

Content, Interpretation, and Consciousness

I. FIRST-PERSON OPERATIONALISM AND HIGHER-ORDER THOUGHTS

We are all familiar with situations in which memory distorts some current experience. I may see a person I don't know at all, but my memory of an old friend causes me to misperceive that person as my friend; the conscious experience that results is of seeing the friend. Perhaps, for example, my friend wears glasses and the person now before me doesn't. Although I see a person without glasses, my memory of the friend intrudes and I seem, so far as consciousness is concerned, to see a person wearing glasses.

Folk psychology accommodates two distinct explanations of such cases. My memory of the earlier experience might contaminate the current visual information before it even reaches consciousness; if so, I in effect hallucinate the glasses. But perhaps, instead, I begin by consciously seeing the person as having no glasses, but the memory then immediately revises the experience by adding the glasses and also overwrites any current memory of the new visual experience without glasses.

These are the two scenarios that Dan Dennett and Marcel Kinsbourne labeled *Stalinesque* and *Orwellian*, respectively, famously arguing that the distinction between them is spurious. When it comes to consciousness, they urge, we cannot distinguish between appearance and reality. So, if the two scenarios are indistinguishable to consciousness, they are indistinguishable in reality, as well. Since consciousness cannot fix the time of contamination as being before or after the new visual information reaches consciousness, the two hypotheses differ only verbally. Folk psychology, by allowing for distinct scenarios, misleads us into thinking that this kind of case can occur in two different ways.¹

As Dennett puts their view, "there is no reality of consciousness independent of the effects of various vehicles of content on subsequent action (and hence, of course, on memory)" (CE 132). So "there are no fixed facts about the stream of consciousness independent of particular probes" (CE 138; cf. 275). Dennett usefully calls this view *first-person operationalism* (CE 132), since it holds that the facts about consciousness are wholly fixed by the effects consciousness has on other things.

¹ Daniel C. Dennett and Marcel Kinsbourne, "Time and the Observer: The Where and When of Consciousness in the Brain," *The Behavioral and Brain Sciences*, 15, 2 (June 1992): 183–201; Open Peer Commentary, pp. 201–234, and Dennett and Kinsbourne's Author's Response, "Escape from the Cartesian Theater," pp. 234–247; and Dennett, *Consciousness Explained*, Boston: Little, Brown & Co., 1991 (henceforth CE), chs. 5 and 6.

Because theory outstrips observation, a satisfactory theory can often settle questions that won't yield to observation alone. So even if the Stalinesque and Orwellian scenarios are indistinguishable to consciousness itself, perhaps a reasonable theory of consciousness will, in principle at least, show us how to tell which scenario any particular case conforms to.

I have argued elsewhere² that we can do just this on a theory according to which a mental state is conscious just in case it is accompanied by a higher-order thought (HOT) to the effect that one is in that state. We do not regard as conscious any mental state of which we are wholly unaware. So we must in some way be conscious of every conscious state, and having a thought about a state is one way of being conscious of it. Intuitively, it seems that the way we are conscious of our conscious states is direct. We can explain this intuition by hypothesizing that we remain unaware of any inferences or other antecedent factors that might lead to HOTs or explain their occurrence; HOTs seem to arise spontaneously.³ Indeed, HOTs need not themselves be conscious, and typically won't be.⁴

On this theory, a mental state becomes conscious at the onset of the relevant HOT. So whether my unrevised visual sensation reaches consciousness depends solely on whether the contamination occurs before or after the onset of some HOT. Since HOTs are determinate states, their exact moment of occurrence is determinate, whether or not we can discover it in practice.

At the same time, however, the HOT model explains why we should feel a certain reluctance to classify particular cases as being Stalinesque or Orwellian. Suppose a sensation occurs and very quickly becomes conscious. But then, almost immediately, the sensation changes, and a moment after that one becomes conscious of the change. Is this revision Stalinesque or Orwellian? That depends on whether we focus on the sensation in its original or changed form. The case is Stalinesque relative to the changed sensation, since that change occurred before the sensation became conscious in its new form. But relative to its original form the case is Orwellian, since the sensation was conscious in that original form before it changed. The case looks Stalinesque if we regard the revised sensation as a new, distinct state, and Orwellian if we see the revised sensation as just a later stage of the original state.

² "Multiple Drafts and Higher-Order Thoughts," *Philosophy and Phenomenological Research*, LIII, 4 (December 1993): 911–918; "Multiple Drafts and Facts of the Matter," in *Conscious Experience*, ed. Thomas Metzinger, Exeter, UK: Imprint Academic, 1995, pp. 275–290; and "First-Person Operationalism and Mental Taxonomy," ch. 8 in this volume.

³ Though HOTs are noninferential from a first-person point of view, they might still result sometimes from inferences we are unaware of. We need not suppose that the way we are conscious of our conscious states actually is unmediated to explain our intuitive sense that it is.

⁴ They will be conscious only when we are conscious of target states in the deliberate, attentive way we call introspection.

For development of this HOT model, see "A Theory of Consciousness," in *The Nature of Consciousness: Philosophical Debates*, ed. Ned Block, Owen Flanagan, and Güven Güzeldere, Cambridge, Massachusetts: MIT Press, 1997, pp. 729–753; "Thinking that One Thinks," ch. 2 in this volume; "State Consciousness and Transitive Consciousness," *Consciousness and Cognition*, 2, 4 (December 1993): 355–363; and "Two Concepts of Consciousness," ch. 1 in this volume.

But the choice between these two descriptions is in most cases unprincipled, since it will typically rely on artificially precise identity conditions for mental states. Whether such a case is Stalinesque or Orwellian hinges on arbitrary questions about taxonomizing our mental states themselves. Still, this does not show that it's not determinate when states with specific content properties become conscious, nor that the facts about when revision occurs and when states become conscious are exhausted by how things appear to consciousness.

II. THINKING, SPEECH, AND PROBES

"[T]here are," according to Dennett, "no fixed facts about the stream of consciousness independent of particular probes." One of the most intriguing applications of this challenging claim has to do with the connection between thinking and speech.

When we speak, we express the thoughts we have. Our speech acts, moreover, reflect the content of those intentional states. It is natural to hold that this correspondence of content is exact; whenever we say anything, the speaker's meaning of our speech act is the same as the content of the intentional state we express.

We all have experienced how putting our thoughts into words can appear to tighten up those very thoughts. It's usually assumed that this happens because the process of finding suitable words for one's thoughts clarifies the thoughts themselves. On that account, the clarifying speech act does not actually outstrip, in respect of content, the antecedent intentional state it expresses. It's just that fixing on the right words results in a new intentional state, whose content is more fine grained than one's original thought. So goes our ordinary, folk-psychological description of these cases, and so it seems to us from a first-person point of view. We have a robust first-person sense that our speech acts exactly match in content the antecedent intentional states they express. Whether or not the match is exact, it seems that way to us.

We will see below that folk psychology and our first-person sense of things overestimate the exactness of this match of content. But it is worth considering certain ostensible counterexamples to the idea that our first-person impression is always that an exact match obtains. We do sometimes discover as we say something that what we're saying does not really reflect our thoughts after all. We may have changed our mind, or even find that we never actually thought that thing at all, but said it only conversationally or from habit. But when this type of thing happens, our first-person sense is not of really saying the thing in question, but something like "as if" saying. We don't sense ourselves saying the thing with full illocutionary force, but simply producing the relevant utterance. In Wilfrid Sellars' apt metaphor,⁵ the utterance is produced parrottingly, as with mere recitations, from causes that are tangential to what one thinks. We do sense in these cases a divergence between our speech and what we

⁵ First used in a letter to Roderick Chisholm, published in Roderick M. Chisholm and Wilfrid Sellars, "Intentionality and the Mental," *Minnesota Studies in the Philosophy of Science*, II, ed. Herbert Feigl, Michael Scriven, and Grover Maxwell, Minneapolis: University of Minnesota Press, 1958, pp. 507-539, at p. 524.

think, but the speech productions do not seem to us, from a first-person point of view, to be full-fledged illocutionary acts.⁶

As we saw, an exact match in content between full-fledged illocutionary acts and the thoughts they express can accommodate cases in which our speaking seems to clarify our thoughts. But these cases seem to fit equally well with another interpretation, which itself lends support to Dennett's first-person operationalism. Perhaps putting our thoughts into words clarifies our thinking not because it results in our having clearer thoughts, but because speech acts themselves actually fix the content our thoughts have. Speech acts, on this view, are one sort of probe that determines the facts of consciousness. This is the "pandemonium model" of speech production that Dennett develops in chapter 8 of *Consciousness Explained*,⁷ on which there are no determinate intentional states prior to the occurrence of verbal expressions. It is not simply that our choice of words often influences the content of our thoughts.⁸ Rather, many forces occur in the intentional arena and compete for expression. The speech act that ultimately wins out in effect results from the vector product of those forces, rather than from some single, antecedently existing state with the relevant determinate content.

The folk-psychological view that speech acts always mirror the content of the intentional states they express reflects our subjective impression of these situations. It always seems, from a first-person point of view, that what we say exactly matches the content of some intentional state we are in. But Dennett's pandemonium model also does justice to this subjective impression. After all, if speech acts do fix the content of the intentional states they express, we will sense a perfect fit between them.

In addition to squaring with our subjective impressions, these two models both capture important aspects of the connection between thinking and speech. But the aspects they capture are different. One aspect has to do with what it is for a speech act to be meaningful. Two things are needed. One is that the sentence uttered have semantic meaning. But even so, no utterance is meaningful if it's a mere recitation without underlying thought, if, as Sellars put it, it is produced parrotingly. We can, of course, distinguish tolerably well between meaningful and parroting speech production, independent of any appeal to theory. But if we want to explain what that difference consists in, we must appeal to the idea that meaningful speech acts express intentional states, whereas parroting utterances do not.

The folk-psychological picture of the relation between speech and thought reflects this explanation. Nonparroting speech productions express antecedent intentional states, and expressing an intentional state means that the semantic meaning of

⁶ Insincere speech and the speech productions of actors playing a part are also like this. See my "Intentionality," ch. 3 in this volume, §V. Thus J. L. Austin notes that if I insincerely say 'I promise', I don't strictly speaking promise, but only say I do ("Other Minds," in *Philosophical Papers*, 2nd and 3rd edns., Oxford: Oxford University Press, 1970 and 1979, pp. 76-116, at pp. 101-103 [1st edn., pp. 44-84, at pp. 69-71]). Similarly, Frege remarks that "stage assertion is only sham assertion" (Gottlob Frege, "Thoughts," in *Logical Investigations*, tr. P. T. Geach and R. H. Stoothoff, New Haven: Yale University Press, 1977, pp. 1-30, at p. 8).

⁷ Page 240. See also pp. 245 and 247, and ch. 10, §5, esp. 315.

⁸ Which Dennett suggests also happens (CE 247).

the speech act matches to some suitable degree the content of that intentional state. Such an explanation is unavailable on the pandemonium model. If speech acts fix the content of our intentional states, it cannot be that expressing an antecedent intentional state with roughly the same content is what makes a speech act nonparroting.

One might reply that the pandemonium model can, after all, explain the difference between parroting and nonparroting speech productions. Speech is nonparroting if it results from the forces competing in the intentional arena for expression; otherwise it is parroting.⁹ Whether this reply works depends on just what those intentional forces are that the pandemonium model posits. If they are merely subpersonal events that resist folk-psychological taxonomy as full-fledged intentional states, the reply fails, since parroting utterances also result from the interaction of subpersonal events that resist folk-psychological taxonomy.

But perhaps the pandemonium model actually posits full-fledged intentional states operating in the intentional arena. Speech acts, then, simply settle which of these states wins out in the competition for expression, rather than converting subpersonal events of content fixation¹⁰ into genuine intentional states. Then the model can explain speech as nonparroting if it results from such intentional states. But the pandemonium model so construed does not differ relevantly from the folk-psychological picture, which also posits antecedent intentional states, and regards nonparroting speech as that which results from such states. Any satisfactory explanation of how parroting and nonparroting utterances differ must invoke the intentional states posited by the folk-psychological model.

There is good reason, in any case, to construe Dennett's pandemonium model as positing only subpersonal events of content fixation, rather than full-fledged intentional states. Thus he writes:

We 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.¹¹

The forces whose interaction results in speech acts are properly speaking proto-intentional, and precipitate into intentional states only when they issue in speech acts or in reactions to some relevant probe. Moreover, just which speech acts and intentional states these subpersonal forces issue in depends to some extent on factors irrelevant to the intentional content at hand. Events of content fixation determine speech performances more in the manner of instructions in aleatory music than fully developed scores.

⁹ I owe this idea to Tim Kenyon, in conversation.

¹⁰ Dennett speaks throughout *CE* of such nonconscious, subpersonal events that subserve mental states taxonomized folk psychologically. Because they occur nonconsciously, "their onsets do not mark the onset of consciousness of their content" (113; emphasis Dennett's). And he holds that, unlike mental states taxonomized folk psychologically, "content-fixations . . . are [each] precisely locatable in both space and time" (113).

¹¹ *CE* 319, emphasis Dennett's. Cf. "The Message is: There is no Medium," *Philosophy and Phenomenological Research*, LIII, 4 (December 1993): 919–931, pp. 930–1.

Nonetheless, the folk-psychological model arguably exaggerates the match in content between speech act and intentional state, and the pandemonium model provides a useful corrective. If the content of an intentional state corresponds exactly to the semantic meaning of the speech act that expresses it, thinking must itself exhibit a language-like structure; perhaps thinking even takes place in a language of thought whose syntax and semantics echo those of overt natural language.¹² Such a picture, however, arguably underestimates the extent to which what we think and how we think it are affected by the way we express our thoughts in speech. The pandemonium model seeks to capture the effect our putting words to our thoughts has on the thoughts themselves.

Is there a view that preserves the virtues of both models, allowing us to explain the difference between parroting and nonparroting speech without tempting us to adopt an unqualified language of thought? Arguably yes. Perhaps meaningful, nonparroting speech acts always express intentional states with antecedently fixed content, but the speech acts are nonetheless often richer and more fine grained in content than the intentional states expressed. The content of the speech acts rules out more possibilities and invokes more distinctions than the less refined content of the intentional states those speech acts express.

How would this work? Which speech acts we perform is of course largely determined by the intentional states they express. But speech acts express those states in words that have, to some extent, an independent semantic life.¹³ So, although the content of speech acts derives mainly from the intentional states expressed, the words used in expressing that content often go beyond it, making for speech acts whose semantic meaning is correspondingly enriched and refined.

Our thoughts, by contrast, seldom need to respect the fine-grained semantic distinctions inherent in natural language. Any intentional state can typically be expressed equally well by a range of speech acts whose semantic meanings are not exactly the same. So it's reasonable to regard the content of our intentional states as neutral among the distinct semantic meanings of those various possible speech acts. Any particular choice of words, then, produces a speech act with semantic meaning more fine grained than the content of the corresponding intentional state. Some match of content is required for a speech act to count as expressing an intentional state, but the match need not be exact.

What words we choose to express an intentional state may itself sometimes be due to some other intentional state. In these cases we might regard the resulting speech act as expressing both intentional states: that which led to some such speech act's being

¹² Most forcefully and impressively defended by Jerry A. Fodor in, e.g., *The Language of Thought*, New York: Thomas Y. Crowell, 1975; "Propositional Attitudes," *The Monist*, LXI, 4 (October 1978): 501–523; "Why There Still Has to Be a Language of Thought," Appendix in *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge, Massachusetts: MIT Press/Bradford Books, 1987); "A Theory of Content," in *A Theory of Content and Other Essays*, Cambridge, Massachusetts: MIT Press/Bradford Books, 1990; and *The Elm and the Expert: Mentalese and its Semantics*, Cambridge, Massachusetts: MIT Press/Bradford Books, 1994.

¹³ Partly for reasons developed in compelling detail by Tyler Burge, "Individualism and the Mental," *Midwest Studies in Philosophy*, IV (1979): 73–121, "Individualism and Psychology," *The Philosophical Review*, XCV, 1 (January 1986): 3–45, and elsewhere.

performed and also that which influenced the particular choice of words. The speech act would be more fine grained in content than either intentional state, but might not outstrip the two combined. But choices of words are doubtless often arbitrary, resulting from no mental factor at all. Mere habit or word pattern might determine what words we use; indeed, it is likely that this is what typically happens. In such cases, our speech act does outstrip in content any relevant antecedent intentional states. And, since the content of intentional states can be less fine grained than the meaning of the speech acts that express those states, there is less temptation on this view to suppose that thinking fully mirrors the syntactic and semantic properties of natural language.

Indeed, it may sometimes be difficult to capture our more coarse-grained thoughts in words, especially with the intentional states we ascribe to nonlinguistic animals. Dennett therefore concludes that the " 'thought' [of such an animal] might be inexpressible (in human language) for the simple reason that expression in a human language *cuts too fine*." But Dennett concedes that "we may nevertheless exhaustively describe what we can't express."¹⁴ But if we can exhaustively describe the content of a thought, we can use those very words to frame a sentence that would, however awkwardly, express that thought. If the content of some thought elides various distinctions inherent in human language, suitable disjunctions can provide neutrality in respect of those distinctions. Even when no straightforward sentence could express the content of thought, some complex compound should succeed.¹⁵

A speech act expresses an intentional state only if its content matches that of the intentional state, but the match need not be exact. How close must that match be? No precise answer is possible, but neither should we expect precision. For one thing, it is unclear just how finely we can differentiate intentional states and speech acts in respect of content. In addition, the distinction between parroting and nonparroting speech productions itself doubtless admits of an intermediate gray area. Nor, finally, should we expect to be able to specify with precision just which speech acts express which intentional states. Some reasonably close match of content is required even though the precise degree of correspondence eludes specification. It is an advantage of the present model that it does not aim for such precision.

III. CONSCIOUSNESS AND VERBALLY EXPRESSING OUR THOUGHTS

Let's call this third picture of the relation between thinking and speech the *refinement model*. Arguably, it avoids the disadvantages of both the folk-psychological and

¹⁴ *Kinds of Minds: Toward an Understanding of Consciousness*, New York: Basic Books, 1996, p. 42. See also his "How to do Other Things with Words," in *Thought and Language*, ed. John Preston, Cambridge: Cambridge University Press, 1997, pp. 219–235.

¹⁵ It is possible that many human thoughts have content that makes them better suited to being expressed by specific nonverbal actions than by speech; doubtless this is true of the thoughts of nonlinguistic animals. One might have a thought, e.g., that is expressed by one's getting in out of the rain. Deliberate cases of such actions are not merely automatic behavior, since they have among their causes intentional states with suitable content. It may not be clear in these cases just what that content is—whether it is, e.g., to get out of the rain, to seek shelter, to stop getting wet, or what. But it would be surprising if that content could not be captured, in some gerrymandered way, in human language.

pandemonium models, by allowing us to explain the difference between parroting and nonparroting speech performances while circumventing the temptation to hypothesize a full-fledged language of thought. The refinement model shares with the pandemonium model the recognition that speech often outstrips our thoughts in content. But unlike the pandemonium model, the refinement model preserves both our folk-psychological taxonomy and the traditional view that speech acts express antecedent intentional states.

Nonetheless, the refinement model appears to face a difficulty that both the pandemonium and folk-psychological models avoid. As noted earlier, whenever we express our thoughts in words, our subjective impression is that what we say exactly matches the content of the intentional state we express. It never seems, from a first-person point of view, that our speech act is richer or more fine grained in content than our antecedent intentional state. The folk-psychological and pandemonium models both reflect this, since both posit an exact correspondence of content between speech act and intentional state, differing only about whether the intentional state fixes the content of the speech act or the other way around. Either way we would seem to sense a perfect fit between them.

Folk psychology generally trades, of course, on such conformity to our subjective impressions. And in this case the pandemonium model follows suit. Can we defend the refinement model despite its departure from our subjective impressions about thinking and speech?

A slight detour here will be useful. As noted earlier, putting our thoughts into words sometimes seems to clarify those very thoughts. By itself, however, the clarifying effect of verbally expressing our thoughts does not tell against an exact match of content between thought and speech. It could simply be that the process of finding words to express our thoughts forces us to clarify those very thoughts. As we mentally hear ourselves say what it is that we think, we find it confused or unclear, and so revise on the fly what we think. But by adjusting our words as we go, we get the right thought out. The content of the resulting speech act could still match exactly that of the suitably revised thought.

There is, however, another way in which putting words to our thoughts might have a clarifying effect on those thoughts. Rather than leading us to clarify the thoughts themselves, it could result instead in our becoming clearer about just what those thoughts are. We would know better, having spoken, just what it was that we had been thinking all along. We discover what we think only as we say it.

Dennett takes this kind of case to support the pandemonium model, on which the content of our thoughts is fixed only when we speak (*CE* 245); hence his striking epigram from E. M. Forster: "How do I know what I think until I see what I say?" (*CE* 193) But our discovering what we think only as we say it does not, by itself, support the pandemonium model, since it does not show that our thoughts lacked fixed content until we spoke. It could instead simply be that our thoughts often are not conscious in the relevant way until we express them in words. Even if the speech act results from an antecedent intentional state whose content exactly matches that of the speech act, perhaps it's only as we speak that we become conscious of the intentional state *as having that content*.

These considerations help with the problem the refinement model faced. On that model, our speech acts often outstrip the intentional states they express in respect of content. But that is never how it seems from a first-person point of view. Our subjective impression is always that our speech acts exactly match in content the intentional states they express.

But these first-person impressions may not accurately reveal the content of our intentional states. It could be that our speech acts do often outstrip our intentional states in content, but that we are nonetheless conscious of those intentional states *as having the richer content*. We might subjectively sense an exact match in content not because such a match obtains, but because that is how we are conscious of our verbally expressed thoughts. Because we are conscious of such a thought as expressed by a particular speech act, we in effect read back the content of the speech act onto the intentional state it expresses. We interpret the intentional state as having the content exhibited by the speech act.

The idea that we interpret our thoughts in the light of our speech acts fits well with the spirit of Dennett's pandemonium model. To some extent at least, we rely on the same considerations others do in determining what we think. As Dennett notes, in cases when we discover what we think only as we say it, "we are . . . in the same boat as our external . . . interpreters, encountering a bit of text and putting the best reading on it that we can find" (CE 245). Hence Dennett's heterophenomenological method, on which our theorizing about mind seeks to do justice to the verbal reports people make about their mental states.¹⁶

The pandemonium model, however, takes these self-interpretations to be the last word about the content of our thoughts.¹⁷ This squares with Dennett's first-person operationalism, which rejects the idea that there can be a difference between how things seem and how they seem to seem (CE 132). Intentional states are subjective states; they are a matter of how things seem to us. And how things seem to us in virtue

¹⁶ CE 72-85; also ch. 4, *passim*, and Dennett's "Two Approaches to Mental Images," in his *Brainstorms*, Cambridge, Massachusetts: MIT Press/Bradford Books, 1978, pp. 174-189.

¹⁷ The heterophenomenological method relies on the reports people make of their mental states; such reports in effect constitute self-interpretations. Dennett also regards cases in which we learn what we think by seeing what we say as involving a kind of self-interpretation (CE 245). But there self-interpretation occurs by way of speech acts that verbally express our intentional states, rather than explicitly reporting them.

Seeing verbal expressions of intentional states as self-interpretations as well as reports suggests that Dennett may be tacitly assimilating reports to verbal expressions. And that, in turn, may help explain why he regards self-interpretation as the last word about the content of our thoughts. Even if speech acts can differ in content somewhat from the intentional states they express, one might plausibly regard such expressions as better than any other evidence could be about the content of those states. So, assimilating reports of intentional states to their verbal expressions will make such self-interpreting reports themselves seem to be decisive about the content of the states those reports are about.

Assimilating the reporting of intentional states to their verbal expression also encourages the idea that mental states are all conscious, since the content of every mental state would itself then affirm, as reports do, that one is in that state. This idea may, moreover, be implicit in first-person operationalism, on which a state's mental properties are determined by how that state appears to consciousness.

of our being in some intentional state is a matter of the content that state has. So there is no difference, on first-person operationalism, between the content one's intentional states seem to one to have and the content they actually have. If it seems to one that an intentional state has a certain content, that's the content it has.

This approach saves an explanatory step. Why, if our speech acts outstrip in content the intentional states they express, do the intentional states subjectively seem to us to have the richer content of the speech acts? The pandemonium model avoids having to answer that question by positing that the intentional states really do have that richer content.

But even if the content of our speech acts does outstrip that of the thoughts they express, we can readily explain why those thoughts and speech acts seem, subjectively, to have the same content. As remarked earlier, a speech act counts as expressing a thought only if it has roughly the same content.¹⁸ But we also sometimes convey our thoughts by actually reporting them—by saying literally that we have those thoughts. Suppose I think that it's raining. I verbally express that thought by saying, simply, that it's raining, whereas I report that very same thought by saying, instead, that I think that it's raining.

Indisputably, two such speech acts differ semantically, since they have distinct truth conditions. Still, they are easily conflated, since, with minor qualifications about degree of conviction that won't matter here, we can appropriately say one thing whenever we can appropriately say the other. This equivalence of performance conditions, moreover, is second nature for us; any time I actually do say it's raining, I might as easily have said that I think that it's raining, and conversely.

Suppose, then, I think that p and I express my thought with the somewhat richer, more refined statement that p' . The richer statement that p' is, then, performance conditionally equivalent to the statement that I think that p' . And because this equivalence is second nature, I might as easily have said that I think that p' . But if I had made the higher-order remark that I think that p' , my statement would have expressed a thought that has roughly the same content; it would have expressed the HOT that I think that p' . And, since I might just as easily have made that statement, I must have had the HOT that the statement would have expressed.

The upshot is that, whenever I say that p' , I have a HOT that I think that p' . And that HOT determines how I am conscious of the thought my speech act expresses. So, whatever the actual content of my thought, I am conscious of that thought *as having the content that p'* . We can explain why, even though my speech act may be somewhat richer and more refined in content than the thought it expresses, I am conscious of the thought *as having* the richer, more refined content.¹⁹

¹⁸ It is arguable that, in addition, the intentional state must be causally implicated in producing the speech act. See my "Intentionality."

¹⁹ These considerations explain, more generally, why it is that all verbally expressed cognitive intentional states are conscious. Whenever I express my thought that p by saying that p , I could as easily have said that I think that p . So I must have had the HOT I think that p , and on the HOT model my thought that p will accordingly be conscious. It is unlikely that any other model can explain why verbally expressed cognitive states are conscious. For more, see my "Why Are Verbally

But a difficulty looms for this explanation. If first-order conscious thoughts can be less fine grained than the speech acts that express them, why can't the same happen with HOTs as well?²⁰ And if our verbal report of a first-order thought might outstrip in content the HOT that report expresses, the HOT need not, after all, reflect the richer content of that report. And then we wouldn't be conscious of the first-order thought as having that richer content.

Perhaps this is all so. But recall that we are not trying to show that speech acts invariably have more fine-grained content than the thoughts they express, but only that this may sometimes happen, despite our subjective sense that it never does. And we have no reason to think it ever does actually happen with HOTs.

Indeed, it is highly unlikely that it ever happens there. We think that first-order thoughts have less fine-grained content than their verbal expressions because it seems clear that what we think could have been expressed equally well by distinct, semantically nonequivalent speech acts. But that's not the case with the HOTs that accompany our verbally expressed intentional states. Suppose I verbally express my first-order thought that *p* by the more fine-grained remark that *p'*. How I express my HOT about my less fine-grained, first-order thought that *p* is now dictated by the performance-conditional equivalence between saying the more fine-grained *p'* and saying that I think that *p'*. So I will now express that HOT only by saying that I think that *p'*. Verbally expressing my thoughts constrains the way I am conscious of them.²¹

IV. THE REFINEMENT MODEL AND FIRST-PERSON OPERATIONALISM

On the refinement model, speech acts often outstrip in content the intentional states they express. To explain how this squares with our subjective impressions, we must distinguish the content those states have from the content they seem to us to have. But the content of intentional states is a matter of how things seem to us. So this explanation conflicts with Dennett's first-person operationalism, which denies any

Expressed Thoughts Conscious?", ch. 10 in this volume. In "Consciousness and its Expression," ch. 11 in this volume, I show how these considerations also explain why affective states, unlike cognitive states, are often verbally expressed without being conscious.

²⁰ This echoes a challenge Dennett raises for the HOT model, that the pandemonium model should apply at higher levels as well; "the second-order state," he claims, "comes to be *created* by the very process of framing the report" (315, emphasis Dennett's). Also: "The emergence of the [verbal] expression [of a higher-order thought] is precisely what creates or fixes the content of higher-order thought expressed" (315).

²¹ Dennett notes with approval Elizabeth Anscombe's argument that it is "wrong to claim that we *know* what our intentions are; rather we just *can say* what our intentions are" (CE 315 n. 10; see G. E. M. Anscombe, *Intention*, 2nd edn., Oxford: Basil Blackwell, 1963). The present explanation reflects this primacy of verbal expressions. On that explanation, the verbal expression of a thought constrains, by the performance-conditional equivalence, the content of our HOT about the target first-order thought, and hence how we are conscious of that target thought.

difference between things' seeming a certain way and their seeming to seem a certain way.²²

But there is reason to reject that denial. Consider the game Dennett describes of "Hide the Thimble," which dramatizes how we can look straight at an object we're trying to find and yet fail consciously to register it (CE 334). This kind of case invites us to distinguish our seeing something consciously from our seeing it without being conscious of seeing it; things may seem a certain way even though they don't consciously seem that way.

The striking and subjectively surprising limits on parafoveal resolution that Dennett cites take us even farther. Parafoveal vision can produce only low-resolution sensations of most of the Warhol Marilyns,²³ but it seems subjectively that we are aware of them all in a clear and focused way. What it's like for one to have a particular conscious sensation is a function of how one is conscious of that sensation. So the best explanation of this case is that we have blurry parafoveal sensations of most Marilyns, but the way we are aware of those blurry sensations represents them as having high resolution.²⁴ What our sensations of the Marilyns is like for us is a function of how we're conscious of those sensations.²⁵ And, if the way we're conscious of our sensations sometimes goes beyond their mental properties by in effect refining them or

²² The hierarchy of levels ends up, he claims, in our "having to postulate differences that are systematically undiscoverable by any means, from the inside or the outside," distinctions that are "systematically indiscernible in nature" (CE 319).

²³ CE 354. See CE 53–4 for Dennett's striking illustration of the limits on parafoveal vision in attempting to discern the color of playing cards seen parafoveally at arm's length.

²⁴ Similarly, in the thimble case; there is a conscious sensation in our visual field corresponding to the location of the thimble, since no subjective gap occurs there, but we are not conscious of that sensation *as* a sensation of a thimble.

We may also be conscious of sensations in ways that leave out aspects of their qualitative character, as when a throbbing pain is conscious only as painful and not as throbbing, or a sensation of red is conscious as red but not in respect of any particular shade. So, even though visual sensations appear to us to be "ultimately homogeneous," in Wilfrid Sellars' useful phrase, it may well be that their sensory qualities are actually composed of many pixels representing specific characteristics; their ultimately homogeneous appearance may be due only to the way our HOTs represent collections of such pixels. We are conscious of our sensations in a way that smooths them out, so to speak, and elides the details of their particulate, bit-map nature. (Wilfrid Sellars, "Philosophy and the Scientific Image of Man," in *Frontiers of Science and Philosophy*, ed. Robert G. Colodny, Pittsburgh: University of Pittsburgh Press, 1962, pp. 35–78; reprinted in *Science, Perception and Reality*, pp. 1–40, at p. 36.)

For more on sensations' diverging from the way we are conscious of them and the way HOTs function in that connection, see "The Independence of Consciousness and Sensory Quality," "Sensory Quality and the Relocation Story," and "Sensory Qualities, Consciousness, and Perception" (chs. 5, 6, and 7 in this volume), "Consciousness and Metacognition," in *Metarepresentation: A Multidisciplinary Perspective*, Proceedings of the Tenth Vancouver Cognitive Science Conference, ed. Daniel Sperber, New York: Oxford University Press, 2000, pp. 265–295, and "Explaining Consciousness," in *Philosophy of Mind: Classical and Contemporary Readings*, ed. David J. Chalmers, New York: Oxford University Press, 2002, pp. 406–421.

²⁵ Consider the striking results John Grimes reports, in which subjects fail consciously to notice dramatic changes of color and shape in a salient object if the change occurs during a saccade (John Grimes, "On the Failure to Detect Changes in Scenes across Saccades," *Perception*, ed. Kathleen Akins, New York: Oxford University Press, 1996, pp. 89–110). Our subjective sense that conscious

touching them up, it is highly likely that the same happens with our intentional states, as well.²⁶

On the pandemonium model, our intentional states come to have determinate content only when expressed by a speech act, or fixed by some other sort of probe. This applies equally, Dennett argues, to the higher-order states in virtue of which we are conscious of our conscious states; those higher-order states themselves “[come] to be *created* by the very process of framing the report.”²⁷ Such “heterophenomenological reports” are the verbal pronouncements people make about their own mental states, to which Dennett’s heterophenomenological method seeks to do justice. Since framing those reports fixes the higher-order content, Dennett holds that sincere heterophenomenological reports are constitutive of what it’s like for the subject at that time and, hence, constitutive of that subject’s consciousness. It is those heterophenomenological reports, rather than any antecedently occurring intentional states, that fix the contents of consciousness.

This conclusion might seem tempting because we are seldom conscious of the higher-order states in virtue of which we are, in turn, conscious of our conscious mental states. So our first-person impression will be that heterophenomenological reports alone determine how we are conscious of those states. But this first-person appearance is unreliable. We are in many intentional states of which we are not conscious, and we should expect that we typically remain unaware of the HOTs in virtue of which we are conscious of our conscious states. Such HOTs would be conscious only if we had third-order thoughts about them, and we can safely assume that that seldom happens.

Moreover, we are occasionally conscious of our HOTs, wholly independently of whether we express them in heterophenomenological reports. Focusing introspectively on some particular state makes us aware not only of that state but also of the reflective, higher-order state of being aware of the lower-order target. The only reason to hold that heterophenomenological reports, rather than the HOTs they express, fix the contents of consciousness is Dennett’s rejection of the folk-psychological model, on which speech acts generally express antecedent intentional states.

HOTs not only determine what it’s like for us to be in various mental states; they also are, in effect, subjectively spontaneous interpretations of the mental states we are in.²⁸ Indeed, there is an important connection between consciousness and

experience is continuously responsive to changes in what we see outstrips the extent to which we actually track changes in our visual sensations. This lends support to the conclusion, argued for by Dennett (*CE* ch. 11), that our subjective sense of great detail throughout our visual field is also illusory. The way we are conscious of our sensations goes beyond those sensations themselves.

²⁶ Because what it’s like to be in conscious states is sometimes informationally richer than the states themselves and that additional informational content is occurrent, we must posit occurrent states of higher-order awareness of the states, and not just dispositions to be aware of them.

²⁷ *CE* 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.*).

²⁸ Subjectively spontaneous because we are unaware of any antecedent factor that might explain their occurrence. Because HOTs are not themselves typically conscious, we will seldom be explicitly aware of engaging in the self-interpretations these HOTs embody.

our seemingly spontaneous self-interpretations. When we spontaneously interpret ourselves as being in certain states, we are conscious of ourselves *as being in those states*. What it's like for us to be in particular states is a function of how we spontaneously interpret those states. In the absence of any such spontaneous self-interpretations, our mental states simply aren't conscious; there is nothing it's like for us to be in them.

First-person operationalism is also interpretationist, but it imposes a special constraint all its own. Mental states are conscious, on first-person operationalism, only when accompanied by intersubjectively accessible probes. Because we have no access to the mental states of others apart from these probes, such probes provide our sole basis for interpreting what states others are in. So tying the consciousness of mental states to the occurrence of probes means that mental states are conscious only when they are available for interpretation by others. First-person operationalism holds that the consciousness of mental states is a function of interpretation, but only third-person interpretations matter.²⁹

But this third-person constraint is gratuitous. The self-interpretations that HOTs provide are by themselves sufficient for one's mental states to be conscious. Indeed, one might well wonder how third-person interpretations could have any bearing at all on the consciousness of mental states. The key is the denial by first-person operationalism of any difference between mental states and their being conscious—between how things seem and how they seem to seem. Given that denial, an interpretation of somebody as being in a particular state serves indifferently as reason to think that the person is in that state and that the state is conscious. First-person operationalism restricts itself to third-person interpretations because it denies the distinction between mental states and their being conscious.

Rejecting first-person operationalism allows us to distinguish being conscious of our mental states from the states we are conscious of. And that opens the way to explaining the consciousness of mental states along lines different from those invoked in explaining other mental properties. In particular, it may well be that mental states have determinate content even though the way we are conscious of those states is a function of the subjectively spontaneous self-interpretations embodied in HOTs. So, too, for the HOTs themselves; what it's like for us to be in various mental states is just a matter of the self-interpretations our HOTs embody, but those HOTs can themselves be determinate in content.³⁰ Interpretationism can be true about what it's like

²⁹ And such third-person interpretations cannot distinguish between a state's never coming to be conscious and its coming to be conscious too briefly to affect memory.

³⁰ As well as determinate as to time of occurrence, which makes it determinate when our mental states come to be conscious.

The HOT model may sometimes count a state as conscious even though there is nothing it's like for one to be in that state, e.g., in some cases where we find it hard to decide between Stalinesque and Orwellian explanations. The occurrence of a HOT would be decisive in any event, regardless of first-person impressions; it is standard for well-established theories to resolve problem cases when the empirical input, including first-person impressions, does not suffice.

for us to be in mental states without thereby holding for their content or other mental properties.³¹

³¹ Work on this paper began in 1998 at the Center for Interdisciplinary Research (ZIF), University of Bielefeld, Germany. I am grateful to the Center for congenial and stimulating surroundings and to a PSC-CUNY Research Award for supporting my stay there. I am also grateful for helpful reactions to an earlier draft from the participants in the November 1998 Memorial University of Newfoundland Conference on Dennett and from members of the CUNY Cognitive Science Symposium.

CONSCIOUSNESS AND MIND

DAVID M. ROSENTHAL

CLARENDON PRESS • OXFORD