conscious of it in a way that isn't at all attentive or reflective. On the higher-order-thought hypothesis, one is conscious of being in a conscious state by virtue of having a roughly contemporaneous thought to the effect that one is in that state. Because that thought is about another mental state, it's convenient to call it a *higher-order thought* (HOT).

HOTs must be episodic and their mental attitude must be assertoric, because dispositional states and nonassertoric states do not make us con-

² See Rosenthal 1986a; 1990; 1991; 1993; 1996.

conscious of the things they are about. Moreover, we sometimes infer that we're in some mental state even when that state isn't conscious; so we must stipulate that the HOT is independent of any inference of which we're aware. That restriction will also account for our sense that we are conscious of our conscious mental states in some way that is intuitively immediate. It's enough here to say why it seems that nothing mediates between our conscious states and our being conscious of them. We need not adopt a model on which nothing actually mediates between the two, since we have no reason, independent of our having the intuition about immediacy, to think that that intuition is true.

Normally, of course, we aren't conscious of any HOTs of the sort this theory posits. But that's to be expected, given the HOT hypothesis. A HOT wouldn't be conscious unless one has a third-order thought about it, and it's reasonable to think that that rarely happens. Indeed, cases in which that does happen are cases in which we are introspectively conscious of our mental states, rather than conscious of them in the ordinary, nonintrospective way.

The HOT hypothesis accommodates the various temporal anomalies no less well than the MDM. For example, the initial flash in colour phi presumably causes a sensory state of a stationary red spot. But it may well be that no HOT about that sensory state occurs; so the sensory state never becomes conscious. After the second flash occurs on the right, a new sensory state occurs, of a moving spot that changes colour. A HOT about this second sensory state then occurs, giving one the conscious experience of a moving, changing spot.

There's also, however, a second possibility. A HOT might occur about the initial sensory state of a stationary red flash, but be replaced so fast by the new HOT about the sensory state of a moving spot that the first HOT doesn't register mentally. It doesn't last long enough to affect our memory of what we saw, and it has no other noticeable effects. In particular, the HOT goes out of existence before there's time to report that one is in the first sensory state, that is, to *express verbally* the HOT in question. So the first sensory state is technically conscious, but so briefly as to make no mental difference.

These two possibilities instantiate the Stalinesque and Orwellian models, respectively. Because the two scenarios are introspectively indistinguishable, Dennett is right that introspection can't determine which of the two has occurred. And he's also right that nonverbal behavioural reactions won't help, since, as he convincingly argues, any such reaction could be due equally to a state that's conscious or to a state that's not (124).

It doesn't follow, however, that there is no way to tell which model is operative in a particular case. Suppose, for example, we have a theory about consciousness that draws the right distinctions in unproblematic cases. We might then be able to apply that theory to our puzzle cases to determine, independently of how it seems to the subject, whether a particular stimulus

does or does not make it to consciousness. Indeed, the HOT hypothesis would do just that, since the Orwellian model in effect posits an initial HOT that on the Stalinesque model doesn't occur. We don't now know, of course, which model explains colour phi or the other temporal anomalies. Perhaps some anomalies are Orwellian and others Stalinesque; perhaps some have both Stalinesque and Orwellian instances. Still, explaining these phenomena doesn't require us to deny that there's a fact of the matter about which is operative in any particular case.³

III. Other Explanatory Advantages

Arguably, the HOT hypothesis explains the many other mental phenomena Dennett considers at least as well as the MDM. Moreover, apart from first-person operationalism, the MDM and the HOT hypothesis agree about most theoretical issues. Consider, for example, consciousness and unity. On the Cartesian Theater model, our mental lives are unified because all conscious states occur in a single place. The MDM insists instead that consciousness is a distributed phenomenon; there's no one place where consciousness occurs, and no single location to direct probes in determining what conscious state a subject is in (136). So there's no unique 'Central Meander' and no unique 'Author of Record' (228). The HOT hypothesis agrees. Distinct mental states are conscious because of distinct HOTs, which presumably occur in different locations.

Dennett explains the apparent unity of consciousness as due to narratives we construct that represent our mental states as belonging to a single mental life. The HOT hypothesis takes a similar line; unity results from HOTs whose content subsumes groups of mental states, and from third-order thoughts whose content connects, in turn, various second-order thoughts, and makes one conscious of the reference those thoughts ostensibly make to a self.⁴ Indeed, Dennett's narratives are, in effect, just the expressions in speech of these HOTs.

On the MDM, a mental state is conscious if it leaves significant traces in memory, and has other substantial mental and behavioural effects. But not all the effects a mental state has will in fact be relevant to whether it's conscious. As Dennett notes, conscious and nonconscious mental states can have exactly the same effects on nonverbal behaviour. Similarly, most of

³ We often adopt first-person operationalism in designing experimental paradigms to test a theory, but once established the theories can leave such methodological operationalism behind.

⁴ The self to which HOTs refer may be whatever minimal self is involved in a creature's distinguishing itself from everything else — roughly, the creature's body. Or it might instead be just a notional intentional object of those thoughts, much as the assumption of a unique Author of Record by heterophenomenological reports is, according to Dennett, just an interpretive idealization (228).

the mental traces left by conscious states could equally well have been left by mental states that aren't conscious. This is true even for effects on memory, which Dennett counts as criterial for consciousness. One sometimes sees something without being at all conscious of seeing it, even though one may later, perhaps to one's surprise, realize that one saw it. In such cases, nonconscious perceiving has a significant, lasting effect on memory.

A mental state may have many mental effects without becoming conscious, but not if it causes a HOT. Having an assertoric thought about something implies being conscious of it. So if one comes noninferentially to have a thought that one is in a particular mental state, that state is then conscious. The HOT hypothesis thus focuses more closely than the MDM on just which mental effects are necessary for a mental state to become conscious.

If a state's being conscious hinges on its having certain effects, then it isn't conscious unless certain subsequent events occur. The temporal anomalies are cases in which it seems evident that a state is conscious only if some subsequent event does occur. But we can describe this without any air of paradox. Because many mental states aren't conscious at all, it's implausible that the property of being conscious is an intrinsic property. All mental states have some sort of content properties — intentional content in the case of intentional states and sensory content in the case of bodily and perceptual sensations and most emotions. Such content properties are arguably intrinsic to mental states.⁵ By contrast, mental states can be conscious at one moment and not at another; so we have no reason to regard the property of being conscious as being intrinsic to such states. Accordingly, a state's being conscious requires the occurrence of something extrinsic to it. And it may well be, therefore, that no mental state is conscious when it first occurs. But this doesn't mean there are no facts of the matter about consciousness; states are conscious when, and only when, the relevant events occur.

Suppose a person has dispositions to make conflicting reports or, equivalently, has conflicting HOTs. One HOT, or disposition to report, might represent the person as being in a particular state, while the other represents the person as not being in it. That still wouldn't show, however, that there's no fact of the matter about whether the state is conscious. A state is conscious just in case we're able to report it noninferentially, whether or not there's also a concurrent disposition to deny its existence. Or the HOTs and dispositions to report might compete in a different way, by attributing distinct properties to the state. The facts of consciousness then embrace both ways of representing the state.

⁵ Also, some mental attitude, such as that of believing, desiring, hoping, wondering, doubting, and the like, is presumably intrinsic to each intentional state.

When one has conflicting HOTs or conflicting dispositions to report mental states, an experimental probe may settle things by causing one of the two to stop existing. The probe would in that way determine the facts of consciousness. Does that mean that a subsequent event determines, after the fact, whether a particular mental state had earlier been conscious, and with respect to which of its mental properties? The probe does retrospectively settle what conscious states one takes oneself to have been in, at least if one doesn't recall the earlier, conflicting HOT or disposition to report, which no longer exists. But the probe does not alter what the facts were prior to it. Subsequent events may make a difference to how the prior facts of consciousness now seem, but not to what those facts were.⁶

Dennett emphasizes that it's not always clear, even from a first-person point of view, whether one is conscious of something. And he vividly illustrates this by the game of 'hide the thimble' (336), in which one may look straight at an object one's trying to find and yet fail to register it. This kind of case poses a problem. The sensory states at the centre of one's visual field are normally conscious; so if one's looking at the hidden thimble, how can one fail to see it consciously? The difficulty we have in describing this kind of case from a first-person point of view seems to underwrite Dennett's claim that, independent of particular probes, there's no fact of the matter about what conscious experiences we have.

We're seldom if ever conscious of all the detail that's represented in our sensory states, even states at the centre of our visual field. And how much detail we're conscious of often changes. Moreover, it needn't be the sensory states that change, but only how we're conscious of them. The HOT hypothesis enables us to explain these things. HOTs represent sensory states in greater or lesser detail. So, one HOT might represent a sensory state as being just of a bookcase with lots of things on it, while another HOT a moment later might represent that very sensory state in greater detail, for example, as including a thimble. In the first case one is conscious of seeing the bookcase but not the thimble; in the second case one is conscious of seeing both. All this is independent of first-person operationalism.

As Dennett notes, training in such things as piano tuning and wine tasting can change what conscious experiences one has. It's hard to believe that this training, by itself, results in new kinds of sensory states; rather, we acquire new discriminatory concepts for our experiences, and the resulting HOTs provide finer detail in the way we're conscious of those experiences. Dennett speculates that if blindsight patients became self-cuing, the states in their 'blind' hemifield would become conscious. The HOT hypothesis suggests why that might happen. If one must guess about the contents of one's blind hemifield, one's intentional states about the sensory contents of that hemifield won't have the assertoric mental attitude required for HOTs.

⁶ Memory helps determine how the facts of consciousness appear to one, even though "writing it down" in memory is not 'criterial for consciousness'.

The ability to be self-cuing would dispel the hesitant attitude involved in just guessing, leading thereby to one's having assertoric HOTs. Moreover, guessing lacks the intuitive immediacy characteristic of the way we're conscious of our conscious states. When we guess, we regard our tentative judgments as mediated by something, even if we have no idea what that mediating factor is.

Dennett convincingly argues that, given the low resolution of parafoveal vision, we see a wide area of wallpaper as being all Marilyns by representing a few foveal Marilyns and, in effect, judging that the parafoveal shapes are 'more of the same' (355). The HOT hypothesis suggests how this might happen. One's sensory states represent foveal shapes as Marilyns and peripheral shapes as largely indistinct. One's HOT cleans things up, then, by representing the shapes in both foveal and peripheral sensory states as being all Marilyns.⁷ And more generally, the HOT hypothesis fits comfortably with Dennett's view that so-called 'filling in' takes place not by the brain's manufacturing the missing sensory states, but by its forming suitable judgments. (See Dennett 1993: 205–10.)

IV. The 'Objectively Subjective'

Dennett recommends his heterophenomenological method in part because it helps undercut the apparent opposition between first- and third-person viewpoints. It does this, he urges, by being neutral about whether a subject's report truly describes that subject's first-person viewpoint or simply expresses beliefs about the subject's mental states, states which may be merely notional. We'll count those reports as describing real events if what they say accords reasonably well with events we know about independently, say, brain events.⁸ The HOT hypothesis achieves the same methodological neutrality, since heterophenomenological reports express HOTs, and HOTs aren't factive.⁹

⁷ For all we now know, of course, the sensory state with foveal Marilyns might actually produce peripheral Marilyns in the sensory state itself. But as Dennett emphasizes, we typically take parafoveal vision to be far more crisp and detailed than it could be, even given saccadic movements. So it's more likely that the apparent parafoveal Marilyns result in part from the brain's interpretive activity. The same may even be true of the influence foveal colour has on what we take surrounding colours to be.

⁸ Taking heterophenomenological reports to be evidentially basic might seem at first sight to make those reports incorrigible. And that in turn suggests first-person operationalism, since how consciousness subjectively seems would then be the last word about its nature. But heterophenomenological reports aren't actually incorrigible in the way needed for first-person operationalism if, as Dennett urges, we judge their truth by way of brain events (85).

⁹ Accordingly, the MDM and HOT hypothesis agree about zombies (and zimboes [310–11]). Dennett denies that anything 'in principle . . . indistinguishable from a conscious person' could lack conscious states (405; cf. 282). And on the HOT hypothesis, since having HOTs is sufficient for a state to be conscious, nothing could be in all the intentional states we are in but lack conscious states. Again, this is independent of first-person operationalism.

Dennett propounds first-person operationalism in part because its denial 'creates the bizarre category of the objectively subjective — the way things actually, objectively seem to you even if they don't seem to seem that way to you' (132). Moreover, he argues that it's a failing of the HOT hypothesis that it accepts that category (316).

There are, however, good reasons to sustain a distinction between seeming and seeming to seem. Being in a sensory state defines, in one respect, how things seem to me. That's because of the connections the state has to other aspects of my mental life, even if the state fails to be conscious. But when a sensory state isn't conscious, I'm not aware of being in it; so it doesn't then seem to me that I am in it. That's the second level of seeming.¹⁰ Dennett might deny that being in a sensory state that isn't conscious can define how things seem to one, since he holds that a state's being conscious consists in its leaving significant traces and having wide-spread mental effects. But as I argued earlier, mental states can have many mental effects without being conscious. Moreover, Dennett accepts that we're sometimes mistaken about what mental states we're in, and it's hard to see how that could happen if the way things actually seem to one always coincides with the way they seem to seem.

Dennett might urge that ascribing representational properties to sensory states incurs all the problems that face qualia and the mental 'pigment' (346) he argues against so effectively. But it's arguable that mental pigment and qualia are problematic only because we conceive of them as being intrinsically conscious. (See Rosenthal 1991.) On the HOT hypothesis, the representational properties of sensory states need not occur consciously.

Dennett notes that distinguishing mental states from the corresponding HOTs makes room for an unexpected kind of error; my HOT might misrepresent what kind of mental state I'm in (317). Moreover, any third-order thought I have might in turn be wrong about what my second-order thoughts are.¹¹ Even conceding that we're sometimes mistaken about our mental

¹⁰ Dennett seems to say as much: the 'onsets [of content-fixations in the brain] do *not* mark the onset of consciousness of their content' (113; emphasis Dennett's). (But see p. 12 ?? below.) Not distinguishing these two levels, moreover, risks representing consciousness as an intrinsic property of mental states, which would accord poorly with Dennett's idea that consciousness is a distributed phenomenon, and with his denial that there's any fact of the matter about when a mental state becomes conscious. And, if being conscious were an intrinsic property, it's unlikely we could explain what it is for a state to be conscious without appealing to the very consciousness we were trying to explain. (One might even interpret the Cartesian Theater model as holding that consciousness is an extrinsic property: a mental state is conscious just in case it occurs in the right location, viz., the Cartesian Theater. Cf. 144.)

¹¹ Dennett takes the HOT hypothesis to posit not just HOTs, but also higher-order beliefs distinct from those HOTs, apparently because he assumes that that reports of conscious states express these beliefs (307; cf. 317). He concludes that error might also occur between HOT and higher-order belief. But since reports of conscious states directly express HOTs, the HOT hypothesis doesn't posit distinct higher-order beliefs at all.

states, Dennett sees these proliferating levels of possible error as another reason to reject the HOT hypothesis, since he thinks it's idle to distinguish among mistakes at all these different levels.¹²

The view that introspection 'must give exhaustive and infallible information' is, as C.D. Broad remarked, 'a curious superstition' (Broad 1925: 284). There are always features of our conscious states we might be aware of but aren't; consider our sensation of the scene that includes the thimble we don't notice. Why, then, shouldn't the way we're conscious of mental states sometimes represent them as having features they lack? There's evidence from clinical contexts and from social psychology¹³ that this sometimes occurs. And many of the repressed states familiar from clinical contexts and Freudian theory are very likely not literally unconscious states, but rather states whose intentional properties subjects represent in a radically disguised way, so as to avoid having to face what states they are actually in.

Even introspective impressions, moreover, may occasionally be erroneous if one's third-order thought misrepresents the content of a second-order thought, perhaps because a theory had influenced what one expected to find introspectively. Once again, it doesn't matter if introspection can't distinguish this kind of error from errors at the first level, since in any particular case theoretical considerations should enable us to tell which has occurred.

Dennett sees the hierarchy posited by the HOT hypothesis as reflecting our folk-psychological conceptions, and so regards both as committed to these distinct levels of possible error. It's unclear, however, that these distinct levels of error are unequivocally part of our folk psychological picture.¹⁴ In any event, the HOT hypothesis is a theoretical proposal about what it is for mental states to be conscious, and so can both supplement and depart from our folk-psychological conceptions.

¹² Dennett is concerned that if the HOT hypothesis allows an indefinite hierarchy of HOTs, there will be indefinitely many distinct errors. But it's unlikely that there's any reason to posit thoughts at a level higher than third-order.

Dennett also urges that a hierarchy of HOTs undermines 'the subjective intimacy or incorrigibility that is supposedly the hallmark of consciousness' (319). But the HOT hypothesis preserves the appearance of subjective immediacy, both because HOTs typically aren't conscious, and because HOTs aren't due to any conscious inference, and so we can't tell from a first-person point of view how we come to be conscious of our conscious states.

¹³ Nisbet & Wilson (1977), who present evidence that subjects confabulate stories not only about what the causes are of their intentional states, but also about what intentional states they are in.

¹⁴ Folk-psychological categories clearly allow distinguishing hierarchical levels of HOTs. But HOTs are theoretical posits not demanded by folk psychology, in part because we're typically unaware of them from a first-person viewpoint. Moreover, folk psychology seems to allow individuating intentional states solely by way of the mental analogue of performance conditions, ignoring the truth conditions and mental attitudes of the states. And that would mean that when a HOT is about an intentional state, we would identify the two.

Since subjects' reports express the HOTs that accompany conscious states and HOTs can misrepresent those states, subjects' reports may be inaccurate. So, as Dennett argues, there's no privileged moment at which a subject's report reveals the true nature of a conscious experience, since the report issued at any particular moment might be erroneous. But the occurrence of erroneous HOTs doesn't sustain first-person operationalism. The facts about what mental states we're in are distinct from the facts about how we're conscious of them, and HOTs pertain only to the latter.

Partly because we often discover what we think only as we say it (245), Dennett argues that our choice of words can influence the content of our thoughts (247). He concludes that HOTs don't always underlie reports of being in mental states (315). But our discovering what we think as we say it needn't show that our words affect our thoughts; rather, it may simply be that our thoughts often aren't conscious until we express them. And even when one's choice of words does affect what one thinks, one's speech act still expresses the resulting intentional state. So reports of mental states will always express HOTs in virtue of which those states are conscious.¹⁵

V. Multiple Drafts and Mental Taxonomy

Dennett sees as artificial the way we ordinarily carve consciousness and mind into discrete mental states. Worse, he thinks the folk-psychological taxonomy in terms of which we do so requires 'postulating differences that are systematically undiscoverable.' It's at bottom because the HOT hypothesis and the Stalinesque and Orwellian models all presuppose a folk-psychological taxonomy of mental states that he rejects both the HOT hypothesis and the distinction between Stalinesque and Orwellian explanations (319).¹⁶

¹⁵ Dennett in effect agrees, when he writes that '[t]he emergence of the [verbal] expression is precisely what creates or fixes the content of the *higher-order thought expressed*' (315, my emphasis).

In Rosenthal 1993, I assumed a causal model of what it is for a speech act to express an intentional state on which, roughly, an intentional state causes the speech act that expresses it. (I argue for this in Rosenthal 1986b.) That model runs counter to Dennett's Pandemonium model of speech production, on which speech performances affect the content of the thoughts expressed. But this doesn't matter for the HOT hypothesis, which requires only that speech acts express an intentional state with the same content and a mental attitude corresponding to the speech act's illocutionary force. The hypothesis makes no assumptions about what causal relation, if any, speech acts bear to the intentional states they express.

One might urge that, although a state is conscious just in case it's reportable and reporting a state is the same as expressing a HOT that one is in that state, the HOT needn't exist until an actual report occurs. If so, unreported conscious states need not be accompanied by HOTs. But the ability to report must be based on something occurrent, and an actual HOT is the most reasonable candidate.

¹⁶ Since a state's being conscious begins, on the HOT hypothesis, at the moment a HOT comes to exist, we can in principle tell whether revision occurs before or after that

As noted above, the HOT hypothesis isn't wedded to folk-psychological categories; rather, it uses them provisionally, in the way any theory works off our commonsense conceptions. Still, the issue about taxonomy will help to get a deeper understanding of Dennett's view.¹⁷ Dennett often writes of such things as 'events of content-fixation' (365) 'information-bearing events' (459), and 'content-discriminations' (113). More important, these 'content-fixations . . . are [each] precisely locatable in both space and time' (113). Since Dennett denies that we can precisely locate conscious mental phenomena, content fixations plainly aren't conscious. Exactly what are they, then, and how do they relate to those mental phenomena that are conscious?

In the early stages of vision, colour, form, orientation, and motion are represented independently. There's presumably no problem about locating these representations precisely. Moreover, except possibly for pathological cases, independent representations of colour and shape do not occur consciously.¹⁸ So it's natural to take Dennett's 'content-fixations' as being something like these early representations. Consciousness arises, then, in the course of subsequent integrative processes that represent these fragmentary representations as being interconnected. At any one time, there may be more than one interpretive process that serves to integrate a particular group of representations; the several editorial processes then constitute multiple drafts of the contents of consciousness. Each integrative process, moreover, will involve dispositions to produce narratives about one's mental life. When there are several concurrent editorial processes, the narratives they dispose one toward may well conflict. Only when 'particular probes' (138) occur will one integrative process drive out the others, thereby settling the facts of consciousness.

moment. But an example will make clear how taxonomy can be crucial. Suppose that these events occur in very close succession. A sensory state occurs, it becomes conscious by virtue of a HOT's coming to exist, the sensory state then undergoes an alteration, a second HOT with correspondingly revised content then comes to exist, and when it does the first HOT ceases to exist leaving no trace in memory. Does the HOT hypothesis count this as Stalinesque or Orwellian revision? Relative to the revised sensory state, the revision is Stalinesque, since the revision occurred before the revised state was conscious; relative to the original state, the revision is Orwellian. In effect, we'll count the revision as Stalinesque if we regard the revised state as a new state, distinct from the original, and Orwellian if we regard the revised and original states as two stages of a single state. Not only is there arguably no preferred way to take things; either choice reflects, in addition, a commitment to artificially precise identity conditions for mental states. This doesn't show, however, that there aren't facts of the matter about exactly when revision occurs and when states with particular content properties become conscious. Rather, the Stalinesque and Orwellian models aren't, as they stand, sufficiently precise to sustain an unambiguous contrast.

¹⁷ For more on the following issues, see Rosenthal 1995.

¹⁸ The evidence for independent representations of colour and shape in early visual processing is inferential, mainly from experiments in which subjects report seeing illusory conjunctions of the colours and shapes of distinct, simultaneous stimuli. Subjects do not report seeing either shape or colour independently of the other. See Treisman 1982; 1986; 1988.

Mental states, conceived folk psychologically, are whatever states, if any, conform to descriptions that occur in our heterophenomenological reports. Such states may turn out to be merely notional objects of these reports, and Dennett seems to think that all that exists are the locatable early representations and the subsequent editorial processes that integrate them.¹⁹ If that's right, our folk-psychological taxonomy of ordinary mental states involves a highly artificial and unrealistic determinateness; in particular, there would then be no HOTs. More important, if no mental phenomena exist other than content fixations and editorial processes, there will be no facts of the matter about conscious mental states conceived in folk-psychological terms. First-person operationalism would then be true for conscious states so taxonomized.

But taxonomy aside, integrative processes do not always result in consciousness. For example, the complex processes posited by cognitive theories often serve to integrate, but are seldom if ever conscious. And many of the states these processes lead to also aren't conscious. Moreover, integrative processes won't help explain why we seem to be in conscious states, folk psychologically taxonomized. Such integrative processes may help us understand why certain properties come together, but not why we seem to ourselves to be in conscious states with those concurrent properties.

HOTs, by contrast, do help with these things. Even if there are no conscious states as folk psychology conceives of them, we could explain why we think we're in such states by positing nonconscious HOTs that have such states as their intentional objects. There would then be determinate facts about consciousness, though they would consist solely of the facts about our nonconscious HOTs. It's reasonable, therefore, to hope that we can explain consciousness by a version of the MDM that does not appeal to first-person operationalism.

References

- Broad, C.D. (1925). *The Mind and Its Place in Nature*. London: Routledge & Kegan Paul.
- Dennett, D.C. (1991). *Consciousness Explained*. Boston, MA: Little, Brown & Co.
- Dennett, D.C. (1993). Back from the drawing board. In Dahlbom 1993.
- Nisbett, R.E. & T.DeC. Wilson (1977). Telling more than we can know: Verbal reports on mental processes, *Psychological Review*, LXXXIV, 231–59.
- Rosenthal, D.M. (1986a). Two concepts of consciousness. *Philosophical Studies* 49, 329–59.
- Rosenthal, D.M. (1986b). Intentionality. *Midwest Studies in Philosophy*, 10, 151–84.

¹⁹ The eliminativist strain here is offset by Dennett's methodological strategy of taking heterophenomenological reports being as evidentially basic, and using brain events to assess what, if anything, those reports are true of (85).

- Rosenthal, D.M. (1990). *Why Are Verbally Expressed Thoughts Conscious?* Report 32, Center for Interdisciplinary Research (ZiF), Research Group on Mind and Brain, University of Bielefeld.
- Rosenthal, D.M. (1991). The independence of consciousness and sensory quality. In Villanueva 1991.
- Rosenthal, D.M. (1993). Thinking that one thinks. In Davies & Humphreys 1993.
- Rosenthal, D.M. (1995). First-person operationalism and mental taxonomy. *Philosophical Topics*, 22, 319-50.
- Rosenthal, D.M. (1996). A theory of consciousness. In Block *et al.* 1996.
- Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology*, 8, 194-214.
- Treisman, A. (1986). Features and objects in visual processing. *Scientific American*, 255, (November), 114-25.
- Treisman, A. (1988). Features and objects: the fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*, 40A, 201-37.

Conscious Experience

edited by

Thomas Metzinger

Imprint Academic
Schöningh
1995