First-Person Operationalism and Mental Taxonomy
Author(s): David M. Rosenthal
Published by: University of Arkansas Press
Stable URL: http://www.jstor.org/stable/43154661
Accessed: 11-02-2017 13:45 UTC

REFERENCES
Linked references are available on JSTOR for this article:
You may need to log in to JSTOR to access the linked references.
First-Person Operationalism
and Mental Taxonomy

David M. Rosenthal
City University of New York, Graduate School

I. MULTIPLE DRAFTS AND FIRST-PERSON OPERATIONALISM

Any acceptable theory of consciousness must plainly satisfy two main constraints. It must, first of all, do reasonable justice to our common-sense, folk-psychological intuitions about consciousness and mentality. Some of these intuitions will be helpful in identifying just what phenomena the theory seeks to explain. Others the theory will predict, using its explanatory machinery. Still others the theory may be neutral about, avoiding any conflict between theory and intuition. Also, a theory may jettison certain intuitions, ruling in effect that they do not reflect the nature of conscious phenomena; but the theory must then explain why those intuitions, despite being mistaken, still strike us as compelling.

A theory of consciousness must also square with what scientific research tells us about the brain and about the conscious functioning of people and other animals. The theory must, in particular, help make it intelligible how brain mechanisms operate in producing conscious experiences, and it must take account of experimental and clinical findings.

Theories often seem to do a lot better with one of these tasks than the other. Many theorists, for example, would take Thomas Nagel's account of
consciousness in terms of subjective points of view as setting a standard for successfully capturing our folk-psychological intuitions. And this may well lead to the kind of skepticism Nagel himself holds about whether those common-sense intuitions can be made to fit with a scientific account of these things. Similarly, whatever the merits of neurobiological explanation, such as the suggestion by Francis Crick and Cristof Koch that consciousness arises from neuronal oscillations close to forty hertz, it isn’t easy to see how such explanations could do justice to our folk-psychological intuitions about consciousness. Indeed some, such as Patricia Smith Churchland, have argued that a scientific theory of consciousness will require that many of these intuitions be rejected.

Against this background, the theory of consciousness developed in Daniel Dennett’s impressive and important book, Consciousness Explained, occupies a useful middle ground. One of Dennett’s main concerns is to describe a model for explaining consciousness which takes account of relevant results in experimental cognitive psychology and the neurosciences. Consciousness, he argues, results from a number of interacting brain processes, which are constantly changing due to new stimuli and feedback from other brain processes. Because the interactions among the relevant processes are continually being updated, the way consciousness represents our mental lives is not fixed from moment to moment.

At any particular moment, therefore, there may be many competing interactions among the brain processes, each capable of giving rise to consciousness. And the interactions competing at any one time may represent the contents of consciousness in different ways. Which interactions lead to conscious results, moreover, may often be a matter of chance factors, factors, that is, extrinsic to the interacting processes. Some external stimulus, or “probe,” may push things in one direction or another. And often there will be a succession of interactions that yield conscious results, producing successive versions—or “drafts”—of the conscious goings-on within the person.

Dennett calls this model for explaining consciousness the Multiple Drafts model (MDM). He contrasts it with what he calls the Cartesian Theater model, on which a mental state’s being conscious is a matter of its being observed, somehow, in a “theater of consciousness.” On the MDM, consciousness is a function of processes occurring in a distributed way throughout much of the brain, whereas the Cartesian Theater model holds that states are conscious just in case they occur at a single privileged location. Dennett argues forcefully against the Cartesian Theater model, which, he plausibly maintains, tacitly underlies the explanatory strategies used by many theorists.

Dennett is concerned not merely to make intelligible how consciousness arises in the brain, but at the same time to do justice to our folk-psychological intuitions about consciousness. But he departs from many
theorists in how he gets at these intuitions. People seem to have direct access to their own introspective data; so it's tempting to claim that such data are decisive about mental reality. But introspective data differ notoriously from person to person, and even for single subjects their reliability is questionable. Dennett therefore takes the principal evidence about conscious mental phenomena to be the verbal reports people make about their mental lives. These "heterophenomenological" reports provide the common-sense information about our mental lives to which a theory must do justice. Because introspectible events aren't directly accessible to others, heterophenomenological reports provide a measure of objectivity that direct reliance on introspection cannot.

Discounting occasional insincerity, heterophenomenological reports are authoritative about how people's mental lives seem to them. Dennett emphasizes that this doesn't mean that these reports are also authoritative about the states and processes that underlie those appearances. Even the mental events to which a subject's heterophenomenological reports ostensibly refer may not always exist. We should take heterophenomenological reports to refer to actual events only if those reports are corroborated by what we know independently, say, about brain events. Dennett's methodological reliance on heterophenomenological reports is therefore neutral about whether such reports truly describe mental events that go into a subject's first-person viewpoint, or whether they simply express beliefs about the subject's mental events, events which may be entirely notional. It's a special strength of Dennett's approach that this neutrality enables the heterophenomenological method to bridge the traditional gulf between first- and third-person accounts of mind and consciousness. We need not, accordingly, have a theory based just on first- or third-person considerations. This approach therefore allows Dennett to weave together heterophenomenological data with the findings of neuroscience and cognitive psychology in a way that does reasonable justice to both.

Heterophenomenological reports may well provide the best evidence about how people's conscious mental lives appear to them, though these reports may not always tell us about the states and processes that underlie those appearances. But Dennett goes further and insists that, whatever the case with underlying processes, there is nothing more to people's conscious experiences than how those experiences appear. When it comes to people's conscious mental lives, there is no distinction to be drawn between the reality of conscious experiences and how those experiences appear to those who have them. Consciousness consists in things' appearing in certain ways; so appearance is all there is to the reality of consciousness. Dennett calls this view first-person operationalism (FPO). It's a form of operationalism because appearance determines reality; but it's operationalism restricted to the first-person case, that is, to "the realm of subjectivity."
Dennett regards his MDM as a form of FPO. But it's possible to separate Dennett's denial of the distinction between the appearance and reality of conscious experience from other aspects of his model that are independent of this denial. In what follows, I'll refer to Dennett's denial of that distinction as FPO, and I'll apply the label 'MDM' more restrictively to those aspects of his model that are independent of FPO. I'll argue that, although Dennett's MDM narrowly understood in this way is as promising a model as we now have for explaining consciousness, we need not also adopt his thesis of FPO. I'll do this by defending a model of consciousness very similar to the MDM, narrowly construed, and by arguing against the addition of FPO.

II. TEMPORAL ANOMALIES AND FACTS OF THE MATTER

Perhaps the most important of the empirical findings against which Dennett tests his model has to do with the remarkable temporal anomalies he describes in chapters five and six of *Consciousness Explained* and in his earlier article with Marcel Kinsbourne. The ability to deal satisfactorily with these anomalies is one of the greatest strengths of Dennett's MDM. So in trying to assess whether FPO is a necessary aspect of Dennett's view, it's useful to begin with these curious results.

Two anomalies will suffice to give the flavor. In color phi, a subject is presented with alternating red and green flashes on the left and right, respectively, but seems to see a single spot that moves and changes color. In the so-called cutaneous rabbit, three successive bunches of physical taps are administered, at the wrist, elbow, and upper arm, but the subject feels a sequence of single taps along the arm, evenly separated by small distances. Subjects presented only with the initial red flash on the left, or only the bunched taps at the wrist, consciously sense these things in just that way. Why is it that when these initial stimuli are followed by the others, subjects do not consciously perceive the initial red flash nor the bunched wrist taps?

Dennett believes the MDM has the answer. On that model, conscious mental states are due to the interaction of many brain processes. These interactions are of course not instantaneous, but take some time to occur. So an interaction that would ordinarily lead to some particular conscious experience can, in effect, be derailed by a suitably timed stimulus. If presented with only a red flash on the left, the subject consciously experiences it. But the subsequent stimulus of a green flash on the right interferes with the interaction that would have led to a conscious experience of a red flash, replacing it with the conscious experience of a moving spot that changes color. Similarly with the cutaneous rabbit.
Among the intervening stimuli that may alter the course of the interactions leading to a conscious experience is the eliciting of some reaction from a subject. And a small change in timing can of course be crucial. At successive moments we may well “precipitate different narratives... different versions of a portion of ‘the stream of consciousness.’” Because of this, there is no privileged moment at which eliciting a report would reveal the true nature of the subject’s conscious experience. Any event that serves as a probe for a subject’s heterophenomenological report may affect the contemporaneous interactions leading to specific conscious results.

Even in the temporal anomalies, when an event following the initial stimulus derails things, it might still have been possible to elicit a report of the initial stimulus if we timed our probe sufficiently precisely. In color phi, the probe would presumably have to occur after the stationary red flash affected the visual cortex, but before the first green flash interrupted the normal resulting processes. Such a probe might block the processes leading to the conscious experience of a moving spot that changes color, allowing the subject to consciously experience the initial stationary flash. But if no such probe intervenes, there will be no conscious experience of the initial stimulus.

Dennett compares the situation to the revising of texts. When a text changes through successive drafts, some features typically persist through many drafts; others may be so transitory as to altogether escape one’s notice. Similarly with consciousness. Successive interactions among brain processes lead to different ways in which consciousness represents our mental lives, that is, to different versions of contents of consciousness. It may seem that reports of our experiences must be the last word on this matter, much as publishing a text fixes what words do and don’t occur in it. This idea is especially inviting if we take heterophenomenological reports to be the best evidence about people’s mental lives. But even publication fixes a text only relative to a social context, and only for a while; post-publication revision can and does occur. Similarly with reports of conscious experiences. We sometimes withdraw earlier remarks about our experiences, replacing them with claims we take to be more accurate.

It seems clear that temporal anomalies such as color phi demand some explanation along these lines. The initial stimulus sets processes in motion that, if uninterrupted, give rise to a conscious sensation of a stationary red flash. But if certain stimuli follow after a suitable temporal interval, those processes are side tracked, and lead instead to the conscious experience of a moving spot that changes color.

When color phi occurs, there is, from a first-person viewpoint, no conscious experience of a stationary red flash. Plainly this is in some way due to the subsequent stimulus. But it seems that we can imagine two distinct types of mechanism by means of which the subsequent stimulus could have
that effect. Perhaps the subsequent stimulus derails the process leading to a conscious sensation of a stationary red flash, so that no such conscious sensation ever occurs. But there’s another sequence of events that would make it seem from a first-person point of view that no such conscious sensation had occurred. Perhaps the initial stimulus does reach consciousness, so that a conscious sensation of the initial flash does occur, but that conscious sensation does not last long enough to have any noticeable mental effects; it commands no attention, and when it ceases any traces in memory are immediately expunged. The first mechanism, in which the stimulus is edited out before any conscious sensation occurs, Dennett calls Stalinesque; the second sequence, in which the stimulus reaches consciousness but is immediately edited out of memory, he calls Orwellian.

It’s here that FPO leads to radical results. According to FPO, there is no distinction between the reality of a conscious experience and how that experience appears. The reality of conscious states consists simply in how they seem, from a first-person point of view. Accordingly, Dennett, 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).

If it seems from a first-person point of view that there has been no conscious sensation of a stationary red flash, then one has not occurred. But if the reality of conscious experiences consists in how they appear from a first-person point of view, there can be no difference between Stalinesque and Orwellian mechanisms. In color phi, the second stimulus prevents the content of a stationary red flash from occurring within the subject’s first-person viewpoint. That stimulus is, at some point or other, edited out. But if “‘writing it down’ in memory [is] criterial for consciousness,” editing a stimulus out of memory will, on criterial grounds, be indistinguishable from editing it out before it reaches consciousness. There will simply be no difference between the two. Accordingly Dennett denies that there’s any difference between the Stalinesque and the Orwellian models. The difference can only be verbal, a difference between two equivalent ways of describing the same thing. When no early reaction is elicited from the subject, there is simply no fact of the matter about whether the initial stimulus ever becomes conscious. So “there are no fixed facts about the stream of consciousness independent of particular probes.”

The temporal anomalies by themselves, however, do not imply these conclusions. Explaining the anomalies does not require us to deny that there’s a real difference between Stalinesque and Orwellian mechanisms nor, more generally, to adopt FPO. We can do justice to the phenomena by a more modest explanation along the lines of the MDM sketched above, on which subsequent stimuli interrupt the processes normally initiated by the original
stimulus. Moreover, this interruption can be Stalinesque, preventing the occurrence of any conscious sensation corresponding to the initial stimulus, or Orwellian, in which case it will cut short the conscious sensation that occurs and remove any trace of it from memory.

In the end, the two mechanisms yield the same subjective appearances. But this doesn’t mean that we cannot determine which is operative, since the order of events they posit is different. Either the stimulus is edited out before a corresponding conscious experience occurs or it isn’t. And on the MDM, construed narrowly without FPO, there is an objective temporal order in which the various events occur. So the MDM thus narrowly construed suggests not only that the initial stimulus is edited out but also that there are two distinguishable ways in which that might happen, either before or after the stimulus leads to the corresponding conscious sensation.

According to FPO, however, there are no facts of the matter about consciousness beyond those which make up one’s first-person, subjective point of view. If a subject cannot distinguish two situations introspectively, then with respect to consciousness there’s no difference between them.

And it may seem that the constraints of FPO are not unreasonable. By hypothesis, Stalinesque and Orwellian mechanisms do not differ in their introspectible results. Moreover, which mechanism is operative makes no difference in subjects’ verbal behavior, including their heterophenomenological reports. Nor can nonverbal behavior help distinguish the two. As Dennett notes, when mental representations result in nonverbal behavior, the very same behavior may occur whether the representation occurs consciously or not.14 So, for example, neither verbal nor nonverbal behavior can determine whether or not the mental representation of a stationary red flash in color phi is conscious.

It is arguable that common-sense folk psychology distinguishes conscious mental phenomena only by way of their introspectible differences, and by the verbal and nonverbal behavior they result in. But even if that is so, theories often considerably expand our ability to discriminate among phenomena that are indistinguishable independent of theory. Suppose that a particular theory about consciousness always draws the right distinctions in problematic cases about whether particular mental states are conscious or not conscious. The theory agrees, that is, with our folk-psychological convictions. We could then apply this theory to the problematic temporal anomalies to determine, regardless of how it seems to the subject, whether a particular stimulus does or does not make it to consciousness.

We can be more specific about how a theory of consciousness can help in this way. Intuitively, it’s a distinguishing mark of conscious states that whenever a mental state is conscious, we are in some way conscious of that state.15 To avoid confusion, I’ll refer to our being conscious of something, whether a mental state or anything else, as transitive consciousness. And I’ll
call the property mental states have of being conscious state consciousness. A state’s being conscious does not, of course, require that we’re attentively or introspectively conscious of it. We don’t introspect or pay attention to most of our conscious states. Indeed, we forget the overwhelming majority of those states moments after they occur. Still, if one is in no way transitively conscious of a particular mental state, that state is not a conscious state.16

Any theory of what it is for mental states to be conscious must explain in what way we are transitively conscious of our conscious states. There must be some event that constitutes one’s being transitively conscious of any conscious mental state, and a theory of consciousness must tell us what event that is. Such a theory would therefore be able to tell us, when a particular temporal anomaly occurs, whether the mechanism responsible for it is Stalinesque or Orwellian, since Orwellian mechanisms involve an event of transitive consciousness that on the Stalinesque model simply doesn’t occur.

In advance of a reasonably well-confirmed theory of this sort, we have of course no way of telling which model explains color phi or the other temporal anomalies. Perhaps some anomalies are Orwellian and others Stalinesque; perhaps some have both Stalinesque and Orwellian instances. Because we cannot distinguish Stalinesque from Orwellian cases by appeal to introspection, speech, and nonverbal behavior, and we now have no suitable general theory, current experimental paradigms reflect a provisional methodological acceptance of FPO. We operate as though FPO were the case. But this methodological operationalism would be unnecessarily restrictive once we had such a theory. We would then be able to frame more fine-grained experiments based on knowing which mechanism is operative in each kind of case.

On the Orwellian model, the initial stimulus in color phi reaches consciousness but leaves no further mental traces; a conscious sensation occurs but makes no mental difference to the subject. This may seem to conflict with the common-sense observation that no mental state is conscious unless one is transitively conscious of it, even if in the most casual and inattentive way. The Orwellian model claims that sensations become conscious even though it never seems to the subject as though those sensations occur. And if the sensation doesn’t, from a first-person point of view, seem to occur, how can the subject be transitively conscious of it? It seems that such a sensation could be conscious in name only, that is, in some technical sense that fails to make contact with our intuitive conception of consciousness.17

The Orwellian model may therefore seem to be an artificial contrivance, somewhat like Descartes’s unflinching insistence that “we do not have any thoughts in sleep without being conscious of them at the moment they occur; though commonly we forget them immediately.”18 Could either claim be more than a mere verbal conceit? It doesn’t help here simply to note that theories often go beyond common sense. What could we mean by a mental
state's being conscious if it could be conscious without one's being in any way whatever transitively conscious of it?

But the Orwellian model does not, in fact, conflict with our common-sense conceptions. Consider the fleeting auditory and visual sensations that occupy the periphery of our consciousness. These sensations are seldom if ever introspectively conscious. But they also do not occur outside our conscious field of vision. They are conscious sensations, though we take note neither of them nor of our being conscious of them. These sensations are so transitory, moreover, that we ordinarily have no memory at all of what conscious contents occupied the peripheries of our perceptual fields, even a moment earlier. Neither the sensations nor our transitive consciousness of them typically leave any trace in memory.

This account of things accords with common sense, which countenances a wide if somewhat indeterminate area for our conscious visual field. Because we have a strong conviction about roughly how far the field extends, and that it's visual through and through, we feel convinced that many sensations near the periphery of that field are conscious. That is so despite our inability to say what sensations occur near the periphery of that field.

The Orwellian model posits states with roughly this status. The model maintains that stimuli reach consciousness but remain there so briefly that, from a first-person point of view, it doesn't seem that any such conscious sensations occur. All that's necessary for this to happen is a momentary event of being transitively conscious of the sensation, albeit too briefly to register as part of the subject's first-person point of view. Presumably this happens all the time with fleeting peripheral sensations.

Because they leave no trace in memory, it can be argued that we have no reason, from a first-person point of view, to think such conscious sensations exist at all. Of course, when we shift our attention to them, we are subjectively certain that they exist; but those are the cases that do leave traces in memory. Here, again, the appeal to theory is irrelevant, since what's at issue is our common-sense, folk-psychological view of these things.

Is it possible for a state to occur consciously even though to the subject it doesn't seem to occur? There is compelling reason to think so. We're often aware of things that don't make it into our first-person point of view. We are conscious in daily life of endless details of which we don't seem to ourselves to be conscious. The same is true with mental states. Like any other mental construct, our first-person view of ourselves leaves out much detail, enabling us to concentrate on the big picture. So there's no reason to think that every mental state we're conscious of occurs as part of our subjective view of ourselves. Still, when we're conscious of states that don't figure in that first-person picture of ourselves, those states are conscious states. The denial that the reality of conscious sensations can differ from their appearance, though central to FPO, is not part of our folk-psychological picture.
III. TRANSITIVE CONSCIOUSNESS
AND THE TWO MODELS

Because a theory of consciousness must explain what events occur when we are transitively conscious of our conscious states, such a theory will help distinguish Stalinesque from Orwellian mechanisms. So explaining the temporal anomalies does not require that we adopt FPO. Nor must we deny that for each occurrence of a temporal anomaly there's a fact of the matter about whether a Stalinesque or an Orwellian mechanism is operative.

Dennett draws the contrast between Stalinesque and Orwellian mechanisms in terms of the temporal order of events: Does the editing out come before the initial stimulus reaches consciousness or after? I've argued in the previous section that there's no reason to doubt that a suitable theory can answer that question. Even so, the contrast between Stalinesque and Orwellian cases turns out to be somewhat more complicated than that. In this section, I'll argue that there's a possible mechanism for the temporal anomalies that resists ready classification either as Stalinesque or as Orwellian. What's crucial is recognizing the difference between our sensations and our transitive consciousness of those sensations. Once that distinction is clearly in place, we'll see that a firm distinction between Stalinesque and Orwellian models cannot, in general, be sustained. This will, in effect, vindicate Dennett's rejection of that distinction, though not quite for the reasons he put forth.

There is compelling reason to hold that our transitive consciousness of sensations is something distinct from the sensations themselves, even when the sensations are conscious sensations. For one thing, not all sensations are conscious. In peripheral vision and subliminal perception, and in some dissociative phenomena such as blindsight, sensations occur without our being in any way transitively conscious of them. The sensations that occur in these processes are not conscious. It's natural to conclude that sensations are distinct from our transitive consciousness of them, which occurs only when our sensations are conscious. Considered apart from our transitive consciousness of them, the sensations by themselves are not conscious states; only the two together—sensation plus one's transitive consciousness of it—constitute a conscious state.

The heterophenomenological method may make it seem as though all mental states are conscious. Whatever may be so when mental states go unreported, the states we do report are always conscious states. Since heterophenomenological reports provide the best evidence about those states, it may be tempting to conclude that no mental states could fail to be conscious. No evidence other than heterophenomenological evidence is, one may think, nearly strong enough to justify the existence of mental states that aren't conscious.
But this is too quick. Although heterophenomenological reports are our best evidence about mental states, they are not our only evidence. Nor does heterophenomenological evidence always trump other considerations. For example, other evidence can override a person’s heterophenomenological denials that that person is in a particular mental state. The mental states thus established would not be conscious states; heterophenomenological denials show that the person is in no way transitively conscious of those states. And, since mental states needn’t be conscious, we have reason in the conscious cases to distinguish the states from our transitive consciousness of them.

There are, in addition, theoretical reasons to distinguish sensations from our transitive consciousness of them. We distinguish sensations by reference to their sensory content. A sensation may, for example, be a sensation of a stationary red flash, whereas another is a sensation of a moving spot that changes color. Each such sensation may be conscious or not conscious.

When a sensation of a stationary red flash is conscious, one is transitively conscious of that sensation. But even when one is conscious of the sensation in the way required for it to be conscious, one’s consciousness of the sensation can be more or less detailed and can represent the sensation in different ways. The way one’s transitive consciousness of the sensation represents it, moreover, determines how it appears to one from a first-person point of view; it determines, that is, what it is like to have the sensation.

Consider, for example, the game Dennett describes of “hide the thimble,”23 in which people may look straight at the thimble they’re trying to find and yet fail to register it consciously. Dennett uses this phenomenon to illustrate that it is not always clear, even from a first-person point of view, whether one is conscious of some particular thing.24 Cases of this kind plainly occur, but they seem to cause difficulties. Whatever is true about the periphery of one’s visual field, the sensory states central to that field are normally conscious. So if one is looking straight 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 lend plausibility to Dennett’s claim that, independent of particular probes, there isn’t any fact of the matter about what conscious experiences we have.

Distinguishing our sensations from our transitive consciousness of them helps explain this kind of case. Mental states often have more detailed content than we’re transitively conscious of. This is true even of the visual sensations that occur at the center of our visual field; we’re seldom if ever aware of all the sensory content such sensations contain, as casual shifts of focus reveal. This should come as no surprise. In general, being transitively conscious of something doesn’t mean being transitively conscious of every aspect of the thing. We would need some special reason to make an exception of mental states and count consciousness as transparently revealing every aspect of their nature.
When I look straight at the thimble, I may well be conscious of the sensations at the center of my visual field, even though I’m not conscious of seeing a thimble. How can we explain this? It’s a mistake to suppose that the sensory content of these central sensations includes no representation of the thimble. It can happen that, even though one doesn’t consciously see an object, one later recalls just where it was and what it looked like. This is strong evidence that the content of our earlier visual sensations contained a representation of the object. That aspect of our sensations wasn’t conscious, but the content was nonetheless there.

But if one is conscious of sensations whose content includes a representation of the thimble, why doesn’t one consciously see the thimble? The only explanation is that, although one is conscious of those sensations, one is not conscious of their content as representing a thimble. One is transitively conscious of the sensations in a way that leaves out that aspect of their content.

This sort of thing happens in many other cases as well. Consider the process by which we acquire the ability to recognize different wines or to pick out the various instruments playing in an orchestra. Normally, the two kinds of sensation are conscious even before one can tell consciously the difference between them—that is, even before they’re distinguishable from a first-person point of view. It’s just that the two types of sensation don’t yet differ consciously. How can a sensation of an oboe and another of a clarinet both be conscious without differing consciously?

Even before one acquires these discriminative abilities, one’s sensory contents must reflect the qualitative differences one is trying to learn; otherwise, one could never learn those differences. So even before one can distinguish an oboe from a clarinet, one’s auditory sensations of the two instruments must differ. There will be some aspect of the sensory content of the two types of sensation that differs, even though that aspect doesn’t register consciously. The two kinds of sensation are typically conscious sensations before one can consciously tell the difference between them—that is, before they’re distinguishable from a first-person point of view. Even before one learns to discriminate the two sensations, one is conscious of them, though not in respect of the relevant qualitative differences. Only afterwards does one become conscious of them in respect of those differences. Again, we have reason to hold that sensations can be conscious even when one isn’t transitively conscious of every aspect of their content.

Since we can be transitively conscious of our conscious sensations in different ways, there are two levels at which we must distinguish content. The sensations of which we’re transitively conscious have sensory content of one sort or another, depending largely on the nature of the relevant stimuli. But even holding the sensory content of a sensation constant, we must distinguish the different contents that our transitive consciousness of those
sensations can have. In the situation just considered, our transitive consciousness may represent a particular auditory sensation either as an indiscriminate woodwind sensation or in a more refined way, say, as a sensation of an oboe. In these two cases, the content of our transitive consciousness of the sensation will differ accordingly, even though the sensory content of the sensation remains unchanged. Moreover, one's first-person point of view is a function of the way one is conscious of one's sensations and other mental states. So it is the content of one's transitive consciousness of the sensation that determines how things are from a first-person point of view.

These considerations suggest a mechanism for the temporal anomalies that Dennett doesn't consider. Suppose, in color phi, that the initial stationary red flash produces in the subject a sensation of that flash. The sensory content of that sensation is of a stationary red flash. But the subject need not be transitively conscious of the sensation in that way. When distinct stimuli follow rapidly one upon another, one often isn't conscious of much detail in the resulting sensations.

Suppose, now, that the subject in color phi is transitively conscious of the initial sensation only as a sensation of a flash and not as something stationary nor even red. After the green stimulus causes a second sensation, then the subject becomes conscious of both sensations, but still not as sensations of stationary flashes. Rather, the subject becomes transitively conscious of the two sensations together, as though fused into a single sensation of a moving spot that changes color. The content of the subject's transitive consciousness of the two sensations is that there's a single moving sensation that changes color. Strictly speaking, there is no editing here, since there's no revising of the content of the sensory states nor of the content of the subject's transitive consciousness of those states. Rather, the subjective appearance results simply from the way one comes to be transitively conscious of those states.

This mechanism resists easy classification as Stalinesque or Orwellian. On the Stalinesque model, the initial stimulus of a stationary red flash never makes it to consciousness. That's what happens in this case, since the subject never becomes transitively conscious of the initial sensation by itself, nor in respect of its sensory contents of color or motion. On the Stalinesque model, editing occurs prior to consciousness. And though strictly speaking there's no editing—that is, no revising—something like editing does occur before the initial sensation makes it to consciousness. For when the subject's transitive consciousness of that initial sensation does occur, it misrepresents the sensory content of that sensation. The transitive consciousness edits the sensations in the attenuated sense that it misrepresents them.

But this mechanism counts equally well as Orwellian. The Orwellian model stipulates that the subject becomes conscious of the sensation that results from the initial stimulus before any editing occurs. That's what
happens in this case; the subject begins by being conscious of the sensation as a flash, albeit one that's indeterminate with respect to color and motion. When the second stimulus is received, the subject becomes transitively conscious of both sensations in a way that more fully reflects the sensory contents of color and motion, as a sensation of a moving spot that changes color. But on the Orwellian model, editing does occur after the first stimulus reaches consciousness. And in the mechanism under consideration, there's editing of a sort after the initial consciousness of the first sensation, since the subject's transitive consciousness changes from that of a flash with no color or motion represented to that of the spot that changes color and position.

When Dennett describes the Stalinesque and Orwellian models, he sometimes seems to allow for sensory contents' being distinct from one's transitive consciousness of those contents. But Dennett holds that we need not regard the sensory contents we're transitively conscious of as distinct existences. They can, instead, be merely notional objects of the relevant transitive consciousness. What matters for consciousness is how one is transitively conscious of sensory contents, not whether distinct sensory contents exist. And if it could be that no sensory contents exist distinct from our transitive consciousness of them, we must avoid the idea that our transitive consciousness of sensory content can vary independently of the sensory content itself. For this reason, the mechanism under consideration doesn't fit comfortably with Dennett's discussion.

But if, as I've argued, sensory states can occur without being conscious states, and hence independently of one's being transitively conscious of them, the third mechanism is at least a theoretical possibility. And, because that mechanism conforms to some extent to both the Stalinesque and Orwellian models, and to neither better than the other, it is arbitrary to describe this mechanism as exemplifying either model more than the other. So the third mechanism blurs the contrast between Stalinesque and Orwellian models. Even if we suppose the objective temporal order of events fixed, it's wholly arbitrary whether to regard this mechanism as Stalinesque or Orwellian. This gives us reason, albeit different from Dennett's, to reject a firm distinction between Stalinesque and Orwellian models.

IV. TRANSITIVE CONSCIOUSNESS AND FIRST-PERSON OPERATIONALISM

The foregoing argument shows that there are mechanisms it is arbitrary to count as Stalinesque or Orwellian, even when the order of events is known. But the argument does not appeal to any indeterminacy about the order of
events; rather, it relies solely on the different ways we may be transitively conscious of our sensations. So the argument gives us no reason to conclude that there’s no fact of the matter about the temporal order of those events. And if the argument of section I is sound, we can expect suitable theoretical developments to pin down any lack of clarity about the order of events. So even though the argument undermines a firm distinction between Stalinesque and Orwellian models, it does not support FPO.

Suppose I have a sensation of red and I’m transitively conscious of that sensation. The content of the sensation determines the sensation’s character, whereas my transitive consciousness of it is responsible for there being something it’s like from a first-person point of view to have that sensation. The distinction between a sensation and one’s being transitively conscious of that sensation warrants a distinction between how conscious sensations appear and the way they really are.

Dennett rejects this conclusion. Seeing things this way, he urges, “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.” Indeed, the main appeal of FPO is, he urges, that it blocks that “bizarre” consequence. Conscious experiences are a matter of things’ appearing in certain ways. And according to FPO, there is no more to the reality of consciousness than the appearances our experiences present from a first-person point of view.

Dennett holds, moreover, that distinguishing between the appearance of states with content and their reality is of a piece with the Cartesian Theater model. On that model, when “vehicles of content . . . ‘arrive at’ the theater of consciousness, . . . [they] ‘become’ conscious.” A mental state’s being conscious consists in its being observed in the theater of consciousness. The Cartesian Theater model must therefore distinguish between the reality of a mental state and how it appears. Its appearance is a function of how it’s observed, whereas its reality consists in its nature independently of any such observation.

Perhaps adopting the Cartesian Theater model does commit one to distinguishing between the appearance and reality of mental states; but the converse does not hold. For one thing, the Cartesian Theater essentially involves the idea that a state’s being conscious is a matter of its being located at that single place in the brain. But there can be a difference between the appearance of conscious states and their reality even if no unique location in the brain is involved.

Such unique location to one side, Dennett stigmatizes the picture he rejects as involving the notion of something’s seeming to seem a certain way. But whatever initial air of oddity there is to this idea, there is good reason to sustain the distinction between how things seem and how they seem to seem. The content of one’s sensory states defines how things seem to one, even when
the sensory states aren’t conscious. Even when they aren’t conscious, sensory states have various connections with other aspects of one’s mental life, both conscious and not. We sometimes see things without being conscious that we do, and our seeing things in these cases often affects us mentally. For example, seeing a truck by peripheral vision may cause one to feel startled and swerve one’s car, even when one is in no way transitively conscious of seeing the truck. Such things contribute to the way things seem to one, even when one is in no way transitively conscious of their seeming that way.

When one is not at all transitively conscious of being in some sensory state, however, it will not, from a first-person point of view, seem that one is in it. That’s where the second level of seeming comes in. The unconscious seeing of the truck is the first level of seeming; if one saw it consciously, that would be a second level of seeming. So, when the seeing isn’t conscious, it’s natural to say that things seem to us a certain way, but without seeming to seem that way. Saying this is just a way of describing the distinction between how mental states really are and how those states seem, from a first-person point of view.

These considerations allow also for a distinction between the appearance and reality of mental states that are conscious. Suppose I see the truck consciously and then, reflecting on my close call, I attend to my conscious experience of seeing the truck. I am now introspectively conscious of seeing the truck; that is, I am conscious of my sensation and conscious also that I am conscious of that sensation. This higher-order transitive consciousness defines how my conscious sensation appears to me.

Dennett might deny that this sort of thing establishes a full-fledged distinction between the appearance and reality of mental states. Rather, it shows only that the way things seem to us is sometimes conscious and sometimes not. To distinguish between the reality of mental states and their appearance, we need cases in which their appearance and reality diverge. As long as the content of my sensation determines how I am transitively conscious of it, the reality of the sensation determines its appearance. And if that always happens, it’s arguably idle to distinguish two levels of seeming.

But that does not always happen. Consider again the woodwind and thimble examples. The best explanation of those cases is that the sensory content a sensation has does not fully determine how one is transitively conscious of that content. There are, for example, two ways one might be transitively conscious of one’s sensation of an oboe. One might be conscious of it indiscriminately, as a sensation of some woodwind or other. Or one might be conscious of it specifically as a sensation of an oboe. The sensory content of the sensation does not fix how one is conscious of that content, and so the two can diverge. There is sometimes a difference between a sensation’s sensory content and the way one is conscious of that content—between how the oboe seems and how it seems to seem.
These cases show that our transitive consciousness of a sensory content may fail to capture everything about that content. Can we also be transitively conscious of a sensory content in a way that actually misrepresents the content? Can we in effect be mistaken about what mental states we’re in? The woodwind case does not clearly involve such misrepresentation. Being conscious of an oboe sensation as a sensation of an indiscriminate woodwind is not an error in that case, especially since it only happens before one has learned to discriminate the various woodwinds. Nor does outright error occur in the thimble case; rather, the way we are transitively conscious of the sensations central to our visual field simply leaves out an important detail of their content. A clear case of error would have to involve our transitive consciousness representing the sensations as having some content they don’t have, and this doesn’t happen in the woodwind and thimble examples.

Dennett seems tempted to adopt the traditional view that we cannot be wrong about our mental states. If we don’t, he thinks, “we lose the subjective intimacy or incorrigibility that is supposedly the hallmark of consciousness.” Such incorrigibility seems also to be connected with Dennett’s use of his heterophenomenological method. The only neutral method for studying consciousness scientifically relies on heterophenomenological reports. And if such reports were definitive about the mental data of investigation, perhaps scientific results could never show such reports to be mistaken. Perhaps, as Richard Rorty has argued, any reason for thinking that some such report is untrue would equally be a reason to think we had misconstrued the reporter’s words.

But adopting the heterophenomenological method does not commit us to rejecting the possibility of real error about one’s mental states. We might well have sufficient success in pinning down the use of the words used generally in somebody’s heterophenomenological reports that we could simply rule out certain misuses of language. We could then conclude that particular reports were untrue because they expressed mistaken judgments.

It’s natural in any case to assume that error about our mental states does occur, and indeed that it is not all that rare. The thimble and woodwind cases show that when mental states are conscious, there can be features of those states that we could be conscious of but aren’t, and it’s plausible that this occurs frequently. And if that happens reasonably often, why shouldn’t the way we’re conscious of mental states sometimes represent them as having features they do not actually have?

One might object that distinguishing the appearance of a sensation from its reality commits us to a hierarchy of such distinctions. The reality of a sensation is independent of its being conscious, whereas its appearance is due to the way we’re transitively conscious of it. And we can even be transitively conscious of our being transitively conscious of the sensation, as we are when we introspect. Here we distinguish the true nature of the
appearance of the sensation from how that appearance seems to us to be. But why would things stop here? If we take these first steps, won’t we risk an endless hierarchy of appearance-reality distinctions?

One might raise this worry in connection with the higher-order-thought hypothesis about consciousness that I’ve developed elsewhere. On that hypothesis, our being transitively conscious of our conscious mental states consists in our having occurrent thoughts to the effect that we are in those mental states. So a mental state is conscious just in case it is the intentional object of a roughly contemporaneous thought—what I call “a higher-order thought.” This higher-order thought must, I argue, have an assertoric mental attitude, and it may not be the result of any inference of which we are transitively conscious. This last requirement, that higher-order thoughts be independent of any conscious inference, is meant to ensure that our transitive consciousness of the mental states the higher-order thoughts are about will, from a first-person point of view, seem immediate.

This concern need not, of course, be tied to the higher-order-thought hypothesis, but can be raised independently of the way any particular theory accounts for our transitive consciousness of our conscious states. The only way to avoid an endless hierarchy of distinctions between appearance and reality, on this worry, is to collapse the initial distinction between mental states and our transitive consciousness of them.

As already noted, introspective consciousness involves two levels of being transitively conscious of our mental states. We’re transitively of the state, and also transitively conscious that we are transitively conscious of it. For a distinction between appearance and reality to apply at this second level, error must again be possible at that level. And indeed introspection is often unreliable. The failure of introspectionist psychology was due less to theoretical objections than to the conflicting results that continually issued from introspectionist experiments. It’s also plain from everyday experience that expectations and preconceptions distort our introspective awareness of our mental states, sometimes to the point of error.

Does our ability to distinguish between mental states and being transitively conscious of them at these two levels imply an endless proliferation of levels at which we might draw that distinction? In principle yes, but not in practice. We can of course conceive of higher applications. But it’s pretty clear that there’s no empirical warrant for drawing that distinction at higher levels, at least in the mental life of our species. After all, it is relatively seldom that being introspectively conscious of one’s mental states plays any useful role; it’s far less likely that an even higher level of transitive consciousness would play any role distinct from that of introspective consciousness itself. So it’s natural to suppose that such higher levels seldom if ever occur.
V. FACTS OF THE MATTER AND MENTAL TAXONOMY

The argument for distinguishing mental states from our transitive consciousness of them relies mainly on noting that mental states are not always conscious, and that no state of which we are not at all transitively conscious will count as a conscious state. Since mental states occur both when we’re transitively conscious of them and when we aren’t, events of transitive consciousness are distinct from the mental states we’re transitively conscious of.

We can reinforce the idea that mental states are distinct from the events of transitive consciousness in virtue of which those states are conscious by appealing to the content of these states. Every mental state has some distinguishing content. But the content of one’s being transitively conscious of a state perforce differs from the content of that state. Suppose I think it’s raining; the content of that thought is simply that it’s raining. If my thought is conscious, I am transitively conscious of it; so the content of that transitive consciousness will be that I have the thought that it’s raining. Similarly for other cases.

It is occasionally argued that we should not individuate mental states by way of their content. After all, the thought that it’s raining and the thought that I think it’s raining seem to amount to much the same thing. That is because they are the same in respect of the mental analogue of conditions of assertibility. Any conditions in which it’s appropriate to have the thought that it’s raining are also conditions in which it’s appropriate to think that I have that thought. And if we individuate thoughts not with respect to their content but by way of the mental analogue of their conditions of assertibility, one’s conscious thought that it’s raining will not be distinct from one’s transitive consciousness of the thought. On this picture, such transitive consciousness turns out to be internal, somehow, to the thought.

Dennett, also, resists individuating mental states the way folk psychology does, by way of content, but for different reasons. Individuating mental states that way, he notes, results in our distinguishing mental states from our transitive consciousness of them, and hence in a potential hierarchy of levels of such transitive consciousness of mental states. “[W]e end up having to postulate differences that are systematically undiscoverable by any means, from the inside or the outside,” distinctions that are “systematically indiscernible in nature.”

Dennett emphatically does not, however, propose to individuate mental states by the mental analogue of their performance conditions. Rather, he urges that

\[\text{[w]}\text{e replace the division into discrete contentful states—beliefs, meta-beliefs, and so forth—with a process that serves, over time, to ensure a good fit between an entity’s internal information—}\]
bearing events and the entity's capacity to express (some of) the information in those events in speech.

Describing things in terms of such processes is doubtless the right way to capture what happens at a subpersonal level of analysis. The subpersonal brain events that subserve our conscious mental lives probably are not organized in any way we could predict by relying on our folk-psychological taxonomy of ordinary mental states, whether conscious or not. Still, even if that is right at the subpersonal level, it does not follow that there is no level of description at which we should taxonomize things in terms of the folk-psychological notion of mental content.

Dennett himself occasionally seems to be committed to describing things in terms of such content. He describes the Stalinesque and Orwellian models, for example, in terms of when a stimulus reaches consciousness. And the idea of a stimulus's reaching consciousness presumably means that it is the content the stimulus produces that becomes conscious. This appeal to content is hardly decisive, however, since Dennett rejects the Stalinesque and Orwellian models; so his descriptions of them may well invoke notions he also rejects.

There is a somewhat stronger reason, however, to think Dennett is committed to some notion of mental content, and thus to a distinction between mental states and our transitive consciousness of them. Throughout *Consciousness Explained*, Dennett speaks of a content's being present in the brain even when it isn't conscious, that is, even when we're not conscious of it. Typically he does not use our ordinary folk-psychological terminology for these purposes. Rather, he talks of such things as "events of content-fixation," "information-bearing events," "content-discriminations," and "vehicles of content." These phrases, moreover, evidently refer to the occurrence of content of which we need not be transitively conscious. The "onsets [of content-fixations in the brain] do not mark the onset of consciousness of their content."

These events of content fixation must, according to Dennett, differ in various respects from mental states as conceived of by folk psychology. For one thing, "content-fixations ... are [each] precisely locatable in both space and time"; by contrast, Dennett argues that we cannot locate conscious mental phenomena precisely in time. But conscious states are states conceived of in folk-psychological terms. So Dennett must hold that events of content fixation are not the sorts of event that could be conscious.

Events of content fixation are the brain events responsible for conscious mental phenomena; they are subpersonal events that subserve mental phenomena as folk psychology taxonomizes them. Nonetheless, they carry content in some way or other. Can we say anything more about exactly what kind of events they are and how they relate to mental phenomena, folk-psychologically taxonomized?
Here is one hypothesis. In the early stages of visual processing, the properties of color, form, orientation, and motion are represented in the brain independently of one another. There is no special problem, moreover, about precisely locating such representations, either spatially or temporally. These independent representations of color, shape, and so forth do not occur consciously. Except perhaps for pathological cases, we never are visually aware of color or motion without shape, shape or orientation without color, and so forth. There is reason to think things are similar with other sensory modalities. It is therefore inviting to suppose that Dennett’s events of content fixation may be something like these early representations.

How, then, would such events of content fixation lead to consciousness? In the early stages of vision—so-called early vision—the properties of color, motion, orientation, and shape occur independently of one another. In conscious visual states, however, taxonomized folk psychologically, these properties are combined. So it is tempting to suppose that consciousness may arise somehow in the course of subsequent integrative processes that represent those properties as unified.

This picture fits well, in a number of respects, with Dennett’s MDM. Dennett holds that no individual states occur that literally exhibit the contents of distinct events of content fixation, such as those in early vision. Consider any group of events of content fixation that represent independent visual properties in early vision. At any one time, there may well be several processes that might integrate the members of that group. Each process would yield a kind of draft of the contents of consciousness. As with the revising of a text, the relevant integrative processes would perform an editorial or interpretive role in bringing together the fragmentary representations of properties in early vision. Those editorial processes would serve “to ensure a good fit between an entity’s internal information-bearing events and the entity’s capacity to express (some of) the information in those events in speech.”

According to Dennett, probes at different moments may “precipitat[e] different narratives . . . ; [different] versions of a portion of ‘the stream of consciousness.’” That would occur on the present model. When distinct integrative processes coexist, each may involve a disposition to produce a different narrative about one’s mental life. Editorial processes that exist concurrently might even dispose one toward conflicting narratives. Only when some particular probe intervenes will one integrative process drive out the others, thereby settling, for that moment, the facts of consciousness.

Heterophenomenological reports give us our best evidence about how people’s conscious mental lives appear to them. But things aren’t always as they seem. So Dennett’s methodological appeal to these reports is neutral about whether sincere reports truly describe the conscious events that go into a subject’s first-person viewpoint or simply express the subject’s beliefs about those mental events, events which may be entirely notional.
If there are any states that do conform to the descriptions that occur in these heterophenomenological reports, they are mental states, folk-psychologically conceived. And it is reasonable to follow Dennett in holding that these reports refer to actual events only if such reports are independently corroborated by what we know by objective, third-person means, for example, by what we know about such things as brain events.

Dennett apparently believes that the situation is at best mixed. Brain events exist that can be reasonably regarded as bearing content, but their content will be dramatically unlike the integrated content that folk psychology ascribes to conscious states. Certainly that is so if we understand events of content fixation on the model of the independently occurring properties of early vision, and it seems equally so however we construe events of content fixation. So perhaps what exists is simply the precisely locatable events of content fixation—the representations of early vision and the like—and the editorial processes that integrate those early representations. There is nothing, then, corresponding to the folk-psychologically taxonomized mental states to which our heterophenomenological reports refer.

On the present hypothesis, editorial processes do not integrate content by producing actual states with unified content. Rather, they integrate by referring to each of the relevant component events of content fixation. These events can be located precisely in time; and presumably the same holds of the processes that appear to integrate those disparate events.

Folk psychology assumes that conscious states, individuated by way of their unified content, can also be located precisely in time. If so, we could locate a particular state by reference to the color, shape, location, and motion it represents. But no such unified states occur, on the present model. Indeed, the distinct events of content fixation that represent color, shape, location, and motion may well occur at distinct times. And in any case, we can assume that these independent events of content fixation will all occur earlier than the editorial process that appears to unify them. So there is no unique, privileged moment at which content occurs that represents all these visual properties together.

If this model is correct, our folk-psychological taxonomy of mental states is inaccurate in certain important ways. Folk psychology posits mental states that represent in a unified way the various visual properties that are represented separately in early vision. And folk psychology supposes that these states can be located relatively precisely in time. But on this hypothesis, no such states exist. Representational events occur that we can locate precisely in time, but relative to our folk-psychological taxonomy, those representations are fragmentary. There are, in addition, processes that appear from a first-person point of view to integrate those fragmentary representations. But these processes do not result in the unified states of folk psychology.

This picture is not eliminativist with respect to mental states, taxono-
mized folk psychologically. To be sure, no single state or process corresponds to any conscious mental state, as folk psychology describes things. But between the subpersonal events of content fixation and the processes that appear to integrate them, we can save the subjective appearances that folk psychology describes. Events of content fixation are well-defined and precisely locatable, and the subsequent editorial processes provide the apparent folk-psychological integration of representational content. These editorial processes will, as Dennett suggests, "replace the division into discrete contentful states:" It is just that no single state or process satisfies both functions at once. This integrative model does not deny that conscious mental phenomena exist. But, as with Dennett's view, the model sees as artificial the way we ordinarily carve consciousness and mind into discrete mental states.

VI. INTEGRATIVE PROCESSES AND CONSCIOUSNESS

Events of content fixation do not occur consciously. So consciousness must result from the editorial processes that appear to integrate those events. On the integrative model under consideration, there is nothing else that could give rise to consciousness. But why should integrative processes produce consciousness? Given that events of content fixation are not conscious to begin with, why should integrating them yield conscious results?

This problem is particularly pressing, since integrative processes often fail to produce consciousness. Cognitive theories posit many processes that integrate various representational contents, but such unification typically does not result in states that are conscious. Nor is it intuitively obvious why integration should yield consciousness. Integrative processes can explain why shape and color, for example, are represented together, but not why the resulting unified representation should be conscious. By itself, integration seems unable to explain why, from a first-person viewpoint, we seem to be in conscious states with those combined properties.

A related difficulty affects Dennett's discussion. Dennett maintains that "[t]here is no reality of consciousness independent of the effects of various vehicles of content on subsequent action (and hence, of course, on memory)." Perhaps it is correct that all conscious mental phenomena leave suitable traces on action and elsewhere in our mental lives, particularly in memory. But Dennett seems to hold that leaving such traces is not just necessary for states to be conscious, but sufficient as well. Thus he writes:

Consciousness is cerebral celebrity. . . . Those contents are conscious that perseveres, that monopolize resources long enough to achieve certain typical and "symptomatic" effects—on memory, on the control of behavior and so forth."
The difficulty is that mental states have many effects that are independent of whether those states are conscious or not. As noted earlier, Dennett observes that the very same nonverbal behavior may be caused either by a state that's conscious or a state that is not.52 Since conscious and nonconscious mental states can have the same effects on nonverbal behavior, having such effects will not make a state conscious. Similarly, most of the mental traces left by conscious states could equally well have been left by mental states that are not conscious. In the processes posited by cognitive theories, representational states typically have very wide ranging mental effects even when those states are wholly nonconscious. But the general point is independent of theoretical posits. We often seem to solve difficult problems without consciously thinking about them; in these cases, many nonconscious mental states must have substantial mental effects, which in turn remain nonconscious, before the solution occurs to us consciously. The same holds for the effects mental states have on memory, effects which Dennett counts as criterial for consciousness. Occasionally we recall having seen something, and may even have a visual image of it, though at the earlier time we were not in any way conscious of seeing it. In such cases, perceiving that wasn't conscious has a significant, lasting effect on memory.

So, just as the integration of fragmentary content can occur without resulting in conscious states, so can cerebral celebrity and states' leaving traces in memory. Still, the integrative model seems to have promise. Ordinary conscious states, taxonomized folk psychologically, do represent shape, color, orientation, location, and movement as unified. If no individual brain events represent those distinct properties together, consciousness and unification somehow go hand in hand. Why should this be?

On the integrative model, properties represented separately in early vision in some way come to be represented together as a result of various editorial processes. These processes need not result in unified states with the representational properties of the relevant components; rather, they may simply refer back to each of those components. Integration may be achieved by referring to all the early representations in a unified way.

If integration occurs in this second way, the connection with consciousness is clear. Conscious mental phenomena are mental phenomena of which we're transitively conscious, in a way that from a first-person point of view seems to be immediate. By referring in a unified way to the separate representations of early vision and other such events of content fixation, integrative processes not only unify those representations, but also make us transitively conscious of them. And because these processes refer to events of content fixation as suitably integrated, they make us transitively conscious of them in just that way. To integrate events of content fixation, these editorial processes must involve some unifying intentional reference to the events. And referring to something mentally is having some sort of thought about
VII. CONSCIOUSNESS AND SPEECH

Dennett rejects the folk-psychological taxonomy of "discrete contentful states," in favor of "a process that serves, over time, to ensure a good fit between an entity's internal information-bearing events and the entity's capacity to express (some of) the information in those events in speech." What sort of fit is this? Putting aside Dennett's reasons for rejecting the taxonomy of folk psychology, just what connection does obtain between speech and our "internal information-bearing events"?

On the standard picture, speech acts express intentional states, the content of which matches that of the speech acts. This has important implications when we turn to heterophenomenological reports. Since these reports are about the mental states we take ourselves to be in, they express our transitive consciousness of those mental states. So the content of our heterophenomenological reports is the same as the content of the corresponding events of transitive consciousness.

Dennett rejects this standard picture. Speech acts, he argues, typically do not express intentional content that is already in place; rather, our choice of words often influences the content of our thoughts. On this Pandemonium Model of speech production, as he calls it, the content of our speech acts does not generally match that of some previously existing intentional states. Accordingly, the content of our heterophenomenological reports seldom reflects prior events of being transitively conscious of our mental states. Instead, these reports often, perhaps always, determine the content of whatever events of transitive consciousness may occur.

Dennett's principle argument against the standard picture of the relation between thought and speech is that we often discover what we think only as we say it. But it is likely that when we discover what we think only as we say it, that is not because the thoughts do not exist until we speak, but because often our thoughts are not conscious until we express them verbally. Doubtless, the words we use do sometimes affect the content of our thoughts, perhaps often. But even when that happens, this does not show that our heterophenomenological reports do not express prior events of transitive consciousness, but only that those reports diverge to some extent in content from those prior events of transitive consciousness.

It is likely that we often assign content to our thoughts on the basis of what we say; in effect, we read back onto our thoughts the refined distinctions of content drawn so readily in speech. But whenever one speaks, there must have been some inner state—or more likely, as Dennett urges,
interactions among inner states—that are responsible for one’s using the words one does. And it is reasonable to identify the thoughts one’s words express with whatever states or interactions among states end up producing those words. As Dennett usefully emphasizes, we often learn what thoughts we have not by introspection but by seeing what we say. But that does not mean that there is no thought, folk-psychologically conceived.

Dennett urges that “the second-order state (the better-informed state) comes to be created by the very process of framing the report.”58 It is unlikely that this is always so. Second-order states plainly occur without being verbally expressed, even if their content is sometimes less elaborate than that of verbally expressed second-order thoughts. In any case, heterophenomenological reports do indicate the occurrence of events of transitive consciousness with the same content as the reports, whether or not those events occur prior to the reports. Since a state’s being conscious implies that one is conscious of it, it must be these events of transitive consciousness which are responsible for the consciousness of the mental states they are about.

On the model I have been considering, no individual states occur that literally combine such contents as shape and color; there are only nonconscious, fragmentary events of content fixation and integrative processes that refer to those events. Nonetheless, these processes, and the higher-order thoughts they involve, enable us to explain why we seem to be in mental states as folk psychology taxonomizes them. Those mental states are the intentional objects of our editorial processes or higher-order thoughts. They are the states we represent ourselves as being in, even if it turns out that they are simply notional.

Conscious mental states do exist on this model, but they are not the kinds of states folk psychology takes them to be. They are not conscious cases of states with integrated sensible properties, but arrays of events of content fixation of which we are transitively conscious, though we represent those arrays as though they were single states.

Dennett would have limited sympathy with these conclusions. As noted earlier, his heterophenomenological method is neutral about whether the mental events referred to by subjects’ reports really exist. Still, if those reports are sincere, he maintains, they are constitutive of what it’s like for the subject at that time, and hence constitutive of that subject’s consciousness.59

The Pandemonium Model of speech production seems to support this idea. What it’s like for one hinges on how one is transitively conscious of one’s mental life. So, if sincere heterophenomenological reports fix the content of whatever events of transitive consciousness occur, perhaps those reports are somehow constitutive of what it’s like for the subject.

As already noted, consciousness occurs even in the absence of sincere heterophenomenological reports. So in these cases Dennett might urge that what is constitutive of consciousness is the disposition to report sincerely,
rather than the reports themselves. But whenever we can describe things dispositionally, there is some occurrent state or property that is responsible for the relevant dispositional behavior. Since the disposition here is to make sincere a heterophenomenological report, we can assume that the relevant underlying state is simply the higher-order thought that this report would express, a thought whose content is reasonably close to that of the report.

As we saw in the thimble and woodwind cases, events of being transitively conscious of our mental states can be more or less detailed. Moreover, what determines in these cases how the subject is conscious of the mental state in question is the event of transitive consciousness, not the state or states that event of transitive consciousness is about. So, even when an event of transitive consciousness is erroneous, we can assume that that event fixes what the relevant conscious state is like for the subject. If I have a sensation of an oboe but my transitive consciousness of the sensation represents it as a sensation of a clarinet, it will be just as though I have a conscious sensation of a clarinet.

Suppose, now, that the sensation is absent altogether, but that an event of transitive consciousness still occurs, representing me as having a sensation of a clarinet. Since that event suffices for it to seem to me that I have such a sensation when my sensation is actually of a different sort, that event should yield the same result even if I have no relevantly similar sensation at all. Even when the sensations events of transitive consciousness are about do not exist, those events will determine what it’s like for one.

This lends plausibility to Dennett’s claim that a subject’s sincere heterophenomenological reports are somehow constitutive of what it’s like for that subject. Events of transitive consciousness fix what it is like for the subject. So it is representing oneself as being in various particular mental states that is constitutive of one’s consciousness. Heterophenomenological reports do just that.

The same considerations also lend support to the integrative model. On that model, when it seems to me that I have a conscious sensation, there is no unified sensation as folk psychology conceives of these things. Rather, there are various events of content fixation, together with my being transitively conscious of those events as suitably unified. Once again, it is the event of transitive consciousness, not the states that event pertains to, that determines what it’s like for me.

What consequences does all this have for FPO? Folk psychology assumes we can assign a precise temporal location to conscious states with relatively unified representational contents, for example, visual perceptions that represent color, shape, and motion together. But that presupposes that such unified states actually exist. If what exists, instead, are various events of content fixation, we can expect that in any particular case the relevant representations of shape and color will occur at slightly different moments, and that the integrating process will occur at still another moment.
What corresponds to the conscious states posited by folk psychology, then, is an array of events and processes. So it may seem that to assign any temporal location to a conscious state, folk-psychologically conceived, would require an arbitrary choice among the moments at which the relevant component events and processes occur. And the present picture would vindicate FPO; there would be no fact of the matter about the temporal location of those notional folk-psychological states.

But assigning such temporal location to the conscious states of folk psychology is not, in fact, arbitrary. Each integrating process involves being transitively conscious of the relevant component events of content fixation as a unified whole. This points toward a nonarbitrary way to locate conscious folk-psychological states in time. Since it is the event of transitive consciousness which is responsible for a state’s being conscious and determines what it’s like for the subject, that event is all that matters for temporal location. Events of content fixation will occur earlier than the event of transitive consciousness. But at those earlier moments there is nothing it’s like for the subject, since at those times there are only the various nonconscious mental precursors of the conscious state. So the relevant events of transitive consciousness provide determinate, nonarbitrary facts about the timing of conscious states. Since the foregoing integrative model closely resembles Dennett’s MDM, narrowly construed, and we have reason to reject FPO, it seems possible that we can explain consciousness by a view along the lines of the MDM but which avoids appeal to FPO.61

NOTES

5. Ibid., 144.
6. Ibid., 72ff.
7. Reliance on such reports is of course standard in experimental cognitive psychology.
9. Ibid.
12. Ibid., 132.
13. Ibid., 138; cf. 275.

15. Fred Dretske has contested this, arguing that we often are not transitively conscious of our conscious states ("Conscious Experience," Mind 102 [1993]: 263–83, esp. 272–5, and "Are Experiences Conscious?" ch. 4 of Naturalizing the Mind [Cambridge, Mass.: MIT Press/Bradford Books, forthcoming]). Instead of a mental state’s being conscious if one is transitively conscious of it in some suitable way, Dretske proposes that a state’s being conscious is simply a matter of its being a case of transitive consciousness ("Conscious Experience," 280–1). But all mental states are cases of transitive consciousness. So Dretske’s alternative in effect defines all mental states as conscious states, which is implausible. Dretske’s argument that we aren’t always transitively conscious of our conscious states also fails to take account of the fact that we can be conscious of an experience in one respect while not being conscious of it in another. See David M. Rosenthal, "Explaining Consciousness," ms.

John R. Searle denies that it’s even possible to be conscious of our conscious mental states, though his reasons are different. "[W]here conscious subjectivity is concerned, there is no distinction between the observation and the thing observed" (The Rediscovery of the Mind [Cambridge, Mass.: MIT Press, 1992], 97). The context makes clear that Searle is denying not just that we can observe our conscious states, but that we are transitively conscious of them at all in the way we’re conscious of other things: "We cannot get at the reality of consciousness in the way that, using consciousness, we can get at the reality of other phenomena" (ibid., 96–7). Searle argues for this by appeal to the idea that we can describe consciousness only in terms of what it’s consciousness of (ibid., 96). But even if that’s so, it doesn’t follow that there can’t be states in virtue of which we’re conscious of our conscious states.

Searle also urges that, when we mentally represent things, the things we represent must be something ontologically objective. Since conscious states, according to Searle, are ontologically subjective, we cannot mentally represent them and so cannot be transitively conscious of them (ibid., 99; cf. 87–100, 137ff., and 144ff.). Because it’s difficult to make clear sense of Searle’s distinction between the ontologically subjective and objective and, indeed, just what ontological subjectivity amounts to, it’s unclear how to evaluate this argument.


Descartes’s insistence that “[b]eing conscious of our thoughts at the time when we are thinking is not the same as remembering them afterwards” stands in useful opposition to Dennett’s view that “‘writing it down’ in memory [is] criterial for consciousness.”

19. Contrast this with the situation in which one senses that another person is looking at one. Though it’s plain on reflection that we get this information visually, it doesn’t intuitively seem to be visual information.


24. Recall the challenge to the Orwellian model considered at the end of the previous section.

25. E.g., Consciousness Explained, 124.

26. It was this way of partially undermining the contrast between Stalinesque and Orwellian models that I had in mind in “Time and Consciousness” (The Behavioral and Brain Sciences 15 [1992]: 220–21), though that discussion wasn’t as clear as I would like.

27. Consciousness Explained, 132.

28. Ibid., 144.

29. Ibid., 319.

30. According to Rorty, when there is reason to believe somebody’s report of a mental state is not true, we cannot even in principle distinguish between the person’s having just misspelled words and having actually made a factual error about what kind of state it is. (“Mind-Body Identity, Privacy, and Categories,” The Review of Metaphysics 19 [1965]: 424.) This recalls W. V. Quine’s well-known argument that any translation of a language which represents people as asserting bald contradictions is overwhelming evidence that the translation is wrong. (“Carnap and Logical Truth,” in The Ways of Paradox and Other Essays, rev. and enl. ed. [Cambridge, Mass.: Harvard University Press, 1976], 109.)


Dennett has informed me (personal communication) that he is not in fact concerned about an endless hierarchy, but rather about taking the first step in distinguishing the appearance from the reality of mental states.

32. Dennett raises a distinct worry specifically about the higher-order-thought hypothesis. He assumes that this theory must posit not only a higher-order thought about each conscious mental state but, in addition, a distinct higher-order belief about the state (see Consciousness Explained, 307 and 317). As Dennett notes, beliefs are dispositional states that underlie our thoughts; in effect, they are dispositions to have certain thoughts. But being disposed to have a thought about something doesn’t make one conscious of that thing. So higher-order beliefs will not figure in explaining how we are transitively conscious of our conscious mental states nor, therefore, in explaining what it is for mental states to be conscious.


38. Similarly, though Dennett notes that the editing posited by the Orwellian model allows for a certain kind of error about what mental states we’re in (Consciousness Explained, 318–19), his rejection of Orwellian explanation leaves it open to him to deny the possibility of such error.
39. *Consciousness Explained*, 365, 459, 113, 114, respectively.
40. Ibid., 113; emphasis Dennett’s.
41. Ibid.
44. Though for other reasons, partly indicated in section VII below, Dennett would not accept this picture.
46. Ibid., 319.
47. Ibid., 135.
48. This is suggestive in connection with the temporal anomalies. Thus, the explanation of our seeming to see a moving spot that changes color when the red and green flashes occur suitably separated in space and time might have to do with the timing in the brain of the events of content fixation that independently represent color, motion, and location.
50. Ibid., 132.
52. See *Consciousness Explained*, 124.
53. To the extent that the integrative model under consideration resembles Dennett’s MDM, this conclusion fits with my argument elsewhere that the higher-order-thought hypothesis has all the advantages of the MDM without being committed to FPO. (See David M. Rosenthal, “Multiple Drafts and Higher-Order Thoughts,” *Philosophy and Phenomenological Research* 53 [1993]: 911–18.)
54. *Consciousness Explained*, 319; emphasis Dennett’s.
55. Ibid., 247.
56. Ibid., 315. Dennett considers in this context the higher-order-thought hypothesis, on which these events of transitive consciousness are higher-order thoughts about our mental states. It is arguable that the higher-order-thought hypothesis in effect follows from what I am calling the standard picture of the connection between speech acts and intentional states. See Rosenthal, “Thinking that One Thinks,” passim.
57. See *Consciousness Explained*, 245.
58. Ibid., 315; emphasis Dennett’s. Also: “The emergence of the [verbal] expression is precisely what creates or fixes the content of higher-order thought expressed” (ibid.).
59. I am grateful to Dennett for emphasizing this (personal communication).
60. Compare W. V. Quine’s view that dispositional descriptions can, for theoretical purposes, be replaced by descriptions that mention enduring structural traits (Word and Object [Cambridge, Mass.: MIT Press, 1960], 46).
61. Special thanks to Dan Dennett for exceptionally useful reactions on an earlier draft.