

## V. CONSCIOUSNESS, INTERPRETATION, AND HIGHER-ORDER-THOUGHT

DAVID M. ROSENTHAL

### 1. THE PROBLEM OF EXPLAINING CONSCIOUSNESS

Few contemporary researchers in psychology, philosophy, and the cognitive sciences have any doubt about whether mental phenomena occur without being conscious. There is extensive and convincing clinical and experimental evidence for the existence of thoughts, desires, and related mental states that aren't conscious.

We characterize thoughts, desires, intentions, expectations, hopes, and many other mental states in terms of the things they are about and, more fully, in terms of their content, as captured by a sentence nominalization, such as a clause beginning with the word 'that'. The philosophical literature follows Franz Brentano's adaptation of Thomist terminology in referring to all such states as intentional states.

But there is another type of mental phenomena, which lack intentionality and whose mental nature consists instead of some qualitative feature. These states include bodily sensations, such as aches and pains, and perceptual states, such as visual sensations of color and tactile sensations of heat and cold. And these states all exhibit some mental quality or another, such as the mental quality distinctive of pain or the mental quality of red or blue.<sup>1</sup> And even theorists who acknowledge that intentional states can and do occur without being conscious have sometimes insisted that qualitative states cannot. There is, according to these theorists, nothing to a state's being qualitative or exhibiting some mental quality unless that state is conscious – unless it is, as we might metaphorically say, "lighted up".

It's striking that Freud himself seems to have adopted this double standard toward the two types of mental state. In his metapsychological paper, "The Unconscious", for example, he writes that "all the categories which we employ to describe conscious mental acts, such as ideas, purposes, resolutions, and so forth, can be applied to [unconscious mental occurrences]" (Freud 1915e, p. 168). But he seems here to have in

---

<sup>1</sup> It is crucial to distinguish the mental qualities characteristic of perceptual sensations from the distinct, but corresponding physical qualities characteristic of the objects perceived. On this, see Rosenthal (1999a).

mind only purely intentional states, which altogether lack qualitative character.

Later in that same paper, by contrast, he writes that “[i]t is surely of the essence of an emotion, that we should be aware of it. [...] Thus the possibility of [...] unconsciousness would be completely excluded so far as emotions, feelings and affects are concerned” (Freud 1915e, p. 177; see also 1923b, pp. 22–3; 1940a, p. 197). To describe feelings as unconscious, he concluded, is to “speak in a condensed and not entirely correct manner” (Freud 1923b, p. 22) about cases in which the representational character of the feelings is repressed or misrepresented (Freud 1915e, pp. 177–8).<sup>2</sup>

But such a double standard is unjustified. There is ample evidence that qualitative states, no less than intentional states, occur without being conscious. In masked-priming experiments, subjects presented very briefly with two successive stimuli report being wholly unaware of the first stimulus, even though it has a demonstrable effect on mental processing. And blindsight subjects, part of whose primary visual cortex has been destroyed, report that they don’t see visual stimuli in the relevant area of the visual field, but their forced-choice guesses about visible characteristics of such stimuli are startlingly accurate. Because these effects correspond to differences among mental qualities in the absence of conscious sensing, they point to the occurrence of qualitative states that aren’t conscious.<sup>3</sup>

Why, then, do many theorists continue to deny the occurrence of nonconscious qualitative states? One reason stems from a conflation of two ways we use the term ‘conscious’. We use that term to mark a difference between two conditions that a person or other creature might be in and also to mark a difference between two ways that mental phenomena occur.

Consider first mental phenomena. When somebody has some thought, desire, perception, or feeling but is in no way whatever aware of that state, the state plainly is not a conscious state.<sup>4</sup> It follows that conscious states are those we are conscious, in some suitable way, of

---

<sup>2</sup> Freud did allow, however, that something corresponding to conscious qualitative states can occur without being conscious (Freud 1923b, p. 22). So it may be that the issue here is simply verbal, hinging not what states can occur but on how one uses the terms ‘feeling’ and ‘sensation’.

<sup>3</sup> On masked priming see Marcel (1983a, 1983b). On blindsight, see Weiskrantz (1997, 1986) and Merikle et al. (2001).

<sup>4</sup> I use ‘aware of’ and ‘conscious of’ interchangeably here.

being in. In section II I'll take up the question of just what that suitable way might be.

There is no circularity in explaining what it is for a mental state to be conscious by reference to one's being conscious *of* that state. We understand what it is to be conscious *of* something independently of any appeal to mental states' being conscious or not. We're conscious of something when we see it or hear it or sense it in some other way, and we're conscious of things when we have thoughts about them as being present.

In addition to using the term 'conscious' to mark the difference between those mental states we are suitably aware of and those we are not, we also use that term to mark a difference in two conditions that people and other creatures may be in. A person or other animal is conscious if it's awake and mentally responsive to sensory stimulation. This is plainly a different property from that of a mental state's being conscious. I'll use the terms *creature consciousness* and *state consciousness* to refer, respectively, to the property a creature has when it's conscious and the property a mental state sometimes has of being conscious.

It's obvious that creature consciousness does not imply state consciousness, since people while awake are typically in many mental states that aren't conscious. But creature consciousness involves a creature's being mentally responsive to sensory stimulation. So it may be tempting to think that creature consciousness implies at least the consciousness of whatever sensory states a creature is in when that creature is conscious. And perhaps, then, all qualitative states whatever are conscious. A little reflection, however, will dispel that illusion. By itself, a creature's being conscious cannot preclude its sensing some things nonconsciously. Indeed, a creature might be awake and mentally responsive to sensory stimulation without any of its qualitative states being conscious. This never happens in the normal human case, but it might well occur with other creatures, fewer of whose mental states are conscious.

A second explanation for the widespread idea that qualitative states cannot occur without being conscious stems from a feature of perceptual sensations that's unique among mental states. Intentional states, such as thinking, desiring, anticipating, doubting, wondering, and the like can all be expressed in speech, as well as in various forms of nonverbal behavior. The same holds for affective states, such as anger, joy, sadness, and fear, which themselves are in part intentional; when we are angry, joyous, sad, or fearful, we are angry, joyous, sad, or fearful that something is the case, and we can express these feelings both verbally

and nonverbally.<sup>5</sup> Even bodily sensations, such as pains, tickles, and itches, are expressible, by saying ‘ouch’ or writhing, by laughing, and by scratching.

By content, we cannot express our perceptual sensations, such as visual sensations of red or auditory sensations of a violin, either in speech or nonverbally. We express our perceptions of these things, which have intentional content, for example, by saying that something is red or that it’s a violin. But, since perceptual sensations have only mental qualities, and no intentional content, there is accordingly no way to express them. And, because we cannot express these states, it’s inviting to think that the only way we can detect their presence is by being conscious of them. The purely qualitative character of perceptual sensations, it may seem, means that their only detectable effect is in being conscious. And that makes it seem, in turn, as though there is nothing to a perceptual sensation’s occurring unless it’s conscious. And it may then seem natural to extrapolate to all states that have qualitative character of which sort, including bodily sensations and emotions.

But the absence of any way to express perceptual sensations does not actually imply any such result. Though the purely perceptual qualitative character of perceptual sensations prevents them from having, by themselves, any effect on behavior, they do have effects on other mental state states and processes. They characteristically lead, for example, to our having full-fledged perceptions with intentional content and to our having perceptual beliefs. Perceptual sensations are detectable, even when they aren’t conscious, by their causing other mental states. Indeed, this is exactly how the occurrence of nonconscious perceptual sensations in blindsight and masked priming is experimentally detected. The inexpressibility of perceptual sensations does not support the idea that qualitative states cannot occur without being conscious.

There is a third, somewhat more theoretical reason why many writers have held that qualitative states cannot occur without being conscious, which stems from a familiar idea about how to reconcile our commonsense picture of reality with that of modern physics. Physics requires that the properties of physical objects can all be described in mathematical terms, but it seems that commonsense physical qualities, such as color and sound, resist such mathematical treatment. A standard solution is simply to deny that color and sound, conceived as common

---

<sup>5</sup> I return to the expressibility in speech and nonverbal behavior of cognitive and affective states in section 4.

sense does, are properties of physical objects at all, and insist that these qualities occur instead only as mental properties.

This manoeuvre has a striking, if largely unnoticed, side effect. Commonsense physical qualities seem problematic only when we're conscious of them. So we will be tempted to relocate as mental properties only those physical qualities that we're conscious of. But that means that the relocated mental qualities will have consciousness already built in. This device for reconciling common sense with physics actually distorts our very conception of mental reality.<sup>6</sup>

Adjusting the way we square common sense with mathematical physics avoids this effect. We can, instead, simply identify the physical qualities of common sense with physical properties that a mathematical physics can accommodate, colors, for example, with suitable surface reflectance properties.<sup>7</sup> We can then explain the commonsense character of such properties by appeal to the way we're conscious of them, without thereby supposing that this commonsense character is part of the properties we're conscious of. And, once we stop construing mental qualities as relocated versions of physical qualities, there is no remaining reason to see consciousness as automatically built into all mental qualities.

## 2. HOW WE ARE CONSCIOUS OF OUR CONSCIOUS STATES

A mental state is conscious when the individual that's in that state is, in some suitable way, conscious of being in that state. So the main job of a satisfactory theory of what I've called state consciousness is to say just what that suitable way is.

There are two traditional answers to this question. One is the view, often associated with Brentano, that the consciousness of our conscious states is intrinsic to the states themselves. Every mental state, according to Brentano, is not only about its ostensible object; it is also about itself. As Brentano he put it, all mental acts "apprehend [themselves,] albeit indirectly" (Brentano 1884, p. 128). It is inviting to wonder whether the view that all intentional states have dual content, which Freud would certainly have heard in Brentano's weekly lectures, may have influenced Freud's own distinction between the manifest and latent content of thoughts and desires.

---

<sup>6</sup> For more on this see Rosenthal (1999b).

<sup>7</sup> More precisely, suitable ratios of such properties.

The other traditional view is that we're conscious of our conscious states by way of a kind of "inner sense", in Kant's useful phrase;<sup>8</sup> we're conscious of our conscious states by sensing them (Kant 1787, A22, B37, p. 174). It's likely that Freud held some version of this view; "consciousness", he wrote, plays the role "of a sense-organ for the perception of mental qualities" (Freud 1900a, p. 615).<sup>9</sup> Both these traditional models, however, have important failings, though these difficulties combine to point the way toward a more successful theory. Consider first the inner-sense model. One reason it seems inviting to hold that qualitative states are conscious when we sense them is the idea that such higher-order sensing might result in the conscious qualitative states' having their "lighted up" qualitative character. The higher-order sensing would itself contribute the relevant conscious quality. But even if that were so, this wouldn't apply to our conscious thoughts, desires, wishes, hopes, and other intentional states; it's highly implausible that we sense those states in any way whatever.

But there is a deeper and more decisive problem for the inner-sense model. Sensing consists in being in states that have some qualitative character or other. That's why higher-order sensing seems to help with conscious qualitative states but not conscious intentional states. But, when a qualitative state is conscious, only one qualitative character seems to occur, that of the state that's conscious. There doesn't, in addition, seem to be a second qualitative property that belongs to some higher-order sensation. But qualitative character is distinctive of sensations. So, if there's no higher-order qualitative character, there's also no higher-order sensation. As inviting as the inner-sense model may initially seem, it cannot be sustained.

There is another, related difficulty. Every sensation not only has some distinguishing qualitative character, but it also belongs to a particular sensory modality, such as vision, hearing, and so forth. But, as

---

<sup>8</sup> John Locke uses the term "internal Sense" (1700, 2, i, 4, p. 105); cp. his remark that "[c]onsciousness is the perception of what passes in a Man's own mind" (1700, 2, i, 19, p. 115). Locke also seems sometimes to hold that our consciousness of our conscious states is an intrinsic aspect of those states, as when he writes that "thinking consists in being conscious that one thinks" (1700, 2, i, 19); so perhaps he held that we sense our conscious states by way of an intrinsic aspect of those states. See also Armstrong (1980, p. 61) and Lycan (1996).

<sup>9</sup> I have used 'mental' in place of Strachey's somewhat anachronistic 'psychical'. On Freud's perceptual model of consciousness, see also Solms (1997a, 1997b) and Rosenthal (1997a).

Brentano noted, following Aristotle, if we sense our conscious qualitative states, it's not obvious what modality the higher-order sensation could have. The modality of the first-order state won't do, since that modality senses physical qualities, not sensations, and no other modality seems available (Brentano 1884, p. 127).<sup>10</sup>

Brentano actually invokes this very consideration in support of his alternative model of consciousness. If we're not aware of our conscious states by distinct, higher-order sensations of those states, our awareness of our conscious states must, he thought, be due instead to something intrinsic to the states themselves. Whether one's awareness of a mental state counts as being intrinsic to that state will of course hinge on how we individuate mental states generally. And it's not immediately obvious what standard of individuation we could use that wouldn't beg the question at hand.

I'll return shortly to the question of individuating mental states. But I want now to note a more serious difficulty that faces Brentano's intrinsicist model. Many mental states occur without being conscious. Brentano recognized that a mental state's occurring without being conscious is conceptually coherent (Brentano 1884, p. 128),<sup>11</sup> but he denied that this ever happens. But we know that it does; so we must explain how it is that some mental states occur consciously and others not. Building consciousness into the state itself makes that difficult, if not impossible. Mental states sometimes shift between being conscious and not being conscious; how can the intrinsicist story explain that? Intrinsicistism goes naturally with the outmoded view that mental states never occur without being conscious.<sup>12</sup>

It's clear that Freud rejected any such intrinsicist view. If a state's being conscious does consist in one's being conscious of it, it's natural to expect that it could occur without one's being conscious of it and,

---

<sup>10</sup> Aristotle's argument, which Brentano adapts, assumes that we do somehow sense that we see. If so, and if the sense we use to sense that we see is distinct from the sense of sight, then that other sense and the sense of sight would both have color as their proper object. But distinct senses never share the same proper object (Aristotle, *de Anima* III 2, 425b12-4).

<sup>11</sup> "An unconscious consciousness is no more a contradiction in terms than an unseen case of seeing" (Brentano 1884, p. 102), and "[a] presentation of a sound without a presentation of the act of hearing would not be inconceivable" (Brentano 1884, p. 128).

<sup>12</sup> Indeed, Brentano's intrinsicist view seems to be his only reason for holding that mental states are always conscious.

hence, without its being conscious. The default condition for mental states will thus be for them to occur nonconsciously. Freud explicitly drew this conclusion, writing that “[t]he mental, whatever its nature may be, is in itself unconscious.” Consciousness is “an inconstant quality” of mental occurrences, “one that is far oftener absent than present” (Freud 1940b, p. 238).

Since the default condition for mental states is that they occur nonconsciously, no mechanism of repression is needed for such nonconscious occurrence of mental states. In and of itself, every mental state occurs without being conscious. A repressed state is one that occurs consciously, because an accompanying higher-order state occurs in virtue of which one is conscious of the state, but the mechanism of repression then blocks that state’s being conscious. So that mechanism must interfere with the higher-order state in virtue of which the repressed state would otherwise be conscious.

Let’s take stock. The inner-sense model fails because there are no higher-order qualities, distinct from the first-order qualities of the states we’re conscious of, and we know of no modality that such higher-order qualities could belong to. The intrinsicist picture fails because we have no independent way to individuate mental states that yields the right result and, more important, if consciousness is intrinsic to mental states, we cannot explain why mental states are sometimes conscious and sometimes not.

Mental states are conscious when we’re conscious of them. There are only two ways we know of in which we’re conscious of things, by sensing them and by having thoughts about them as being present. Since we aren’t conscious of our conscious states by sensing them or by way of something intrinsic to the state, the only remaining possibility is that we’re conscious of them by having thoughts about them as being present, thoughts that are extrinsic to and distinct from the conscious states, themselves.

We can say more about what these higher-order thoughts (*HOTs*) must be like. When a mental state is conscious, one is conscious of oneself as being in that state. So the content of a *HOT* must be that one is in that very state.<sup>13</sup> *HOTs* need not conceptualize their target states in any very elaborate way; a mental state can be conscious without one’s being

---

<sup>13</sup> More precisely, it must be that I, myself, am in that state; it’s not enough for me simply to ascribe the state to an individual that happens, unbeknownst to me, to be me. On this and related issues, see Rosenthal (2003).

conscious of it in the sophisticated way adult humans think about mental states. Thus children may well lack adult human concepts for thoughts, sensations, feelings, and even the mental itself.<sup>14</sup> All that's needed is that the HOT make one aware of the state in respect of some suitable property, a property that we adult humans would classify as mental.

We are seldom aware of any such HOTs. But that's no problem for the model. We wouldn't be aware of any HOT unless it was conscious, and that will happen only if a HOT is in turn accompanied by another HOT about it, in effect, a third-order thought. This is rare, but it does sometimes happen. There is a difference between mental states' being conscious in the ordinary, everyday way and their being introspectively or reflectively conscious. And mental states are reflectively conscious when they are accompanied not simply by HOTs, but by HOTs that are themselves conscious. The reflective character of introspective consciousness results from one's actually being conscious of being aware that one is in some particular state. Since HOTs are typically not conscious, moreover, we avoid the regress that would occur if we explained one state's being conscious by appeal to another that must itself be conscious.

When a mental state is conscious, one is conscious of that state in a somewhat special way that many theorists have characterized as direct or immediate. Thus Descartes, who builds consciousness into his concept of a mental state, insists that "the word 'thought' applies to all that exists in us in such a way that we are immediately conscious of it" (Descartes 1641, p. 160).

This apparent immediacy doubtless explains much of the intuitive appeal of the intrinsicist and inner-sense models. One's awareness of a mental state would plainly be immediate if it were intrinsic to that state, and sensing things also seems, intuitively, to be direct and unmediated. But the HOT model can also explain this sense of immediacy. We can require that HOTs rely on no conscious inference, no inference, that is, of which one is conscious. Suppose I come to have a thought that I feel a certain way solely because you tell me and I have confidence in you; that HOT won't make the feeling conscious. For the feeling to be conscious, I must have a HOT that's independent of any such conscious inference. It's this independence of conscious inference that results in

---

<sup>14</sup> For developmental work that suggests this, see (Perner 1991; Perner, Leekam & Wimmer 1987; Wellman 1990; Wimmer & Perner 1983). For an alternative view see Fodor (1992), and a reply by Perner (1995).

the subjective sense that nothing mediates between the feeling and my awareness of it.

This raises a theoretically delicate issue. Suppose you tell me that I feel a certain thing and, though it doesn't subjectively seem to me as though I do, I believe you. I come to have a HOT based on consciously inferring from what you said. My feeling isn't conscious, just as the theory provides. But suppose now that my feeling does become conscious; this presumably often happens in psychoanalytic treatment. How can the HOT model explain this shift?

Originally my HOT was based on a conscious inference; I had the thought that I had a certain feeling, and I also consciously thought I had that feeling because you'd said I had it. Later, even though I may consciously remember what you'd said, it comes to seem to me that I'm conscious of the feeling independently of your having said anything. My thought that I have that feeling comes to seem, subjectively, to be independent of your remark. Since my HOT is now independent of any conscious inference, the feeling now is conscious. In what follows, I'll assume these qualifications, using the term 'HOT' to apply only to the noninferential cases and the term 'noninferential' to mean independence of inference of which one is conscious. And I'll return to these issues in closing.

HOTs that rely on conscious inference and those which do not are not two kinds of HOT; the very same HOT might at one point rely on conscious inference and at another point not. The difference is wholly extrinsic to the HOT, itself. Nor, therefore, is the difference a matter of the genesis of the HOT, since a HOT might first occur as a result of conscious inference and subsequently come to be independent of the conscious inference. All that matters is whether it seems, subjectively, as though one would be conscious of the target state even in the absence of any such inference. It does not even matter whether one actually would be conscious of the state even without any such inference, but only whether it seems that way.<sup>15</sup>

---

<sup>15</sup> On so-called conceptual-role theories of intentional content, a state's content is a matter of its inferential connections with other states. So one might object that the HOT model conflicts with a conceptual-role account of content, since HOTs would vary in content depending of whether they are based on conscious inference. (I am indebted to Pierre Jacob, personal communication, for this ingenious objection).

But conceptual-role theories characterize content not as a function not of the actual inferences one makes but as a function of dispositions to infer, of what inferences

HOTs cannot be dispositional,<sup>16</sup> since being disposed to have a thought about something doesn't make one conscious of that thing. More important, HOTs cannot be internal to the states they are about.<sup>17</sup> So the HOT model cannot be combined with Brentano's intrinsicist view. The reason derives, ironically, from something that Brentano noted, namely, that we individuate intentional states by reference to the mental act being performed (Brentano 1884, p.127). But we individuate mental acts, in turn, partly by reference to the type of mental attitude involved, such as believing, desiring, doubting, wondering, expecting, and the like.

Suppose, then, that the HOT in virtue of which a mental state is conscious is internal to that state, and suppose also that the state in question is a case of wondering something. Wondering about something doesn't make one conscious of that thing; an intentional state makes us conscious of the things it is about only when that state has an assertoric mental attitude. So the HOT in virtue of which our state of wondering is conscious must itself be assertoric. But, since the HOT will then have a different type of mental attitude from its target, the HOT and its target will be distinct states. HOTs cannot be internal the states they are about.<sup>18</sup>

It's worth distinguishing this view from the position put forth independently by Daniel C. Dennett and Bernard Baars that a mental state's being conscious consists in its being widely broadcast within our cognitive system. Thus Dennett writes:

“Consciousness is cerebral celebrity. [...] Those contents are conscious that persevere, that monopolize resources long enough to achieve certain typical and “symptomatic” effects – on memory, on the control

---

one might make. And HOTs will bear the same inferential relations so construed. Moreover, what matters to the HOT model is only whether the actual inferences are conscious, and that wouldn't in any case matter to a conceptual-role account of content.

For conceptual-role theories, see Harman (1973, 1999, part III) and Block (1986).

<sup>16</sup> As Peter Carruthers has argued they are; see, e.g., Carruthers (2000).

<sup>17</sup> As argued by Gennaro (1995) and, more recently, Kriegel (2002, 2003).

<sup>18</sup> This may recall the difficulty the inner-sense model faced about a possible mismatch of sensory modality between the higherorder and the target sensings. For more on the HOT hypothesis, see Rosenthal (1986, 1997b, 2002).

of behavior and so forth” (Dennett 1993, p. 929; see also Dennett 1991; Baars 1998, 1997).<sup>19</sup>

Doubtless many conscious states satisfy this description. But many states that meet the description are nonetheless not conscious. A state can be widely broadcast to cerebral areas, monopolize resources, and have the characteristic effects Dennett mentions without being conscious. Indeed, exactly that presumably occurs with many states that guide and influence our everyday behavior. Such states widely influence both cerebral functioning and behavior without our being in any way whatever conscious of them. Indeed, it is tempting to think that cerebral broadcasting corresponds more to Freud’s notion of the preconscious than to any intuitive notion of a mental state’s being actually conscious. Similarly, a state can be conscious despite its having little such influence. Though they often go together, consciousness is not cerebral celebrity.

As noted earlier, Freud evidently adopted a version of the sensory model of what it is for mental states to be conscious states. It’s not clear, however, how that view squares with his conviction that qualitative states are always conscious. It seems to result in the kind of regress just mentioned. The states in virtue of which we perceive things all exhibit some mental quality. On the perceptual model, a mental state’s being conscious consists in its being accompanied by a higher-order qualitative state. But, if qualitative states never occur without being conscious, that higher-order qualitative state will in turn have to be accompanied by a still higher-order qualitative state, and so forth without limit. Perhaps it was this kind of consideration that led Freud to write that “the fact of consciousness” “defies all explanation or description” (Freud 1940b, p. 282).

### 3. CONSCIOUSNESS AND CONFABULATION

Theorists of a traditional Cartesian bent hold that mental states are all conscious. But often they also insist on two other views, which go beyond that Cartesian dictum, namely, that consciousness represents our mental states infallibly and that it represents those states exhaustively.

---

<sup>19</sup> Ned Block’s notion of access consciousness is also of this sort. A state is access conscious if it is “poised to be used as a premise in reasoning, [...] [and] for [the] *rational* control of action and [...] speech” (Block 1995, p. 231).

These three views all stem from a fundamental conviction that the mind is transparent to itself, in Gilbert Ryle's apt, disparaging metaphor, that the mind is "self-intimating" (Ryle 1949, p. 158).<sup>20</sup> And the mind's being transparent to itself would ensure not only that all states of the mind are conscious, but also that consciousness invariably represents those states accurately and fully.

The HOT model denies all these things. Not only are mental states not all conscious; we have, in addition, no reason to suppose that HOTs will represent the mental states they are about in respect of all their mental properties, nor even that they will always represent those mental targets accurately.

This is all to the good. Once the view that mental states are all conscious loses its grip on us, the views about exhaustiveness and infallibility should strike us, in the words of C. D. Broad, as "a curious superstition" (Broad 1925, p. 284).<sup>21</sup> And there is ample evidence that neither view is correct.

Consider first exhaustiveness. Suppose I consciously see something red. In most ordinary circumstances my sensation won't be conscious in respect of that specific shade, but only in respect of some generic red. But attending to the sensation often results in the sensation's becoming conscious in respect of some far more specific shade. This gives us reason to think that, even when the sensation is conscious only as being generically red, the mental character of the sensation itself is typically far more specific. The mental quality of the sensation is all along the more specific shade, of which we're conscious sometimes only generically and at other times with greater specificity. Consciousness does not invariably represent our mental states fully and in respect of all their specific mental properties; indeed, it presumably does so relatively seldom.

How about infallibility? Even if consciousness fails to represent everything mental about our conscious states, perhaps it is, still, reliable about the mental properties it does represent our conscious states as having. But there is reason to reject this as well. Clinical contexts provide examples of consciousness misrepresenting which mental states

---

<sup>20</sup> And, for the case of qualitative states, that the mind is "self-luminous" (Ryle 1949, p. 159).

<sup>21</sup> Broad's remark was specifically about the information that introspective consciousness delivers, and not about ordinary, unreflective consciousness, but the same point applies.

one is in, and there is also experimental evidence from social psychology. Subjects sometimes report having beliefs and desires when there is independent evidence that they don't actually have them, especially when those beliefs and desires fit with an explanatory or motivational story the subjects want to adopt (Nisbett & DeCamp Wilson 1977; Greenwald & Banaji 1995; White 1988; Wilson, Hodges & LaFleur 1995).

Such confabulation can occur not only with intentional states, such as beliefs and desires, but even with qualitative states. A dramatic example involves dental patients, who sometimes report pain when drilling occurs in a tooth in which pain cannot occur, because the nerve is anesthetized or dead. The standard explanation, known as dental fear, is that the patient has sensations of vibration and fear, and misconstrues the unpleasant experience as pain. Giving the patient this explanation typically results in the patient's no longer reporting pain when drilling resumes, but rather sensations of vibration and fear. Still, the patient also continues to have a vivid memory of the earlier conscious experience as being pain. Consciousness sometimes actually misrepresents our conscious qualitative states.<sup>22</sup>

The HOT hypothesis not only accommodates these phenomena; it predicts them. The reason a mental state is conscious when it's accompanied by a HOT is that one's HOT, in representing one as being in that state, makes one conscious of being in it. In having a HOT about a state, one is conscious of oneself as being in that state. But, as a general matter, having a thought makes one conscious of something only in respect of those properties which the thought attributes to that thing. When I have a thought about some object as being present, I'm conscious of that object in just the way that thought describes it. Being conscious of it in any other way would require my having a thought or sensation that represented the object in that other way.

HOTs are simply a special case of this. One is conscious of a conscious state in respect of just those properties which the HOT represents that state as having. When one has a conscious sensation of red, for example, the way one is conscious of the mental qualities of that sensation is a matter of how one's HOT represents it. So the very same sensation of red can be conscious in respect of a relatively generic redness or a more specific shade, depending on how one's HOT describes that

---

<sup>22</sup> For evidence of related phenomena, see Staats, Hekmat & Staats (1998) and Holmes & Frost (1976).

sensation, depending, that is, on what mental quality the HOT attributes to the sensation. And, since no thought captures all the properties of the things it's about, no HOT will represent all the mental properties of its target mental state. Consciousness is not exhaustive.

Similarly, the HOT model also predicts that consciousness is not infallible. Not all our thoughts represent things accurately, and those which don't can make us conscious of things in inaccurate ways. So it is with HOTs. In having an inaccurate HOT, one is conscious of oneself as being in a particular state under an inaccurate description. This kind of thing presumably happens in clinical situations. One may, for example, be motivated not to be conscious of oneself as having a certain desire or feeling. And one's motivation not to be conscious of oneself as having that desire or feeling may lead one to misrepresent the state and to be conscious of oneself as having, instead, a somewhat different desire or feeling. On the HOT model, one's HOTs misdescribe the relevant state so as to make it less unacceptable. One is conscious of the desire or feeling as being a state of a different sort from that which one wishes to avoid acknowledging. It is because this sort of thing can happen that it's open for Freud to claim that the feelings we call unconscious do not literally fail to be conscious. Rather one is conscious of those states in respect of their qualitative character of the feeling but not their intentional content.<sup>23</sup>

In the extreme case, one's HOT might even represent one as being in a mental state that one isn't in at all. But it's important to note that this case differs only in degree from the case just considered, in which a HOT represents one as being in a state but inaccurately characterizes the state. Strictly speaking, when a thought is inaccurate, the thing that thought purports to be about doesn't actually exist.

What matters for conscious subjectivity, then, is how one's HOTs represent things, that is, what states they represent one as being in. It doesn't matter for conscious subjectivity whether one is actually in that state. Consider two cases. In the first, one is in a particular mental state and one also has an accompanying HOT that correctly characterizes that state; in the second, one has that very same HOT but the state is absent. In both cases one is conscious of oneself as being in that state. So the two cases will be subjectively indistinguishable. There will doubtless be functional differences between the two cases; every mental state has

---

<sup>23</sup> Or, in any event, not the intentional content they actually have. See Freud (1915e, pp. 177-8; 1923b, pp. 22-3; 1940a, p. 197).

causal connections with behavior and with other mental states. So, if the target state is absent, that will be detectable. Still, there will be no difference between the two cases so far as conscious subjectivity is concerned.

#### 4. CONSCIOUSNESS, SPEECH, AND INTERPRETATION

A major advantage of the HOT model is that it explains how mental states can be conscious in respect of one of its mental properties but not another. A mental state is never simply conscious, *tout court*; it is conscious *as* a state of a particular mental type. The model explains this because a HOT specifies the mental type in respect of which one is conscious of a state.

HOTs, moreover, do a far better job than inner sense could in capturing the way mental states are conscious in respect of various mental properties. Sensing represents things only by way of properties that are specific to the several sensory modalities, whereas there are no limits to the ways thoughts can represent things. So thoughts are far more flexible than sensing would be in capturing variations in the way mental states are conscious.

Another major advantage of the HOT model is that it explains the close and important connection that holds in humans between consciousness and speech. For creatures, such as adult humans, who can readily say what mental states they are in, a state's being conscious coincides with one's being able to report noninferentially that one is in that state. This connection is widely accepted in experimental psychology as a mark in humans of a state's being conscious.<sup>24</sup>

I report a mental state when I say that I am in that state. But every sincere, meaningful speech act expresses a thought with the same content as the speech act and a mental attitude that parallels its force. That's why one cannot meaningfully say things such as 'It's raining but I don't think it is'; the first part purports to express a thought that the second part denies the existence of.<sup>25</sup> So a noninferential report that one is in a

---

<sup>24</sup> See, e.g., Marcel (1983a, 1983b). Writers seldom if ever make explicit that the report must be noninferential, but it's plainly understood; nobody would hold that reportability that relied on conscious inference coincides with consciousness.

<sup>25</sup> This kind of construction was first noted by G. E. Moore, who pointed out that, although it is not contradictory, one cannot make any statement with such a

mental state always expresses one's noninferential thought that one is in that state. It expresses, that is, a HOT.

It follows that the ability to report noninferentially that one is in a mental state is the same as the ability to express verbally a noninferential HOT to the effect that one is in that state. This allows us to explain why we can noninferentially report any conscious mental state. We can do so because, when a mental state is conscious, it's accompanied by the very HOT that a report of that state would express. Similarly, the reason we cannot report any mental state that isn't conscious is that the HOT that such a report would express doesn't occur. The HOT hypothesis is the best explanation of the connection between consciousness and noninferential reportability.<sup>26</sup>

The HOT hypothesis explains as well another striking aspect of the connection between speech and consciousness, namely, that whenever we verbally express a thought, that thought is conscious. When I say that it's raining, I verbally express my thought that it's raining. I could also express that very same thought nonverbally, for example, by taking my umbrella as I go out.<sup>27</sup> When I express my thought nonverbally, the thought may well remain unconscious; I may take my umbrella without thinking consciously about it. But, when I express the thought verbally, by saying that it's raining, the thought is always conscious. Why is that? Simply positing a necessary connection between consciousness and speech, as some traditional thinkers have (see, e.g., Descartes, *Letters to Newcastle and to More*, 1645 & 1649), amounts to giving no explanation at all.

Whenever I say it's raining, I could as easily have said that I think it's raining. The two statements do not, of course, mean the same; one is about the weather and the other is about my mind. But they do have virtually the same conditions of use.<sup>28</sup> Even more important for our pur-

---

construction (Moore 1942, p. 543; 1944, p. 204). Following Wittgenstein (1953, 2, §10), this type of construction is known as Moore's paradox. See also Rosenthal (1995).

<sup>26</sup> For more on this, see Rosenthal (1993).

<sup>27</sup> One might urge that nonverbal behavior never expresses the thoughts that underlie such behavior in the same way that speech acts do; I am grateful to Josef Perner for pressing this point, in conversation. But that issue does not affect the present argument.

<sup>28</sup> Perhaps saying that I think it's raining conveys a measure of caution not present when I simply say it's raining, but that difference doesn't affect the present argument.

poses, that equivalence in conditions of use is automatic and second nature for us, so much so that we may not even notice on any given occasion which of the two we said, nor which of the two another person said.<sup>29</sup>

Suppose, then, that I say that it's raining. Because I might as easily have said instead that I think it's raining, I must have had the thought that this other remark would have expressed; otherwise the interchangeability of the two statements wouldn't be fully automatic and second nature. But the thought I would express by saying that I think it's raining would be the HOT that I think it's raining. And, since the automatic interchangeability ensures that I have that HOT, my first-order thought that it's raining is itself a conscious thought. The HOT hypothesis explains why, whenever I say anything at all, the thought I thereby express is conscious.<sup>30</sup>

Things are different when one expresses a thought nonverbally. There is no equivalence in performance conditions between the nonverbal expression of a thought and any expression, verbal or otherwise, of the corresponding HOT. So the nonverbal expressing of a thought gives us no reason to posit an accompanying HOT. The HOT model thus also explains why nonverbally expressed thoughts need not be conscious.<sup>31</sup>

One might object that parapraxis shows that verbally expressed thoughts are not, after all, invariably conscious. Suppose that somebody, intending consciously to declare a meeting open, slips and says 'I declare this meeting closed'.<sup>32</sup> One could interpret this speech performance

---

<sup>29</sup> We can explain the automatic and second-nature character of this performance-conditional equivalence by appeal to a folk theory that is well-entrenched in our ways of thinking about thought and speech. On that folk theory, thoughts are the kind of state that can be expressed in speech; so anybody who says that it's raining will have the thought that it's raining. Our operating on that tacit assumption leads to our regarding as indifferent whether one says that it's raining or that one thinks it's raining.

<sup>30</sup> There is an exception. Saying that I think it's raining verbally expresses a HOT that I think it's raining but, as noted earlier, that HOT is seldom conscious. But the HOT model also explains this exception, since the higher-order case differs in a crucial way from the first-order case. Though the statement that I think it's raining is, again, performance conditionally equivalent to the statement that I think that I think that it's raining, in this case the performance-conditional equivalence is hardly second nature. So we needn't suppose that any third-order thought occurs.

<sup>31</sup> For more on this, see Rosenthal (Why Are Verbally Expressed Thoughts Conscious?, forthcoming).

<sup>32</sup> Freud's well-known example in *The Psychopathology of Everyday Life*

as expressing the person's unconscious intention to close the meeting. And perhaps, then, that the speech act expresses a thought that isn't, after all, conscious.

But that construal doesn't do the situation full justice. The person also has some conscious intention to open the meeting; otherwise why was any remark made at all? And, if we see the remark as expressing the unconscious intention not to open the meeting, the connection between the conscious intention and the making of the remark remains obscure. We would do better to see the remark as verbally expressing the conscious intention, and regard the unconscious intention not as expressed by the remark but as interfering with its production, by causing the injection of a stray word. Parapraxis is an interaction effect in which an unconscious state affects the expressing of a conscious state; it is not a counterexample to the regularity that verbally expressed cognitive states are invariably conscious.<sup>33</sup>

Things are strikingly different with affective states, such as joy, anger, and sadness. One might verbally express anger, for example, by asserting moral disapprobation or some other negative evaluation. And in that case one may well also adopt a kind of emotional distance from the situation, even to the point of remaining unaware of one's anger. This can happen because, unlike the case of cognitive intentional states, there is no standard way of verbally expressing one's affective states. So there is no well-entrenched equivalence in conditions of use between one's verbally expressing an affective state and one's describing oneself as being in that state, and hence no reason to posit a HOT whenever one verbally expresses an affective state. An affective state's being verbally expressed is not by itself sufficient for that state to be conscious.<sup>34</sup>

As with all other mental states, however, being able to report an affective state noninferentially is enough for that state to be conscious, since that ability ensures the occurrence of the HOT that an actual report would express. And a noninferential report expresses one's consciousness of oneself as being in the state in question.

---

(Freud 1901b, p. 59-60), in which the President of the Lower House of the Austrian Parliament's said "I . . . declare the sitting *closed*," rather than open. See also Freud (1940b) and elsewhere.

<sup>33</sup> It might also be that both intentions occur consciously, though for social reasons the speaker would resist acknowledging the intention not to open the meeting. In such a case, even if the remark expressed the intention not to open the meeting, it would still express a conscious intention.

<sup>34</sup> On the special case of affective states, see Rosenthal (1998).

A mental state's being conscious consists in one's being conscious of that state in some suitable way. Inner-sense theorists typically construe our being thus conscious of our mental states as the result of some mechanism by which the mind monitors itself (see, e.g., Armstrong 1980; Lycan 1996). Mechanisms can malfunction, and a monitoring mechanism could deliver mistaken results; so the existence of some such mechanism would square with the observation that consciousness is neither infallible nor exhaustive. But consciousness doesn't simply go wrong; it sometimes actually confabulates. And that suggests that self-interpretation plays some role in whatever process leads to our being noninferentially conscious of our mental states. This, once again, points to HOTs rather than inner sense.

Even independent of confabulation, it is inviting to see noninferential reports that one is in some mental state as selfinterpretations; one interprets oneself as being in the mental state in question. Others sometimes offer interpretations of what mental state one is in; but, as noted earlier, accepting another's interpretation does not by itself lead to the state's being conscious. Accepting another's interpretation of one's mental state is coming to think that one is in that state. But that thought will remain inferential as long as it rests on one's accepting the other person's interpretation.

But if one comes to identify with that other person's interpretation sufficiently closely to make that interpretation one's own, the resulting self-interpretation may well become noninferential. One will then be conscious of oneself as being in the state, but it will no longer seem to one as though that awareness of the state relies on any conscious inference or other mediating process. One will seem subjectively to be directly aware of the state, and so that state will have come to be conscious. As mentioned earlier, one need not also cease being aware of one's having previously relied on the other person's interpretation; it's enough that one now takes oneself to be in the relevant state independently of any such inference.

What happens here resembles the process by which all verbally expressed thoughts come to be conscious. In both cases, the performance of a speech act leads automatically and without conscious processing to one's having a HOT that one is in a particular mental state, and so to one's being conscious of oneself as being in that state. The ascription by the other person results in one's having a HOT because one identifies with that person and, hence, with the ascription. One's verbally expressing a thought leads to one's having a HOT about it because of the well-

entrenched equivalence between saying something and saying that one thinks that thing. And we can see that automatic equivalence, in turn, as itself a matter of identifying with one's own speech act.<sup>35</sup>

Having noted that we often infer what mental states others are in, Freud writes that “[p]sycho-analysis demands nothing more than that we should apply this process of inference to ourselves also” (Freud 1915e, p. 169). But, if that process of inference about ourselves is to lead to the inferred mental states’ being conscious, it must come to seem as though that inference no longer figures in our being conscious of ourselves as being in those states.<sup>36</sup>

## REFERENCES

- Aristotle De Anima. In *Aristotle's De Anima. Books II and III*. Oxford: Clarendon Press 1993.
- Armstrong, D.M. (1980). What is consciousness. In *The Nature of Mind*. St. Lucia, Queensland: University of Queensland Press, pp. 55–67.
- Baars, B.J. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- Baars, B.J. (1997). *In the Theater of Consciousness: The Workspace of the Mind*. New York: Oxford University Press.
- Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy X*: 615–678.
- Block, N. (1995). On a confusion about a function of consciousness. *The Behavioral and Brain Sciences* 18(2): 227–247.
- Brentano, F. (1884). *Psychology from an Empirical Standpoint*. London: Routledge & Kegan Paul 1973.
- Broad, C.D. (1925). *The Mind and its Place in Nature*. London: Routledge and Kegan Paul.
- Carruthers, P. (2000). *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.
- Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown and Company.

---

<sup>35</sup> One identifies with one's own speech act by way of the tacit, well-entrenched folk theory on which speech acts all express intentional states. On that folk theory, see n. 29, above. One's nonconscious thought that one thinks the thing one is saying constitutes, in effect, an identification of oneself with the thought one takes oneself to have.

<sup>36</sup> A first version of this paper was presented at the 3rd International Symposium of the Austrian Academy of Sciences “Psychoanalysis as an Empirical, Interdisciplinary Science. On Continental & Anglo-American Research”, Vienna, November 22, 2002. I am very grateful to that audience for useful reactions.

- Dennett, D. (1993). The message is: There is no medium. *Philosophy and Phenomenological Research* 53(4): 919–931.
- Descartes, R. (1641) Geometrical Exposition of the Second Replies. *Oeuvres de Descartes* 7, ed. C. Adam & P. Tannery. Paris: J. Vrin 1964.
- Descartes, R. (1645) Letter to Newcastle. *Oeuvres de Descartes* 4, ed. C. Adam & P. Tannery. Paris: J. Vrin 1964, pp. 573–576.
- Descartes, R. (1649) Letter to More. *Oeuvres de Descartes* 5, ed. C. Adam & P. Tannery. Paris: J. Vrin 1964, pp. 275–279.
- Fodor, J.A. (1992). A theory of the child's theory of mind. *Cognition* 44(3): 283–296.
- Freud, S. (1900a). *The Interpretation of Dreams*. S. E. 4–5.
- Freud, S. (1901b). *The Psychopathology of Everyday Life*. S. E. 6: 1–289.
- Freud, S. (1915e). *The Unconscious*. S. E. 14: 166–215.
- Freud, S. (1923b). *The Ego and the Id*. S. E. 19: 3–68.
- Freud, S. (1940a). *An Outline of Psychoanalysis*. S. E. 23: 141–207.
- Freud, S. (1940b). *Some elementary lessons in psycho-analysis*. S. E. 23: 279–86.
- Gennaro, R.J. (1995). *Consciousness and Self-Consciousness*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Greenwald, A.G. & Banaji, M.R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review* 102(1): 4–27.
- Harman, G. (1973). *Thought*. Princeton: Princeton University Press.
- Harman, G. (1999). *Reasoning, Meaning and Mind*. Oxford: Clarendon Press.
- Holmes, D.S. & Frost, R.O. (1976). Effect of false autonomic feedback on self-reported anxiety, pain perception, and pulse rate. *Behavior Therapy* 7(3): 330–334.
- Kant, I. (1787). *Critique of Pure Reason*. Cambridge: Cambridge University Press 1998.
- Kriegel, U. (2002). Consciousness, permanent self-awareness, and higher-order monitoring. *Dialogue* 41(3): 517–540.
- Kriegel, U. (2003). Consciousness, higher-order content, and the individuation of vehicles. *Synthese* 134: 477–504.
- Locke, J. (1700). *An Essay Concerning Human Understanding*. Oxford: Oxford University Press 1975.
- Lycan, W.G. (1996). *Consciousness and Experience*. Cambridge, Mass.: MIT Press/Bradford Books, pp. 13–43.
- Marcel, A.J. (1983a). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology* 15: 197–237.
- Marcel, A.J. (1983b). Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology* 15: 238–300.
- Merikle, P.M., Smilek, D. & Eastwood, J.D. (2001). Perception without awareness: Perspectives from cognitive psychology. *Cognition* 79(1–2): 115–134.
- Moore, G.E. (1942). A reply to my critics. In *The Philosophy of G. E. Moore*, ed. P.A. Schilpp. New York: Tudor Publishing Company, pp. 533–677.
- Moore, G.E. (1944). Russell's 'theory of descriptions'. In *The Philosophy of Bertrand Russell*, ed. P.A. Schilpp. New York: Tudor Publishing Company, pp. 175–226.

- Nisbett, R.E. & DeCamp Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84(3): 231–259.
- Perner, J., Leekam, S.R. & Wimmer, H. (1987). Three-year olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology* 5(2): 125–137.
- Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, Mass.: MIT Press/Bradford Books.
- Perner, J. (1995). The many faces of belief: Reflections on Fodor's and the child's theory of mind. *Cognition* 57(3): 241–269.
- Rosenthal, D.M. (1986). Two concepts of consciousness. *Philosophical Studies* 49(3): 329–359.
- Rosenthal, D.M. (1993). Thinking that one thinks. In *Consciousness: Psychological and Philosophical Essays*, ed. M. Davies and G.W. Humphreys. Oxford: Basil Blackwell, pp. 197–223.
- Rosenthal, D.M. (1995). Moore's paradox and consciousness. *Philosophical Perspectives* 9: 313–333.
- Rosenthal, D.M. (1997a). Perceptual and cognitive models of consciousness. *Journal of the American Psychoanalytic Association* 45(3): 740–746.
- Rosenthal, D.M. (1997b). A theory of consciousness. In *The Nature of Consciousness: Philosophical Debates*, ed. N. Block, O. Flanagan & G. Güzeldere. Cambridge, Mass.: MIT Press, pp. 729–753.
- Rosenthal, D.M. (1998). Consciousness and its expression. *Midwest Studies in Philosophy* 22: 294–309.
- Rosenthal, D.M. (1999a). The colors and shapes of visual experiences. In *Consciousness and Intentionality: Models and Modalities of Attribution*, ed. D. Fi-sette. Dordrecht: Kluwer Academic Publishers, pp. 95–118.
- Rosenthal, D.M. (1999b). Sensory quality and the relocation story. *Philosophical Topics* 26(1–2): 321–350.
- Rosenthal, D.M. (2002). Explaining consciousness. In *Philosophy of Mind: Classical and Contemporary Readings*, ed. David J. Chalmers. New York: Oxford University Press, pp. 406–421.
- Rosenthal, D.M. (2003). Unity of consciousness and the self. *Proceedings of the Aristotelian Society* 103(3): 325–352.
- Rosenthal, D.M. (2005). Why are verbally expressed thoughts conscious? In *Consciousness and Mind*, ed. D.M. Rosenthal. Oxford: Clarendon Press, in press.
- Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson and Company.
- Solms, M. (1997a). What is consciousness? *Journal of the American Psychoanalytic Association* 45(3): 681–703.
- Solms, M. (1997b). Response. *Journal of the American Psychoanalytic Association* 45(3): 765–778.
- Staats, P.S., Hekmat, H. & Staats, A.W. (1998). Suggestion/placebo effects on pain: Negative as well as positive. *Journal of Pain and Symptom Management* 15(4): 235–243.
- Weiskrantz, L. (1986). *Blindsight: A Case Study and Implications*. Oxford: Clarendon Press.

- Weiskrantz, L. (1997). *Consciousness Lost and Found: A Neuropsychological Exploration*. Oxford: Clarendon Press.
- Wellman, H.M. (1990). *A Child's Theory of Mind*. Cambridge, Mass.: MIT Press/Bradford Books.
- White, P.A. (1988). Knowing more than we can tell: 'Introspective access' and causal report accuracy 10 years later. *British Journal of Psychology* 79(1): 13–45.
- Wilson, T.D., Hodges, S.D. & LaFleur, S.J. (1995). Effects of introspecting about reasons: Inferring attitudes from accessible thoughts. *Journal of Personality and Social Psychology* 69(1): 16–28.
- Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13(1): 103–128.
- Wittgenstein, L. (1953). *Philosophical Investigations*, ed. G. E. M. Anscombe. New York: Macmillan.