

Naturalizing Subjective Character

URIAH KRIEGEL

University of Arizona and University of Sydney

When I have a conscious experience of the sky, there is a bluish way it is like for me to have that experience. We may distinguish two aspects of this “bluish way it is like for me”: (i) the bluish aspect and (ii) the for-me aspect. Let us call the bluish aspect of the experience its *qualitative character* and the for-me aspect its *subjective character*. What is this elusive for-me-ness, or subjective character, of conscious experience? In this paper, I examine six different attempts to account for subjective character in terms of the functional and representational properties of conscious experiences. After arguing against the first five, I defend the sixth.

There is something at least *prima facie* mysterious about conscious experience. The problem of consciousness is the problem of *demystifying* whatever it is that accounts for the *prima facie* mysteriousness of conscious experience. This would involve showing that the *prima facie* mysterious aspects of conscious experience are not super-natural phenomena. That is, it would require “naturalizing” the relevant aspects of conscious experience, presumably by showing how they could exist in a purely physical world.¹

It is useful to refer to the *prima facie* mysterious element in conscious experience in terms of *what it is like for the subject* to have or undergo a conscious experience.² When I have a conscious experience of the sky, there is a bluish way it is like for me to have or undergo my experience.³ I suggest that we distinguish two aspects in this “bluish way it is like for me”: (i) the *bluish* aspect, which we may call the experience’s *qualitative character*, and (ii) the *for-me* aspect, which we may call its *subjective character*. Not only *is* the experience bluish, but I am also *aware* of its being bluish. Its *being*

¹ It is not altogether clear what would count as successful naturalization of a phenomenon. Here I am assuming that showing a certain phenomenon to be physical is a sufficient condition for naturalizing it. This does not settle all the questions, of course, as it is not much clearer what the general criterion for physicality is. But in this paper I will rest content with an intuitive understanding of the project of naturalization, trusting that we have a fairly solid intuitive notion of what is natural and what would be super-natural.

² I am borrowing this phrase from Nagel (1974), of course. In using this phrase I do not commit, however, to any of Nagel’s claims about it.

³ I use the adjectives “bluish” to denote the property of color experiences and “blue” to denote the property of external objects and surfaces, that is, worldly color itself (and similarly for other color terms). In this I follow Levine’s (2001) useful practice.

bluish constitutes its qualitative character, while my *awareness* of it constitutes its subjective character.⁴

A demystification of this “bluish way it is like for me” would thus be a two-front endeavor. It would require the naturalization both of the qualitative character and of the subjective character of conscious experience.⁵ The philosophical literature on consciousness focuses mainly on experience’s qualitative character.⁶ Thus, functionalist and representationalist attempts to naturalize qualitative character are debated routinely.⁷ Discussion of the subjective character of experience, and of the prospects for its naturalization along functionalist and/or representationalist lines, is much more limited.

That is precisely what I want to discuss in the present paper: the prospects for naturalizing subjective character, the “for-me” aspect of experience. I will consider six possible accounts of it, each more promising than its predecessor. In §2, I consider the least viable approach to naturalizing subjective character, which I call *conceptualist*. In §3, I consider two functionalist accounts of naturalizing subjective character, and in §4 two representationalist accounts; all four will be shown to suffer from rather principled difficulties.

⁴ We may use the term “phenomenal character” to signal the compresence of qualitative character and subjective character.

⁵ It is possible to maintain, though, that the naturalization of one of these aspects is sufficient for the demystification of consciousness. Thus, Rosenthal and Lycan argue that qualitative character is quite accidental to conscious experience, and therefore need not be naturalized for consciousness *per se* to be demystified—see Rosenthal 1991, Lycan 1996 chapter 4. Levine (2001, chapter 4) argues that they are wrong, and that the qualitative and subjective characters of experience are in fact inseparable, but then goes on to claim that it is subjective character which is the source of the problem of consciousness. By contrast, Block (1995) may be interpreted as claiming that the naturalization of qualitative character is all important for the project of demystifying consciousness, whereas the naturalization of subjective character is immaterial. Elsewhere, I take the side of Rosenthal and Lycan in this debate (see Kriegel 2003c), but I will not argue for it here in much detail. On my view, the qualitative character of a conscious experience determines *which* conscious experience it is, but it is the experience’s subjective character that determines *whether* it is a conscious experience at all. On this view, then, it is the subjective character of a conscious experience that accounts for it *being* a conscious experience, while its qualitative character only determines *which* conscious experience it specifically is. Once we accept this view, it follows quite straightforwardly that the crux of the problem of consciousness is subjective character. For what is puzzling about consciousness is primarily why we have conscious experiences *at all*, not why we have the specific conscious experiences we do.

⁶ Mainly, but not exclusively. Higher-order monitoring theorists of consciousness, such as Rosenthal and Lycan, seem to target subjective character. This follows naturally from their view (which we mentioned in the previous footnote) that the qualitative character of conscious experience is accidental to it.

⁷ The representationalist approach to qualitative character is the more popular nowadays—see, e.g., Harman (1990); Dretske (1995); Tye (1995, 2000)—probably because of the dead-end into which functionalist accounts ran, in attempt to accommodate inverted and absent qualia. But in my view, the representationalist approach to qualitative character ultimately collapses into the functionalist approach (see Kriegel 2002a).

In §5, I advance a different proposal, which I argue to be superior to the functionalist and representationalist accounts previously considered. But I start, in §1, with a more detailed discussion of the phenomenological nature of subjective character.

1. Subjective Character and its Psychological Reality

If there is something that makes a conscious experience “for me,” then by having the experience, I must be somehow *aware* of having it. For if I am wholly unaware of my experience, there is no sense in which it could be said to be “for me.”

The awareness in question is quite special, however. On the one hand, it must be conceded that we are aware of our conscious experiences. For conscious experiences are not sub-personal states which simply happen *in* us, without our being aware of them. A mental state we are completely unaware of is an unconscious state. On the other hand, the awareness we have of our conscious experiences is, normally, not a very focused awareness. When I have a conscious experience of the sky, I am focused on the sky, not on my experience of the sky. It seems, then, that we do not have a very focused or attentive awareness of our conscious experiences, yet we certainly have *some* form of awareness of them.

Our awareness of our conscious experiences is a fairly elusive phenomenon, then. To describe this awareness more precisely, let us introduce a distinction between *focused* awareness and *peripheral* awareness. The distinction is easy to draw for the case of straightforward visual awareness. In my current visual perception, there is a distinction to be made between my *foveal* vision of my laptop before me and my peripheral vision of a myriad of objects surrounding the laptop (a stack of books on the left corner of the desk, a bonsai on the right, etc.). My awareness of the object I am foveating (i.e., my laptop) is focused and attentive in a way my awareness of the other objects (e.g., the books) is not. What I want to emphasize is that the fact that I do not have a focused and attentive awareness of the books and the bonsai should not mislead us to say that I am wholly unaware of them. Rather, we should say that I have a sort of peripheral awareness of them.

This peripheral awareness takes a visual form in the case of the books and the bonsai, but it can take a different form in other cases. Thus, another peripheral component of my current experience may be the sound of car engines passing outside my window. Peripheral awareness may even take a non-perceptual form in some cases.⁸

⁸ Thus, James’s (1890) notion of the *fringe of consciousness* (which is widely used by cognitive psychologists) is supposed to cover non-perceptual phenomena. Similarly for the notion of *marginal consciousness* commonly discussed in the phenomenological literature (see especially Gurwitsch 1985). For a recent comprehensive discussion of the non-perceptual elements in the fringe of consciousness (such as feeling-of-knowing phenomena,

My suggestion is that our awareness of our conscious experiences is normally a form of such non-perceptual peripheral awareness. Conscious experiences exhibit a subjective character in virtue of the fact that whenever we have a conscious experience, we are peripherally aware of having it. Thus, if my visual perception of the laptop is a conscious one,⁹ then in having it I am peripherally aware not only of the stack of books on the left and the bonsai on the right, but also of the fact that I am having that very experience.

An objector could claim that there is no such peripheral awareness of our conscious experiences.¹⁰ To my mind, it is a little odd to claim that con-

tip-of-the-tongue phenomena, and ‘rightness’ phenomena), see Mangan (2001). The notion of inner awareness is not claimed here to be a simple extension from the notion of peripheral visual awareness. The latter admits of simple extension, for instance, to peripheral awareness in other perceptual modalities. Thus, while foveating my laptop, I may have an *auditory* peripheral awareness of the cars passing by on the street. The notion of inner awareness as a peripheral awareness of one’s current conscious experience is importantly different from these forms of peripheral awareness, not least in being more intellectual than perceptual. But the existence of non-perceptual peripheral awareness has already been variously admitted, among others by the authors cited above.

⁹ That is, if it is not a subliminal perception, or a blindsighted perception, or a perception that is otherwise unconscious...

¹⁰ One way the objection could play out is *epistemologically*. The idea is to claim that the notion that we have a peripheral awareness of all our conscious experiences leads to an overly Cartesian conception of self-knowledge. Let me make a few remarks by way of trying to alleviate this worry.

The notion of inner awareness has had the historical misfortune of being associated with the Cartesian doctrines of infallibility and incorrigibility. In its strongest version, the Cartesian picture holds that at any moment of its waking life, the mind has perfect knowledge of all the mental states it is in and all the mental states it is not in, and is in this sense *transparent* to itself. More weakly, it has sometimes been held that a person always has knowledge of all the mental states she is in, although she may not have knowledge of the mental states she is not in. Yet more weakly, the view has been offered as applying not to *all* mental states, but only to *conscious* mental states: a person always has accurate knowledge of the *conscious* mental states she is in. Let us call this the *doctrine of weak transparency*.

Now, knowledge involves (at least) justified true belief. So according to the doctrine of weak transparency, whenever we have a conscious experience, we also have a justified true belief about it. Let us call the *doctrine of weak infallibility* the thesis that whenever we have a conscious experience, we also have a true belief about it; and let us call the *doctrine of weak incorrigibility* the thesis that whenever we have a conscious experience, we also have a justified belief about it. The doctrine of weak transparency amounts, more or less, to the conjunction of the doctrines of weak infallibility and incorrigibility.

What I want to point out is that the notion of inner awareness does not, by itself, commit us to any of these. At most it may commit us to the following thesis: whenever we have a conscious experience, we also have a belief about it. This thesis does not guarantee that the relevant belief be true or justified. It only guarantees its occurrence. In fact, the notion of inner awareness may commit us to even less than that. Strictly speaking, what it commits us to is this: whenever we have a conscious experience, we also have an *awareness* of it. This is weaker than the previous thesis, because it guarantees only the occurrence of an awareness, not the occurrence of a belief. And belief is doubly stronger than awareness: (i) belief always has propositional content, whereas awareness

conscious experiences may take place when their subject is completely unaware of them. Although we are unaware of many of our mental states, it is something of a conceptual truth that only the mental states we are aware of count as conscious.

Our objector may insist, however, that she finds no such awareness in her phenomenology. Unfortunately, I cannot offer here a positive demonstration of the existence of such awareness, and hence of subjective character. Arguably, such a demonstration is simply unavailable.¹¹ (Observe, though, that the kind of evolutionary/functional advantages commonly attributed to peripheral awareness in general speak in favor of the existence of peripheral awareness of one's conscious experiences.¹²) But even so, there are methodological reasons not to ignore the phenomenon of subjective character. For we certainly *appear* to be aware of our conscious experiences, and to that extent conscious experiences *appear* to have a subjective character. It is thus a constraint on the adequacy of any account of conscious experience that it *explain* this appearance, if only by explaining it *away* as a predictable illusion. And this latter project already requires that we take the for-me-ness of conscious experiences seriously, if only as an appearance. Therefore, delving into the topic

may have non-propositional content; (ii) belief involves a certain *endorsement*, or *affirmation*, of its content, which may be lacking from mere awareness.

The notion of inner awareness commits us, then, to much less than the false promises of Cartesian transparency. In general, pointing out the existence of a psychological phenomenon is not the same as claiming epistemic and semantic privileges on its behalf. The phenomenon of inner awareness is psychologically real, whether or not it enjoys any special epistemic status.

¹¹ How would one go about offering an argument for the *existence* of anything? Suppose someone goes eliminativist with respect to my laptop, and professes to maintain that my laptop does not exist. How can I demonstrate the existence of my laptop? The best I can do is show this person my laptop. If she is blind, or claims to be blind, however, I will find myself in a tight spot. I can let her touch or hear my laptop, but she may claim that the feelings and sounds are very weak evidence for the existence of a portable computing device. The way I see it, the burden of argument that lies on the realist with respect to the existence of my laptop is mainly to show the arguments for the non-existence of the laptop to be weak. That is, the realist is only obligated to making the case for the existence of my laptop *via negativa*. A positive demonstration of its existence may be out of question. Likewise for inner awareness.

¹² Thus, according to Mangan (2001), the function of the 'fringe of consciousness' is to augment the availability of some contents. Our capacities for attentive, focused consciousness are limited, and in order not to overburden consciousness and fill it with 'noise', large parts of the contents a person perceives or thinks must remain outside the focus of consciousness. But some contents are important enough not to be excluded from the sphere of consciousness altogether, where their availability for quick and effortless retrieval will be dramatically diminished. These contents are kept in the fringe of consciousness, that is, they become contents of peripheral awareness. The point to take account of is that, if there is an important advantage in keeping information about the peripheries of one's field of vision in the fringe of consciousness, there is certainly at least as much advantage in keeping there information about one's current states of mind as well.

of subjective character is inevitable, even from the perspective of those who would deny its existence.

Now, it is surely *preferable* to explain a phenomenon within one's theoretical framework than it is to explain it *away*. We should opt for the latter strategy only upon losing confidence in the former. It is therefore incumbent upon us to at least *start* by searching for a positive account of subjective character.

In trying to account for the special awareness we have of our conscious experiences, we should note that the awareness is special not only in being peripheral rather than focused, but also in that it does not seem to constitute a *separate* mental event, independent of the experience of which it is an awareness. Our awareness of our conscious experiences does not seem to require an extra mental act, through which we *become* aware of a previously existing experience. Rather, it seems to be somehow built into that very experience. That is, the experience seems to include within it an awareness of itself.

It is possible to claim (and certainly has been claimed) that, appearances to the contrary, our awareness of our conscious experiences *is* grounded in a separate, numerically distinct mental event or state. The suggestion, which we find in the so-called *Higher-Order Monitoring* approach to consciousness, is that the for-me-ness of conscious experience consists in the occurrence of a numerically distinct (metaphysically independent) representation of the experience.¹³ This suggestion faces a number of difficulties, into which we cannot go here.¹⁴ The main problem can be summarily put as the following dilemma.¹⁵ If my awareness of my experience of the sky was anchored in a numerically distinct representation, that representation would have to be either conscious or unconscious. But an *unconscious* representation would not do, because unconscious representation does not qualify as awareness (in that it is not phenomenologically manifest),¹⁶ and a *conscious* representation would not do either, because that would lead to infinite regress.¹⁷

¹³ This would be to apply a higher-order monitoring approach to the for-me-ness of conscious experience. The Higher-Order Monitoring approach to consciousness is quite popular nowadays—see Armstrong (1968, 1981); Rosenthal (1986, 1990, 2003); Lycan (1990, 1996, 2001); Carruthers (1989, 1996), among others. It is probably meant by its proponent as an account of subjective character.

¹⁴ The higher-order monitoring theory suffers from innumerable difficulties, which we cannot go into here. For criticism, see Rey (1988); Aquila (1990); Dretske (1993, 1995); Goldman (1993); Natsoulas (1993); Block (1995); Guzeldere (1995); Byrne (1997); Neander (1998); Carruthers (2000); Moran (2001); and Kriegel (2002b, 2003a, 2003b).

¹⁵ For a more detailed and more careful presentation of the dilemma argument I present in the text, see Kriegel (2003b).

¹⁶ This point can be made by appeal to an example or by appeal to a more principled consideration. Example: to be *aware* of a table, one must harbor a *conscious* representation of it. The principled consideration is that we have no first-person acquaintance with our unconscious mental states, but we do have first-person acquaintance with the for-me-

The awareness of our conscious experiences can be grounded neither in a numerically distinct unconscious state nor in a numerically distinct conscious state. Therefore, it cannot be grounded in a numerically distinct state. It must be internal to the conscious experience itself. We may signal this special sort of awareness—awareness that is internal to what one is thereby aware *of*—by labeling it *inner awareness*. Joseph Levine describes the challenge presented by inner awareness as follows (2001, p. 168):

[There is a] paradoxical duality of [conscious] experiences: there is an awareness relation, which ought to entail that there are two states serving as the relevant relata, yet experience doesn't seem to admit of this sort of bifurcation. Let's call this the problem of "duality."

That is, the inner awareness built into a conscious experience is paradoxical, inasmuch as the experience exhibits a duality as both the awareness and what one is thereby aware *of*. It is the *act* of awareness and the *object* of awareness all at once.¹⁸ It is presumably for this reason that conscious experiences are sometimes said to be *self-presenting*.

The subjective character, or for-me-ness, of an experience consists, then, in the inner awareness the experience involves. Now, awareness of something involves intentional directedness towards it. So for a mental state to have inner awareness built into it, it must be intentionally directed at itself. When a mental state is intentionally directed at itself, it is both intentional act and intentional object and thus has the sort of dual nature that Levine seeks in conscious experience.¹⁹

ness of conscious experience. So the for-me-ness of conscious experience cannot consist in an unconscious mental state.

¹⁷ If my awareness of my conscious experience was conscious, it would have a subjective character, which would mean that I would be aware of my awareness of my conscious experience, but this awareness of my awareness of my conscious experience would also have to be conscious (lest we fall back to the first horn of the dilemma), and so would have a subjective character, which means I would be aware of my awareness of my awareness of my conscious experience—and so on *ad infinitum*.

¹⁸ This is not to say that it is the *only* object of awareness. When I have a conscious experience of the sky, the experience anchors an awareness both of the sky and of itself. Indeed, it is difficult to imagine what it would be like to have an experience which is its only object. Perhaps this is the state of mind the Buddhist seeks in her meditative practices.

¹⁹ This dual nature as act and object may explain the strange appeal of *sense data*. Sense datum theories (see, e.g., Russell (1922); and Ryle (1949) for arguments against them), as well as other forms of "representative realism" (or "indirect realism"), often construe sense data (or whatever inner objects they posit) as a third entity, which is distinct from both the intentional act and its worldly object, and is interposed between them. This third entity is the direct object of the intentional act, but it also represents the worldly (indirect) object. Thus it performs two functions, being both a representation and an entity represented. Sense data, conceived along these 'third entity' lines, are often dismissed as ontologically suspicious—and justifiably so. However, much of the picture offered by the sense datum theorist can be preserved if we construe a conscious experience as intentionally directed at itself, such that it is both act and object of awareness. For then the

This view—that conscious experiences are intentionally directed at themselves—goes back at least to Brentano (1874/1973, pp. 153-4).²⁰

[Every conscious act] includes within it a consciousness of itself. Therefore, every [conscious] act, no matter how simple, has a double object, a primary and a secondary object. The simplest act, for example the act of hearing, has as its primary object the sound, and for its secondary object, itself, the mental phenomenon in which the sound is heard.

Through Brentano's influence, this view of consciousness had become commonplace in the phenomenological tradition. Different versions of the view have been developed and defended by Husserl (1928/1964), Sartre (1937, 1943), Henrich (1966/1982), Brough (1972), Sokolowski (1974), Gurwitsch (1985), Smith (1986, 1989 ch. 2), Frank (1995), Natsoulas (1996, 1999), Wider (1997), and Zahavi (1998, 1999).²¹

The problem of subjective character is the problem of understanding how a mental state may include within it an awareness of itself. How can a mental state be both a bluish experience and an awareness of that very bluish experience? More specifically, how can we make sense of this possibility within a purely naturalistic, or physicalistic, conception of the mind? We may call the latter problem *the problem of naturalizing subjective character*. It is this problem that I will address in the remainder of this paper.

2. The Conceptualist Approach to Subjective Character

In the passage quoted above, Levine formulates the worry about inner awareness in terms of the duality of conscious experiences as both acts and objects of awareness. The problem, he says, is that the awareness relation “ought to entail” two separate relata, the act and the object, but inner awareness disallows such bifurcation. Now, it is not quite correct to say that the existence of an awareness relation ought to entail the existence of two distinct items as its relata. After all, some relations are reflexive (e.g., being at-least-as-tall-as), in the sense that every entity bears them to itself, and some are a-reflexive, in the sense that not every entity *fails* to bear them to itself (e.g., being the-

experience fulfils the double function of a sense datum, representing the worldly object while being itself represented. The only element in the picture of the sense datum theorist that is not preserved here is the ontological commitment to a third entity distinct from both the intentional act and the worldly object.

²⁰ According to Caston (2002), however, this view can already be found in Aristotle, and Brentano likely inherited it from there. Thus, in *Metaphysics* 12.9, Aristotle writes that conscious “knowing, perceiving, believing, and thinking are always of something else, but of themselves on the side” (1074b35-6). In Brentano, the view is first introduced in Section 7 of chapter II (“Inner Consciousness”) in Book 2 of Brentano (1874/1973), which is entitled “A Presentation and the Presentation of that Presentation are Given in One and the Same Act.”

²¹ Recent proponents of the view in the analytic tradition include Levine (2001), Caston (2002), and myself (Kriegel 2002b, 2003a, 2003b, 2006).

accountant-of).²² Naturalizing subjective character would require showing that the awareness relation, and more specifically the intentionality that underlies it, is an a-reflexive relation, a relation in which a mental state *can* stand to itself.²³ The problem, however, is that it is unclear how naturalist accounts of intentionality can allow for that.

Levine distinguishes two general strategies for securing the inner awareness of conscious experiences. We may call them the *conceptual-relation strategy* and the *real-relation strategy*. In the next few sections, I will discuss the real-relation strategy. In the present section, I want to argue against the conceptual-relation strategy.

The idea behind the conceptual-relation strategy is to take the experience and the awareness of it and *stipulate* that they are to count as two parts of a single state, rather than as two distinct states. The conceptual-relation strategy has actually been pursued in the consciousness literature. It has been developed and defended, for instance, by Gennaro (1996, p. 23):

We can understand consciousness as involving the property of ‘accompanied by a MET [meta-psychological thought]’... But we might individuate conscious states ‘widely,’ i.e., in a way that treats consciousness as an intrinsic property of those states. On this account, the MET is part of the conscious state. I will call it the ‘wide intrinsicity view,’ or WIV.

Gennaro’s view does not deny that the experience and the awareness of the experience are separate. What it denies is that they are separate *states*. It claims that they are rather separate *parts* of a (single) state. According to Gennaro, this does justice to the duality of conscious experiences (1996, pp. 27-8):

The WIV also makes room for the historically influential thesis that conscious mental states are, in some sense, directed at themselves... Some care, however, is needed. Brentano’s view literally says that a conscious mental state CMS is, in part, about CMS. This is not quite what WIV says... Strictly speaking, on the WIV, a CMS is not directed at itself. There is rather an ‘inner directedness’ or ‘inner relationality’ in the CMS... The METs are directed at parts of states of which they are part.

Gennaro’s view seems to consist in attempting to account for the duality of conscious experiences by mandating that they will count as part of a single state the other part of which is the subject’s awareness of her experience.²⁴

²² I follow here the practice of distinguishing a-reflexive relations from anti-reflexive relations, the latter being such that they *never* hold between a thing and itself. Being the-father-of, for instance, is an anti-reflexive relation, since nobody is their own father. By contrast, some people are their own accountants, and so being one’s accountant (or being the-accountant-of) is a-reflexive.

²³ We cannot construe intentionality as reflexive, of course, since some things can be intentionally directed at other things without being intentionally directed at themselves.

²⁴ Gennaro’s view is in fact a little more obscure than this. He holds that the awareness of a conscious experience is an unconscious part of a conscious state. What this means is

There are three problems with the conceptual-relation strategy. The first is noted by Levine (2001, p. 169): “The basic problem with the conceptual constituency strategy is that it substitutes a purely conceptual, or nominal relation for a real one.” If the subjective character of conscious experience is a psychologically real phenomenon, one does not explain it by reshuffling one’s conceptual scheme. The second (not unrelated) problem is that the conceptual-relation strategy fails to capture the duality of conscious experiences. After all, it concedes that the experience and the awareness of it are separate. This allows that each can occur without the other, so there is no genuine inner awareness here.²⁵ The third problem is that whether an experience and the awareness of it are two distinct mental states or two parts of a single mental state is not necessarily up to us. John Williams (1981) has shown that it is possible for a person to believe that *p* and to believe that *q* without believing that *p* & *q*. If so, there is a substantive question involved when we ask whether two contents are parts of the same mental state or constitute two separate states. We cannot just legislate the answer.

According to the real-relation strategy, there is some non-conceptual, real relationship in which conscious experiences stand to themselves that underlies their intentional directedness at themselves. The literature on the matter offers two general approaches to theorizing this real relationship, a functionalist one and a representationalist one.^{26, 27}

Functionalism individuates mental states in terms of their functional role within the subject’s mental life. It appeals crucially, then, to a distinction between a *role* and its *occupant*. The role is an abstraction from the set of all causes and effects of the state, as we shall see in greater detail in the next section. The occupant is ordinarily construed as a neurophysiological state of the brain.²⁸ The neurophysiological state is said to *realize* the mental state in virtue of occupying the functional role distinctive of that mental state.

something I am not altogether clear on. But for the purposes of our present discussion we can proceed without clarifying this aspect of the view.

²⁵ Gennaro may be happy to concede that, but if we are right in construing the subjective character of experience in terms of inner awareness, this would amount to a denial of the existence of a subjective character of experiences.

²⁶ Functionalism is sometimes taken as offering a respectable way of being anti-physicalist. But the aspect of functionalism that is most important to me here is its avoidance of supernatural entities, as well as its compatibility with the physicalist thesis that everything (including every realization of mental properties) is ultimately physical.

²⁷ Although the notion that conscious experiences are intentionally directed at themselves is not new, the project of naturalizing such self-directed intentionality certainly is. Brentano was a dualist, and would therefore oppose the naturalization of subjective character. Similarly for Smith, who argues that intentionality itself is not a material relation but a formal one and thus cannot be naturalized (Smith 1999). I am not going to discuss their arguments here.

²⁸ Again, it is possible to hold that the occupant is a non-physical entity, made of mindstuff. This compatibility of functionalism with dualism is what marked it off from earlier positions on the mind-body problem as it was introduced by Putnam (1967). But what is

The idea behind the functionalist account of subjective character is this. There is a certain functional role characteristic of an experience with bluish qualitative character.²⁹ And there is also a functional role characteristic of an *awareness* of one's bluish experience.³⁰ A neurophysiological state which would occupy both the role characteristic of a bluish experience and the role characteristic of an awareness of a bluish experience would realize a mental state with a dual nature as both experience and awareness-of-experience. Such a state would thereby exhibit inner awareness, and hence subjective character.

The representational theory of mind, or representationalism, individuates mental states in terms of their representational contents. Here the distinction to keep in mind is between the *vehicle* of representation and the *content* of representation, that is, between what does the representing and what is being represented. Materialists hold that both are physical entities: the vehicle is a neurophysiological state of the subject's brain and the content is a feature in the subject's physical environment. Furthermore, the relation of representation holding between them must be a physical relation.³¹

The representationalist account of subjective character is fairly straightforward. Whatever the physical relation underlying representation, a neurophysiological state which would bear that relation to itself would be a physical entity with a dual nature as vehicle and content (act and object), and hence with subjective character. This representationalist account of subjective character will be discussed in greater detail in the section after next.

These two approaches to the naturalization of subjective character are not mutually exclusive.³² In fact, under the so-called Cognitivist Paradigm, we should *expect* an experience with a subjective character to both fit the functional role described above and represent its own occurrence. Under this paradigm, the functional role and representational content of mental states mirror each other, in that mental states that differ in functional role also differ in representational content, and vice versa. There are two versions to this story. One goes by the slogan that the brain is a syntactic engine driving the semantic engine that is the mind (see Haugeland 1978) and holds that the mirroring is due to the fact that the functional role of a mental state *determines* its rep-

important for our present purposes is that it is also possible to hold that the occupant is a physical entity.

²⁹ Thus, a bluish experience tends to produce (cause) a belief that one is in the presence of a blue object, and it tends to be produced (caused) by retinal states that register the relevant light reflectance.

³⁰ Thus, such awareness tends to be produced by bluish percepts and to produce introspective generalizations about one's mental life.

³¹ At least this is the model for mental representation of *concrete objects*. For mental representation of abstract entities, if such there be, certain complications will have to be introduced.

³² This is something Levine (2001, p. 171)—who also distinguishes between functionalist and representationalist approaches to subjective character—notes as well.

representational content. The other reverses this order of explanation and holds that it is in virtue of its representational content that a mental state is recruited to play its functional role (see Dretske 1988). But on both versions functional role and representational content go hand in hand.³³ If so, we should expect subjective states to occupy the sort of functional role characteristic of an experience and the awareness of it, on the one hand, and to have themselves as (part of) their own representational content, on the other. The problem, however, is that it is difficult to see how a neurophysiological state of the brain could do either of these things.

3. Functionalist Accounts of Subjective Character

Functionalist accounts of subjective character can be found in recent work by Shoemaker (1996) and Carruthers (2000, ch. 9). In this section, I will examine the accounts portrayed by Carruthers and Shoemaker, and argue against them.

Recall that the functional role of a mental state is an abstraction from its total set of relations to its causes and effects. Thus, if two mental states M_1 and M_2 bring about the same effect E , but M_1 takes three milliseconds more than M_2 to do so,³⁴ we want to abstract from this psychologically insignificant difference and maintain that the functional roles of M_1 and M_2 are the same (as far as their relation to E is concerned). The question for a functionalist is *how much* to abstract from the concrete causal interactions of mental states. Different functionalist accounts of mental states follow from different abstraction policies.

One popular policy is to abstract from anything that does not bear on the *inferential relations* among mental states. A belief that p tends to bring about a belief that q , when conjoined with a belief that if p then q , and tends to be brought about by a belief that r , when conjoined with a belief that either not- r or p . Not only is it psychologically inconsequential whether it takes half a second or a full second for the belief that p to bring about, and be brought about by, these other beliefs. Anything about it that does not bear on this sort of inferential relation should be abstracted from.

On this view, functional role is *inferential role*. This notion of inferential role has been developed in a rigorous way by Field (1977), according to whom the inferential role of a mental state M is fully characterized by an ordered quadruple of sets $\langle S_1, S_2, S_3, S_4 \rangle$. S_1 is the set of all mental states from which M can be deductively inferred; S_2 is the set of all mental states from which M can be inferred non-deductively (inductively, abductively, by

³³ There are more recent complications in this mirroring conception, emanating from externalism about mental representation (Putnam 1975; Burge 1979), which are not very relevant to my concerns in the present paper. For a thorough discussion of attempts to confront these complications, see Fodor (1987).

³⁴ To take an example from Block (1986).

inference to the best explanation, or the like); S_3 is the set of all mental states which can be deductively inferred from M ; and S_4 is the set of all mental states which can be non-deductively inferred from M . M_1 and M_2 have the same inferential role iff each quadruplet in the inferential role of M_1 is equivalent to the corresponding quadruplet in the inferential role of M_2 .

According to inferential role *semantics*, it is the inferential role of a mental state that determines its intentional content. A belief is the belief that p precisely *because* (in the sense of “in virtue of”) it tends to bring about a belief that q when conjoined with a belief that if p then q and to be brought about by a belief that r when conjoined with the belief that either not- r or p . What it is to *be* a belief that p is to bring about, and be brought about by, just such beliefs in just such circumstances.³⁵

Carruthers (2000, pp. 242-252) appeals to inferential role semantics in an attempt to account for the subjective character of conscious experiences. What he wants to highlight is the fact that, when we have a bluish experience, we are disposed to infer (non-deductively) on its basis not only such first-order thoughts as that we are in the presence of a blue object, but also such second-order thoughts as that we are having a bluish experience.³⁶ That is, it is an element in the inferential role of the bluish experience that it tends to bring about thoughts *about* its occurrence, through which the subject would become aware of it in a focused way. Where Es is the experience of the sky and AEs is an awareness of the experience of the sky, we may say that for Es , $AEs \in S_4$, that is, AEs is a member of the fourth quadruplet characterizing the inferential role of Es (namely, the set of all mental states which can be non-deductively inferred from Es). Against the background of inferential role semantics, this means that AEs is a determinant of the content of Es . And this, Carruthers suggests, is what accounts for the inner awareness (and hence subjective character) implicit in Es itself.

Carruthers’ suggestion is interesting, but it depends on the dubious notion that a mental state *actually* exhibits peripheral inner awareness in virtue of *potentially* bringing about a focused awareness of it.³⁷ This notion we would

³⁵ The program of inferential role semantics comes in two versions, which we may distinguish as *descriptive* and *normative*. For the descriptive version (see Harman 1982), what matters for the inferential role of a mental state is the actual inferences the subject is psychologically inclined to draw from it. For the normative version (see Brandom 1994), it is the inference the subject *ought* to draw from her mental state (that is, the inferences sanctioned by the laws of logic) that determine its inferential role. The version that will interest us here is the descriptive one.

³⁶ This thought is second-order in that its object is itself a mental state—a first-order mental state (i.e., a mental state that does not have another mental state as its object).

³⁷ Recall that inner awareness of our conscious states is awareness that is peripheral, implicit awareness analogous in some respects to the visual awareness we find in peripheral vision. Explicit and focused awareness of our conscious experiences is possible, but it is not the same as the inner awareness we are interested in, because it constitutes a separate mental act.

reject with respect to other kinds of focused awareness the experience can potentially bring about. Thus, my bluish experience of the sky may potentially lead me to believe that the sky has been blue for the past week. The “outer awareness” of the past week is thus a determinant of the content of my bluish experience, according to inferential role semantics. But we would be disinclined to say that this awareness of the past week is somehow implicit in my bluish experience. Nor is this recommended by inferential role semantics. Inferential role semantics suggest determining the content of a mental state M by triangulation from the contents of the states it tends to bring about and be brought about by. It is an extra step—and an implausible one at that—to claim that all these neighboring contents are somehow *implicit in* the content of M .

The suggestion we may extract from Shoemaker (1994, pp. 288-9) is more plausible. The basic idea is that a mental state would qualify as both a bluish experience and an awareness of a bluish experience if it is the realizer, or occupant, of both a functional role characteristic of a bluish experience and a functional role characteristic of an awareness of a bluish experience. Any state that realizes simultaneously both functional roles is a bluish experience which includes within it an awareness of itself.

We can, again, couch the proposal (or rather formulate a similar proposal) in terms of inferential role semantics. Suppose we exhaustively characterize the inferential role of E s (the bluish experience) in terms of the ordered quadruple $\langle S_1, S_2, S_3, S_4 \rangle$ and then the inferential role of AE s (the awareness of the bluish experience) in terms of the ordered quadruple $\langle S^*_1, S^*_2, S^*_3, S^*_4 \rangle$. We can then produce the *intersection* of each pair of corresponding quadruplets. Thus, $S_1 \cap S^*_1$, the intersection of S_1 and S^*_1 , includes all the mental states from which *both* a bluish experience *and* an awareness of a bluish experience can be deductively inferred.³⁸ Similarly for the other quadruplets. The ordered quadruple we thus define, $\langle S_1 \cap S^*_1, S_2 \cap S^*_2, S_3 \cap S^*_3, S_4 \cap S^*_4 \rangle$, captures the inferential role of a mental state that is both a bluish experience and an awareness of a bluish experience.³⁹ It then becomes an empirical question whether any neurophysiological state of the brain occupies this inferential role. If there is such a state, it exhibits a subjective character in a physicalistically kosher way. But in any case there is no *conceptual*, or *principled*, difficulty for naturalism associated with the notion of subjective character.

³⁸ There may not be any—but that is immaterial to the *definition* of the inferential role.

³⁹ A plausible functionalist account could require less than full intersection of the relevant pairs of quadruplets. Thus it could require no more than significant overlap between the functional role characteristic of a bluish experience and the functional role characteristic of awareness of a bluish experience to characterize the functional role distinctive of a conscious experience. That seems to be what Shoemaker (1994) in fact requires—significant overlap rather than full intersection.

This functionalist account of subjective character is more promising than Carruthers', but it suffers from two fundamental drawbacks. One of them I find particularly vexing, but let me start with the other.

The first problem is this. Inferential role semantics is quite plausible when understood as an account of what *underlies* intentional content, but not when understood as an account of what *constitutes* intentional content. That is, perhaps the inferential role of a mental state somehow *underlies*, and even *determines*, the state's intentional content, but it cannot possibly *constitute* that intentional content. The intentional content of a thought about San Francisco is a matter of its relation to *San Francisco itself*, not its relation to other mental states. The thought may *inherit* its intentional content from its different causal/inferential relations to other mental states, but its intentionality *consists in* a relation to the extra-cranial entity that is San Francisco.

The problem is that what seems difficult to understand in the duality of conscious experience is how anything physical can *constitute* it, not so much how anything physical can *underlie* (or determine) it. To dissipate the air of mystery, what would be needed is a physicalist story about how self-directed intentionality can *exist*, not about how it can be *determined*. That is, the most a functionalist account can do is explain what makes it the case that a certain mental state exhibits inner awareness, *given that* there is such a thing as inner awareness. But it cannot explain what makes it the case that there is such a thing as inner awareness. In general, it seems that the dual nature of the experience as both intentional act and intentional object must be explained in *intentional* terms, not in functional ones, since it is an intentional, not functional, feature of it.⁴⁰

The second problem facing the functionalist account under consideration is that it attempts to account for a categorical property in terms of a dispositional one. A mental state's functional role is a merely *dispositional* property of it, whereas the inner awareness present in conscious experience is a *categorical* property. For a token mental state to have a certain functional role, nothing has to actually happen with it. It suffices that the state be of a type

⁴⁰ To see the force of this point, consider by analogy a functionalist "solution" to Moore's paradox (Moore 1942). Moore's paradox is concerned with such propositions as "It is raining and I do not believe that it is raining," which no rational creature could assert or believe. A functionalist might "explain" the paradox as follows. All we need in order to make sense of believing "It is raining, but I do not believe that it raining" is a definition of its distinctive functional role, which can be obtained by intersecting the corresponding pairs of quadruplets in the definition of the functional role of believing "it is raining" and "I do not believe that it is raining." If Moorean propositions cannot be believed, this is simply because no mental state can occupy the thus defined functional role. The reason this "explanation" strikes us as obviously inadequate is that it does not explain *why* the relevant functional role cannot be occupied. It is only by considering the intentional content of beliefs in Moorean propositions that we can appreciate their absurdity.

that *tends* to bring about, and be brought about by, certain other states. But when a mental state becomes conscious, something very real and categorical happens with it. It acquires a feature it did not have before. This feature may well turn out to be the categorical *basis* of a new dispositional property (the relevant functional role), but it is not *identical* to that dispositional property. This problem would apply to any functionalist account of subjective character: functional role is always dispositional, whereas subjective character is occurrent or categorical.⁴¹

Indeed, it appears that the categorical basis of the state's newfound functional role is precisely its duality as both intentional act and intentional object. Once the state comes to be intentionally directed at itself, it acquires certain causal powers, which accordingly alter its functional role.⁴² If we are to naturalize subjective character, then, we must come to terms with the possibility of self-directed intentionality.⁴³

The natural thought is that self-directed intentionality is a matter of the mental state representing itself. If so, the upshot of our discussion is that an account of subjective character would have to concentrate on the notion of self-representation. Mental states which feature inner awareness are self-representing states, states which represent their very own occurrence. This is what *constitutes* the fact that they are the act of awareness and the object of awareness all at once.⁴⁴ Therefore, what we must wrap our mind around, in attempt to naturalize the subjective character of conscious experience, is how a neuro-

⁴¹ Sydney Shoemaker (in conversation) objected to me that in the only sense in which functional role is dispositional, so is subjective character. The sense in question is that functional role properties individuate in terms of causal powers. My response is that there is another sense in which functional role is dispositional, and in which subjective character is not. This is the sense that functional role properties consist in, are nothing but, funds of causal powers. Subjective character, being a phenomenologically manifest feature, is not *just*—not *nothing but*—a fund of causal powers.

⁴² If we accept this approach, then the functional role of conscious experiences does not even *underlie* their self-representational content. Rather, the opposite is true: the experiences assume the relevant functional role precisely because of their self-representational content.

⁴³ Another problem that may face the functionalist account of subjective character is the following. According to Bealer (1996), functionalism cannot account for the intentionality of second-order representations, because it must attribute to them the wrong objects, namely, the physical realizations of the first-order mental states we intuitively wish to regard as their objects. For instance, when a theory featuring a thought that one thinks that *p* is completely functionalized, the thought that *p* is identified with the a certain physical realization R, and so the thought that one thinks that *p* would have to be identified with a thought that one is in R. According to some critics (McCullagh, 2000, Tully 2001), however, Bealer's argument is fallacious and does not yield a genuine threat to functionalism. I am not going to discuss the matter here. Suffice it that we note that *if* Bealer is right, then the functionalist account of subjective character discussed in the text faces a third formidable difficulty, on top of the two presented in the text.

⁴⁴ If the remarks in the previous paragraph are correct, it is also what serves as the categorical basis of their distinctive functional role.

physiological state of the brain could represent its own occurrence in the appropriate way.

4. Representationalist Accounts of Subjective Character

It may be thought curious that understanding the nature of consciousness should come down to making sense of physical self-representation. After all, such physical self-representation is abundant and unproblematic. Consider the sentence “This very sentence is written in Times New Roman.”⁴⁵ This sentence represents itself, and it is a physical entity.

However, this sentence represents itself only thanks to our interpretation of it. There is nothing in the concatenation of symbols which constitutes it that directs it towards itself. That is, it is not *intrinsically* about itself.⁴⁶ The question is how a physical entity could be intrinsically—that is, *in and of itself*—about itself.

In attempt to account for such intrinsically self-directed intentionality, two lines of thought may be pursued, which correspond to two aspects of intentionality, namely, the intentional *content* and the intentional *mode*. The distinction between content and mode is particularly striking for the propositional attitudes.⁴⁷ The difference between believing that it will rain and believing that it will snow is a difference in content, but the difference between believing that it will rain and hoping that it will rain is a difference in mode. A similar distinction applies to experiences. The difference between seeing a lion and seeing a bird is a difference in the content of experience, but the difference between seeing a lion and fearing a lion is a difference in the mode of experience.⁴⁸

⁴⁵ Note that such self-representational sentences can easily be misrepresentations. Thus, “This very sentence is written in Arial” is false, since it is written in Times New Roman, not Arial. This is consistent with my claim, in §1, that nothing in the notion of inner awareness commits one to infallibility or other special epistemic status.

⁴⁶ Perhaps this is because all linguistic intentionality is derivative, and only the intentionality of mental states is intrinsic or original, as Grice (1957) maintained.

⁴⁷ It is often put in terms of a distinction between the ‘representational content’ and the ‘psychological attitude’. For a systematic exposition of this distinction, see Searle (1983) ch. 1.

⁴⁸ In this case the difference is between a particular perceptual mode and a particular emotional mode. But we may extend this notion to differentiate also among different perceptual modes (i.e., visual vs. auditory) and among different emotional modes (i.e., fear vs. anger), as well as in relation to such somatic modes as pain and pleasure. It is an interesting question—into which we cannot go here—whether there is any experiential mode associated with the propositional attitudes. Russell, for one, thought so (1948, p. 108, 146). For him, the difference between believing that it will rain and hoping that it will rain is experiential: the feel of believing a proposition is different from the feel of hoping a proposition (if you will). More weakly, Brentano (1874/1973) thought that the difference between believing a proposition strongly and believing it hesitantly is a difference in experiential feel. If so, the difference between such attitudes as *being sure* that it will rain and *suspecting* that it will rain is a matter of experiential mode.

Corresponding to this distinction between the content and the mode of experience is a distinction between two ways of trying to account for the self-directed intentionality of conscious experiences. Smith argues that this self-directedness is built into the *mode* of conscious experience (1986, p. 150):⁴⁹

The reflexive character [of conscious experience] is thus part of the [mode], and not of the [content], of presentation in the experience. In the modal structure ‘in this very experience’, the experience is not, strictly speaking, presented... Hence, inner awareness does not consist in a second presentation, a presentation of the experience itself...

On this view, there is not only a difference between seeing a lion and fearing a lion, but also between *awarely*-seeing a lion and *unawarely*-seeing a lion.⁵⁰ The difference is that awarely-seeing is seeing in a way that includes inner awareness of itself, whereas unawarely-seeing is seeing without such inner awareness. A conscious experience has subjective character in virtue of employing an intentional mode of a certain kind, the “awarely” kind.

To construe subjective character in terms of the mode of experience is not the same as to offer a physicalist account of it, however. We now need to see how neurophysiological states of the brain can be said to constitute states of, e.g., awarely-seeing. For that matter, we must be told what that would *mean*. The problem is that it is impossible to see how this could be done satisfactorily within either of the two accepted accounts of intentional mode, which we may call the *direction-of-fit* account and the *functional-box* account.

According to the direction-of-fit account, the difference between believing and hoping is a difference in the ‘direction of fit’ between the state and the world. The satisfaction conditions of a belief require that it fit the way the world is; the direction of fit is in this sense mind-to-world. By contrast, the satisfaction conditions of a hope require that the world fit it; the direction of the fit is here world-to-mind. But it is impossible to see how awarely-seeing (or awarely-fearing) can be explicated in terms of a distinctive direction of fit. The inner awareness of a visual experience is veridical if the world is the way the awareness represents it to be. In this respect, it is not different from outer awareness. Therefore, its direction of fit is straightforwardly mind-to-world.

⁴⁹ Smith’s terminology is a little different here, as it is drawn from the phenomenological tradition. Following Meinong, Twardowski, and Husserl, he calls the intentional mode ‘modality of presentation’ and the intentional content ‘mode of presentation’. But in substance his view is the one we are now considering. Thus he also says (1986, p. 150 fn#7) that his terms are interchangeable with Searle’s (1983) ‘psychological mode’ (or ‘attitude’) and ‘intentional content’.

⁵⁰ In some places, Natsoulas seems to suggest something along the same lines. For instance, he writes (1996, p. 385): “[There is a] phenomenology of those visual experiences that are a product and part of a kind of seeing that is different from straightforward seeing: a kind of seeing that takes place usually upon adopting an introspective attitude with reference to one’s seeing. Elsewhere, I have called such seeing ‘reflective seeing’.”

There is no special direction of fit that is peculiar to inner awareness, so inner awareness cannot be cashed out in terms of direction of fit.

According to the functional-box account, the difference between believing that p and hoping that p is that in the former case the proposition p is stored in the so-called ‘belief box’, whereas in the latter case it is stored in a ‘hope box’. The box metaphor is then cashed out in terms of the distinctive functional roles of belief and hope. Thus, the hope that it will rain tends to cause one to check online whether it will, whereas the belief that it will rain tends to spare one the need to check. The problem here is obvious, namely, that the suggestion reduces to a functionalist account of inner awareness, which we have already seen to be unsatisfactory.

It may well be that a different account of intentional mode would provide for the eventual naturalization of the intrinsically self-directed intentionality of conscious experience. But none is in sight at the moment. We are therefore justified in expecting an account of it in terms of the *content* of experience.⁵¹

According to Levine, however, we have no idea how to make sense of this possibility. He develops a trilemma argument to that effect (2001, pp. 171–3). In the remainder of this section, I will argue that his trilemma is ineffective, along two dimensions. First, one of its horns is less than conclusive, and second, the three options he considers are not exhaustive. After arguing against Levine’s trilemma, however, I will present a different argument, which does present a principled challenge to any naturalization of subjective character in terms of self-representation.

Levine’s argument proceeds as follows. When a bluish conscious experience E , which is realized by a neurophysiological state of the brain B , represents itself, there are three descriptions under which it may represent itself. It may (i) represent itself *qua* the brain state B ; (ii) represent itself *qua* the bluish experience E ; or (iii) represent itself indexically, that is, *qua* whatever is doing the representing.⁵² All three options, Levine argues, yield unsatisfactory accounts of inner awareness.

The first option Levine dismisses without much argument.⁵³ But an argument is not hard to come by. Consider our concept of gray matter.⁵⁴

⁵¹ This appears to be the view of Caston (2002). It is also the view I defended in Kriegel (2003b).

⁵² One option which is not even worth listing is to say that the experience represents itself simply in virtue of *being* itself. This would make self-representation too cheap: every table would be self-representing, but obviously it is not the case that every table has a subjective character.

⁵³ He writes (2001, p. 172): “‘B’ [the representation of the experience *qua* brain state] is obviously a non-starter—there is no reason to think first-person self-representation delivers a neurophysiological description...” The argument seems to be simply that this view does not conform to the phenomenology of conscious experience as it reveals itself to introspection. To my mind, this is in fact a quite strong argument.

⁵⁴ This is a variation on an example due to Chris Hill.

When we think about gray matter, our thought is realized in a neurophysiological state of the brain that surely involves gray matter. So when we think of gray matter, gray matter represents itself. Yet this does not amount to the sort of inner awareness we are after. For we may certainly have unconscious thoughts about gray matter.

Levine's argument against the third option is that it cannot account for the way the experience is presented in the inner awareness of it.⁵⁵ An indexical represents what it does in a blank sort of way, not through any specific information about what it represents. Thus, the symbol "I" refers to the person who produced it, *whomever it may be*. Similarly, if the bluish experience represented itself indexically, it would represent itself as the state that is doing the representing, *whatever it may be*. But the inner awareness of the bluish experience is not an awareness of it as *whatever it may be*, but as *bluish*. So the inner awareness built into the experience cannot quite be captured by indexical self-representation.

This brings us to the second option. The bluish experience must represent itself precisely *as* a bluish experience. But, Levine argues, if the representation employs the description "bluish experience" in its representation of itself, there will be nothing special in its self-representation. It will represent itself simply by falling under the extension of the description it employs. It will represent itself in the same sense in which the word "word" may be said to represent itself. And just as the word "word" has no subjective character to it, so the bluish experience will not have a subjective character.⁵⁶

It is not clear to me, however, why the experience must represent itself as *a* bluish experience, rather than as *this* bluish experience. What seems to be missing in the second option, as presented by Levine, is a way for the experience to represent itself *qua particular* experience, not *qua* experience of a certain *type* (e.g., bluish). It seems that when a bluish experience represents itself, it represents itself not just *qua* bluish experience but *qua* one particular bluish experience.⁵⁷ That is, it must represent itself not under the indefi-

⁵⁵ Levine couches this horn of the trilemma in terms of what he calls our *substantive and determinate* awareness of our conscious experiences. This feature Levine attributes to conscious experiences is apparently very important to him. I cannot discuss it in detail here. The general idea is that the mode of presentation of qualia is substantive, in the sense of employing the very quale it presents in presenting it; and it is determinate, in that it presents the quale independently of anything else the quale may be related to.

⁵⁶ One problem with Levine's argument here is that "word" represents itself in a derivative way, whereas a bluish experience which would represent itself as a bluish experience would do so intrinsically. I am not going to dwell on this difficulty, though, because I think Levine is ultimately right to maintain that this sort of self-representation is somehow insufficient for inner awareness.

⁵⁷ It may even be that the experience represents itself, specifically, *qua* itself. On this suggestion, the experience does not just represent itself as particular, but as its-particular-self, if you please. We may say that the description under which it represents itself is not 'this bluish experience', but 'this *very* bluish experience'. This is an intriguing suggestion.

nite description ‘a bluish experience’, but under the definite description ‘this bluish experience’. If this is the description under which the experience represents itself, this form of self-representation may very well account for its subjective character.

Furthermore, the three options Levine considers are not jointly exhaustive. There are at least two other plausible options he fails to consider. One is that the experience represents itself both *qua* bluish experience *and qua* the thing that does the representing (i.e., indexically).⁵⁸ The other is that the experience does not represent itself *qua* anything, because it represents itself *non-conceptually*. In at least one sense of the notion, non-conceptual content is content that features objects and states of affairs not as falling under any description.⁵⁹ The possibility that the experience’s representation of itself is non-conceptual in this sense is therefore a genuine (and to my mind, not at all implausible) option which Levine does not consider.⁶⁰

I conclude that Levine’s trilemma argument is inconclusive. At the same time, it brings out the theoretical burden that rests on the shoulders of representationalists. Proponents of the representationalist account of inner awareness would have to offer quite specific semantics for mental representation before their proposal could be clearly evaluated.

There may also be a more principled problem with the self-representation strategy. It is that all naturalist accounts of intentionality, or mental representation, offer a reductive explanation of it in terms of some more or less complex causal relation, but causal relations are anti-reflexive (they never hold between a thing and itself). Thus no mental state can bring about its own

which would have to be explored more seriously elsewhere. Note, however, that even if it is correct, that is not what is missing from Levine’s second option. What is missing from Levine’s second option is that the experience represent itself as a *particular*. And the particular is of course itself (the experience). But whether the experience represents itself as its-particular-self is not something we can learn from examining the failure of Levine’s second option.

⁵⁸ More generally, then, Levine fails to consider the possibility of self-representation in accordance with more than one of the three options. The particular combination I cite in the text—self-representation along the lines of both the second and the third option—strikes me as particularly plausible, but other combinations are logically possible as well.

⁵⁹ For a thorough account of non-conceptual content, see Evans 1982 ch. 4. An account more in line with the needs of my present argument is Stalnaker’s (1998). Interestingly, Bermúdez (1998) argues for the existence of non-conceptual self-awareness. I am not going to discuss his arguments here, but they appear to support just as well non-conceptual inner awareness.

⁶⁰ I suspect Levine may reply that this option does not capture the *substantive and determinate* awareness we have of the experience (see footnote 55 above). But in any case, *some* reply is called for to encounter this option.

occurrence. And this seems to preclude a physicalistically kosher self-representation.⁶¹

This, I believe, is the fundamental problem for the program of naturalizing subjective character by appeal to self-representation. Let me therefore expand on it a little. According to naturalist accounts of mental representation, the latter involves a physical relation between the representing mental state and the represented environmental state. The environmental state, E, is the representational content of the mental representation, whereas the mental state, M, is the vehicle of representation. Different naturalist accounts differ in how they choose to elucidate the physical relation between M and E.⁶² The simplest version would be a straightforward causal account, according to which M represents E just in case E causes M. A more sophisticated version is Dretske's (1981) informational account, according to which M represents E just in case the occurrence of M-type states is nomically dependent upon the occurrence of E-type states, where this means that, given the laws of nature, the occurrence of M-type states is brought about *only* by E-type states.⁶³ Yet another naturalist theory is the teleological account, developed most comprehensively by Millikan (1984), according to whom M represents E only if M has been selected by evolutionary processes to covary with E.

Because of their pervasive appeal to causal relations, none of these accounts can make sense of self-representation. Consider the simple causal account. According to this account, for M to represent itself would be for M to be *caused* by itself. But, of course, no state can bring itself into existence. So within the framework of the causal account, self-representation is impossible. The other two accounts appeal to more complex physical relations, but these are still causal. Thus, according to the informational account, for M to represent itself is for M-type states to be brought about only by M-type states, given the laws of nature. But for M-type states to be brought about only by themselves, they would have to be brought about by themselves, which is again impossible.⁶⁴ Therefore self-representation is impossible

⁶¹ Levine (2001, p. 173) hints at this problem: "The problem with both materialist strategies for explaining subjectivity is this: all the materialist has out of which to construct the subjective relation to the contents of experience is the relation of cause and effect, or nomological covariation. But appeal to causal/nomological relations can't explain the cognitive intimacy we need for an account of conscious cognition."

⁶² A causal-informational approach is found in Dretske (1981, 1986, 1988); Fodor (1990). A teleological approach is found in Van Gulick (1980); Millikan (1984); McGinn (1989a); Papineau (1993). A long-armed functional role approach is in Harman (1987).

⁶³ Another way to put this is as follows: M nomically depends on E iff E-type states cause M-type states, and nothing else but E-type states causes M-type states, in all nomologically possible worlds.

⁶⁴ Rob Stainton suggested to me that Dretske might respond to this argument by emphasizing the distinction between types and tokens. The idea is that mental states have their representational content in virtue of belonging to a certain type, but the causal relations underlying the representation relation take place at the level of tokens. On this view, a

within the framework of the informational account. And similarly for the teleological account: selection is a causal process.⁶⁵

The general problem can be organized as the following argument: (1) according to all naturalist accounts of mental representation, mental representation implies a causal relation between the vehicle of representation and the content of representation; (2) the causal relation is anti-reflexive; therefore, (3) according to naturalist accounts, a mental state cannot represent itself. This argument appears to mount a principled barrier before materialist attempts to account for the subjective character evidently featured by some mental states.⁶⁶ It suggests that we cannot account for the duality of conscious experiences without positing a non-physical (hence super-natural) relation of self-representation.

mental state of type T_1 represents an environmental feature of type T_2 in virtue of the causal relations holding between tokens of T_1 and tokens of T_2 . So a mental state of type T_1 may represent itself in virtue of being caused by other tokens of type T_1 . The problem with this response is its empirical implausibility. It is quite implausible (and may even lead to infinite regress) that every conscious experience is caused by a mental state of the same type, and it is in virtue of being so caused that it represents itself.

⁶⁵ On the basis of teleological semantics, Chris Hill (in conversation) develops a different problem for self-representation. According to Hill, there is no reason to suspect that self-representation carries any evolutionary advantage, and would therefore be unlikely to be selected for. This problem I find less troubling. It is notoriously difficult to speculate with any accuracy about what may or may not augment an organism's reproductive potential. Moreover, not every biological trait is selected for: some traits stick around accidentally, just because nothing better has formed to supplant them. In some cases, we do not know what the evolutionary advantage of a feature is, if any, but we would never on that basis argue that it does not exist. A good example is the human chin (Sober 1985). Either the chin is one of those useless structures that just lie around un-supplanted, or it has some evolutionary value which we have failed to pin down, but one thing is sure—there are chins!

⁶⁶ Ernest Sosa has suggested to me rejecting premise (1) by construing mental representation in terms of Lewis's (1973) "counterfactual dependence" rather than in terms of causation. (A similar move, of allowing for self-representation by relaxing the representation relation, was suggested to me by Andrew Brook.) If we replace causation by counterfactual dependence in the simple account, for instance, then M represents E in virtue of counterfactually depending on E , which means that there is at least one possible world in which both M and E occur that is "closer" to the actual world than any possible world in which M occurs but E does not. This suggestion suffers, however, from the opposite problem, namely, that the relation of counterfactual dependence holds trivially between every state and itself. Since there is (trivially) at least one possible world in which M occurs that is closer to the actual world than any possible world in which M both occurs and does not occur, M (trivially) represents itself. But in this sense, a table also represents itself, but it certainly does not have subjective character. It is possible to substitute causation with counterfactual dependence in the more sophisticated naturalist accounts, such as Dretske's informational semantics, but given the "weakness" of counterfactual dependence, this is unlikely to lead to a relation that is neither reflexive nor anti-reflexive, but is instead a-reflexive in the way the relation of representation truly is.

5. Subjective Character and Cross-Order Information Integration

It is this sort of trouble that raises the specter of eliminativism in the philosophy of mind. Materialists discouraged by the prospects of accounting for self-representation may come to deny that mental states ever have a subjective character or inner awareness. In this final section, however, I want to consider one more way a materialist might try to account for the subjective character of conscious experience.⁶⁷

Consider a reddish experience of a red circle. The representational content of this experience is quite complex, since it wraps up together an outer awareness of the circle and an inner awareness of itself. Its representational content thus features both a red circle and an experience of a red circle. The approach I want to introduce suggests that this sort of representational content may be produced simply through the integration of information carried by what are initially separate mental states. When the contents of these separate mental states are appropriately integrated, a (single) mental state arises which has just the right sort of representational content.

In the case of the reddish experience, the experience arises through the integration of information carried by two separate mental states, one representing the red circle and one representing that very representation of the red circle. These two mental states (the representation of the circle and the representation of the representation of the circle) have no subjective character, of course. But when their representational contents are integrated, a mental state arises which does have subjective character. The reason this third state has subjective character is that it has the right sort of representational content: it folds within it a representation of an external object and a representation of that representation.

There is nothing inherently mysterious about the process of information integration itself. It is involved in the production of the simplest mental contents. Thus, the mental state representing the red circle involves the integration of information about the circle and information about the redness. We have known for a while that these bits of information are processed in different parts of the brain.⁶⁸ At the end of the initial information processing, then,

⁶⁷ This way emerges from very recent work by Van Gulick (2001, 2004, 2006) and myself (Kriegel 2002b, 2003a, 2006). At least the way I understand Van Gulick's recent account, it follows a similar line as mine. Van Gulick calls his approach the *higher-order global state (HOGS) model*. Similarly, elsewhere I claimed that a "conscious experience is a GW [Global Workspace] state...which incorporates information about the self and its present sensory state" (Kriegel 2002b, p. 529). Here I will avoid the reference to "global," or "global workspace," states, because I find that they are quite immaterial to the main tenets of the approach, especially when considered in the context of accounting for the dual nature of conscious states.

⁶⁸ This is what gives rise to the so-called "binding problem" in the brain sciences. This is the problem of understanding how the brain binds together the information processed in (both

the subject harbors one representation of a circle and a separate representation of red. It is only after a further process takes place, in which the information about the circle and the information about the redness are integrated, that a unified representation of a red circle emerges.⁶⁹

A representation of this representation may then be formed elsewhere in the subject's brain. The same process that is involved in the integration of the information about the circle and the information about the redness (or a similar process) may then be put to work in integrating the information about the red circle and the information about the internal representation of the red circle.⁷⁰ That process, in itself, is not particularly mysterious, then. Yet it yields a unified representation of the red circle and the representation of the red circle. This new representation has precisely the sort of representational content we want a state with subjective character to have.

The model just sketched is based on the integration of first-order and second-order information; hence the unlovely title of this section. The process of information integration is nothing special, and the first- and second-order mental states carrying the relevant information are no more special. But all this leads to the formation of a mental state with a quite special representational content. It leads to the formation of a mental state that is at once a representation of red and a representation of a representation of red, thus acquiring the special translucence we seek in states with subjective character.

It should be stressed that the integration in question must be conceived as leading to a sort of genuine unity or fusion whereby two separate states are superseded by a single state. Consider the following metaphor.⁷¹ When you taste a mushroom soup, you may find that the tastes of mushroom and garlic

structurally and functionally) disparate parts of the brain to produce the unified representational content characteristic of conscious experience.

⁶⁹ This process is what brain scientists call the binding process, then. Currently the popular theory of the binding process is in terms of firing synchrony (von der Malsburg 1981). When two subpopulations of neurons in two disparate parts of the brain fire their electrical pulse in almost absolute synchrony (within the millisecond range), the information they carry is bound into a single, cohesive content. That is, while the brain uses the *rate* of firing to represent the presence of certain environmental features (e.g., redness), it uses the *synchrony* of firing to represent the "togetherness" (if you will) of these features as belonging to the same object.

⁷⁰ Integration of information takes place at numerous levels in the cognitive system. At the very basic level, the representation of, say, an angular shape must be constructed out of distinct representation of the sides of the angle. It is only then that information about the shape can be integrated with other visual information, such as color or motion. Then multisensory information is integrated into a single representation: if our red circle produces a humming noise, for instance, the humming noise, the redness, and the circle are eventually all integrated into one representation. The process of information integration the model discussed in the text relies on is therefore at a very advanced stage. It is difficult to speculate about the similarity and difference of the higher-level processes of information integration to the lower-level processes (but for interesting work along such speculative lines, see Metzinger 1995).

⁷¹ The metaphor is due to Farid Masrouf (Ph.D. Dissertation, University of Arizona).

blend nicely. But a connoisseur might then turn up and claim that the soup is not quite perfect, in that the mushroom and the garlic have not quite fused so that the soup has a single flavor hitting her palate. The distinction which the connoisseur is able to draw, but you cannot, is between a blend of two flavors and a single flavor displacing the blend. Likewise, we might want to distinguish between, on the one hand, the compresence of two mental states, the first-order one and the second-order one, and on the other, the unity and integrity of a single mental state with a first- and second-order component.

How should we make sense of this notion of integration and the integrity that results from it? This is a complicated question that requires a long answer, one which we cannot responsibly delve into here.⁷² The important feature of the cross-order integration model is that the mental state yielded by that integration may not necessarily represent *itself*, at least not in the first instance. At most, we can say that one part of it represents another part.

This could lead to the objection that the account under consideration fails to provide for genuine inner awareness. It might be claimed that the dual nature of conscious experiences as both acts and objects of awareness *requires* that they be construed as self-representing states.

But it could be responded that what is important about states with a dual nature as act and object is not so much that they be self-representing, but that they have the sort of representational content that self-representation would confer on them. If they can acquire this kind of representational content *without* representing themselves, that in itself should not disqualify them from exhibiting a subjective character. After all, we concluded from the discussion of functionalist accounts in §3 that an account of subjective character must be *representational*. So given that the sort of mental state I have in mind has the very same representational content as the one requested by Levine, it would seem unduly obstinate to insist that the present model still does not capture what we want in the dual nature of conscious experiences.

Furthermore, although in the first instance the most we can say of a state that arises from cross-order integration is that one part of it represents another part, in a second instance, we could hold that the second-order part of the state represents the entire state *by* representing the first-order part of the state, just as one can represent an entire tree *by* representing a part of it.⁷³

It may also be objected to the model under consideration that it reduces, in fact, to Gennaro's unhelpful suggestion that subjective character is a matter of one part of a mental state representing a separate part of the same state.

This objection is misguided, however. There is a crucial difference between Gennaro's model and the present one. In Gennaro's model, the relation between the experience and its inner awareness is merely conceptual,

⁷² For some work in that direction, see Kriegel (2003a and 2006).

⁷³ See Kriegel (2006).

whereas in the present model it is a real relation, namely, the relation effected by information integration. That is, in the present model there is a distinction between, on the one hand, the *compresence* of a representation and the representation of that representation, and, on the other hand, the *integration* of a representation and the representation of that representation. Only the latter yields subjective character, and it implies an actual causal (hence temporally extended) process absent from the former. Gennaro's model, by contrast, provides for no such distinction: there is nothing to the "togetherness" of a representation and the representation of that representation (as parts of the same mental state) beyond their compresence.

The objection might be pressed further, however.⁷⁴ It could be claimed that while the present model differs from Gennaro's at the level of implementation, it does not at the level of content. For there is a crucial difference between the integration of information about the circle and the information about the redness and the integration of information about the red circle and the information about the representation thereof. In the former, the integrated information concerns *the same object*, an object which is both circular and red, whereas in the latter, it concerns different objects, a red circle and a representation.

However, although it is natural to construe the information in the latter case as concerning different objects, it is also possible to construe it as concerning the same object, namely, the external circle. The content of the resulting state will be something like "this particular is red and circular and represented (hereby) to be red and circular."⁷⁵ One and the same object is predicated as red, circular, and represented-(thereby-)to-be-red-and-circular.

Another objection may be the following.⁷⁶ It is natural to think that the same mental state can occur both consciously and unconsciously. Thus, a person may believe that the moon is not made of cheese first unconsciously and then consciously. But it is a consequence of the account under consideration that this is impossible. Once a mental state turns conscious, it becomes a different state, with a part it did not have before.

This objection is correct, insofar as the account under consideration does indeed entail that a conscious mental state can never occur unconsciously. But

⁷⁴ I would like to thank an anonymous referee for *Philosophy and Phenomenological Research* for indeed pressing me on this point.

⁷⁵ I do not intend to suggest that the content must be propositional; merely to illustrate the kind of content a state resulting from cross-order integration would bear. Any other illustration would do. The referee for this journal who raised this objection framed it as follows. When the information about the circle and the redness is integrated, content is transformed from <(there is an x) x is red> & <(there is an x) x is circular> into <(there is an x) x is red & x is circular>. My suggestion is that in the analogous cross-order integration, the content is transformed into <(there is an x) x is red & x is circular & x is represented-to-be-red-and-circular>.

⁷⁶ This objection was presented to me by Graham Oddie.

the account is not at odds with our pre-theoretical intuitions, it merely formulates them differently. Thus, the account does allow that a person may believe that the moon is not made of cheese first unconsciously and then consciously. But whereas other accounts of consciousness would describe the situation as one in which the same belief state acquires a new property (i.e., the property of being conscious), our account describes it as a situation in which the person enters a new belief state. The account does not entail that such a situation is impossible, then. It merely describes the transition in more dramatic terms than others do.⁷⁷

The cross-order information integration model is also immune to the drawbacks of the models considered in previous sections. Unlike the conceptualist approach, it does not wrongly assume that the difference between one mental state with two parts and two distinct mental states is merely verbal. Furthermore, it does not hold that the two parts of a conscious state are *separate*, since the two parts must be integrated in order for them to amount to a conscious experience. Unlike the functionalist approach, it does not attempt to account for subjective character in non-intentional terms. Furthermore, it does not attempt to account for it in dispositional terms, but in categorical ones. And unlike the representationalist approach, it does not appeal to a relation of self-representation.⁷⁸

To conclude. The cross-order information integration approach offers, in fact, a representationalist account of subjective character, but one which does not appeal to self-representation. Instead, it suggests a way in which mental states that are not self-representing could have the sort of representational content that self-representing states would have. It also ascribes to these states the distinctive inferential role we described in §3, although it does not take this to account for their subjective character. The mental states in question thus feature the distinctive representational content and functional role of conscious experiences. The approach also withstands the main lines of objection that bedevil the other approaches to naturalizing subjective character.

It would be overly presumptuous to conclude from these remarks that cross-order information integration *must* be the key to the subjective character of conscious experience. But the notion ought to be explored, at a multitude of levels (philosophical as well as cognitive-scientific). If anything, this approach to naturalizing subjective character enjoys the virtue of not being

⁷⁷ I am assuming here that our pre-theoretical intuitions pertain to the situation in which a person believes something first unconsciously and then consciously. This intuition can be accommodated by the account I defend in the text. An insistent objector may hold that the account does not accommodate the intuition that the mental state is the same when unconscious and when conscious. To my mind, however, to the extent that we do have such an intuition, it is not a *pre-theoretical* intuition.

⁷⁸ That is, the second-order part of the representational content of a bluish conscious experience is not “this very bluish experience,” but rather “this bluish experience” (if we use “this very” to signal self-reference).

straightforwardly unworkable as the other functionalist and representationalist accounts we have considered. And while we may acquiesce in the existence of a genuine explanatory gap between consciousness and physical reality, it would probably be unwise to go as far as McGinn (1989b, 1999) and speculate that humans as such are incapable of bridging that gap, due to certain constitutional limitations on their intellect. So the fact that we are nowhere near solving the problem of consciousness should not give us pause in *trying* to. What I would like to suggest is that, in *trying* to solve the problem of consciousness, the cross-order information integration approach may be the most viable one to pursue.

6. Conclusion

Some mental states are conscious experiences and some are not. Those that exhibit a subjective character, or for-me-ness, whereas those that do not. This subjective character consists in a special representational content which folds within it a representation of some environmental feature and a representation of that representation. How to account for this special representational content is the problem facing materialist theories of consciousness.

In this paper, I have discussed six approaches to the problem, so posed. I found the first five wanting, and in rather principled ways. The conceptualist approach substitutes the real relation a conscious experience bears to itself with a merely nominal relation. The functionalist approaches may at best account for what *underlies* the special representational content of conscious experiences, but fail to account for what *constitutes* that content. And the self-representation approaches cannot provide for a physical self-representation relation that is not anti-reflexive.

The approach I have developed and defended hinges on the cognitive integration of cross-order information, more specifically information about an environmental feature and about the internal representation of that environmental feature. This approach, I argued, is our best bet for trying to naturalize the subjective character of conscious experience.⁷⁹

References

- Aquila, R. 1990. "Consciousness and Higher-Order Thoughts: Two Objections." *American Philosophical Quarterly* 27 (1990): 81-87.
- Armstrong, D. M. 1968. *A Materialist Theory of the Mind*. New York: Humanities Press.

⁷⁹ Versions of this paper have been presented in early 2003 at the University of Colorado at Boulder, Carleton University, the University of South Florida, and later at Syracuse University. I am indebted to the audiences there, in particular to the following individuals: Andy Brook, Robert Hanna, Rocco Gennaro, Chris Hill, Michael Huemer, Bill Lycan, Graham Oddie, Gabe Shapiro, Ernest Sosa, Rob Stainton, Cybele Tom, Michael Tully, Robert Van Gulick, Eric Winsberg, and especially Sydney Shoemaker.

- Armstrong, D. M. 1981. "What Is Consciousness?" In his *The Nature of Mind* pp. 55-67. Reprinted in Block et al., 1997.
- Bealer, G. 1996. "Self-Consciousness." *Philosophical Review* 106: 69-117.
- Bermúdez, J. L. 1998. *The Paradox of Self-Consciousness*. Cambridge, MA: MIT Press/Bradford Books.
- Block, N. 1986. "Advertisement for a Semantics for Psychology." *Midwest Studies in Philosophy* 10 (1986): 615-677.
- Block, N. J. 1995. "On a Confusion About the Function of Consciousness." *Behavioral and Brain Sciences* 18 (1995): 227-247. Reprinted in Block et al., 1997.
- Block, N. J., Flanagan, O., and Guzeldere, G., eds. 1997. *The Nature of Consciousness: Philosophical Debates*. Cambridge MA: MIT Press/Bradford Books.
- Brandom, R. 1994. "Reasoning and Representing." In *Philosophy of Mind*, ed. M. Michael and J. O'Leary-Hawthorne. Dordrecht: Kluwer Academic Publishers.
- Brentano, F. 1874/1973. *Psychology from Empirical Standpoint*. Ed. O. Kraus. Ed. of English edition L. L. McAlister, 1973. Translation A. C. Rancurello, D. B. Terrell, and L. L. McAlister. London: Routledge and Kegan Paul.
- Brough, J. B. 1972. "The Emergence of an Absolute Consciousness in Husserl's Early Writings on Time-Consciousness." *Man and World* 5 (1972): 298-326.
- Burge, T. 1979. "Individualism and the Mental." *Midwest Studies in Philosophy* 4 (1979): 73-121.
- Byrne, A. 1997. "Some Like It HOT: Consciousness and Higher-Order Thoughts." *Philosophical Studies* 86 (1997): 103-129.
- Carruthers, P. 1989. "Brute Experience." *Journal of Philosophy* 85 (1989): 258-269.
- Carruthers, P. 1996. *Language, Thought, and Consciousness*. Cambridge and New York: Cambridge University Press.
- Carruthers, P. 2000. *Phenomenal Consciousness*. Cambridge: Cambridge University Press.
- Caston, V. 2002. "Aristotle on Consciousness." *Mind* 111 (2002): 751-815.
- Dennett, D. C. 1988. "Quining Qualia." In *Consciousness in Contemporary Science*, ed. A. Marcel and E. Bisiach. Oxford: Oxford University Press.
- Dretske, F. I. 1981. *Knowledge and the Flow of Information*. Oxford: Clarendon Press.
- Dretske, F. I. 1986. "Misrepresentation." In *Belief*, ed. R. Bogdan. Oxford: Oxford University Press.
- Dretske, F. I. 1988. *Explaining Behavior*. Cambridge MA: MIT Press/Bradford Books.

- Dretske, F. I. 1993. "Conscious Experience." *Mind* 102 (1993): 263-283.
- Dretske, F. I. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press/Bradford Books.
- Evans, G. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.
- Field, H. 1977. "Logic, Meaning, and Conceptual Role." *Journal of Philosophy* 74 (1977): 379-409.
- Fodor, J. A. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press/Bradford Books.
- Fodor, J. A. 1990. *A Theory of Content and Other Essays*. Cambridge MA: MIT Press/Bradford Books.
- Frank, M. 1995. "Mental Familiarity and Epistemic Self-Ascription." *Common Knowledge* 4 (1995): 30-50.
- Gennaro, R. J. 1996. *Consciousness and Self-Consciousness*. Philadelphia and Amsterdam: John Benjamin Publishing Co.
- Gettier, E. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23: 121-123.
- Goldman, A. 1993. "Consciousness, Folk Psychology, and Cognitive Science." *Consciousness and Cognition* 2 (1993): 364-383.
- Grice, P. 1957. "Meaning." *Philosophical Review* 66 (1957): 377-388.
- Gurwitsch, A. 1985. *Marginal Consciousness*. Athens OH: Ohio University Press.
- Guzeldere, G. 1995. "Is Consciousness the Perception of What Passes in One's Own Mind?" In *Conscious Experience*, ed. T. Metzinger. Paderborn: Schoeningh-Verlag.
- Harman, G. 1982. "Conceptual Role Semantics." *Notre Dame Journal of Formal Logic* 23 (1982): 242-256.
- Harman, G. 1987. "(Non-Solipsistic) Conceptual Role Semantics." In *New Directions in Semantics*, ed. E. LePore. London: Academic Press.
- Harman, G. 1990. "The Intrinsic Quality of Experience", in *Philosophical Perspectives* 4: 31-52, rpr. in Block et al. (1997).
- Haugeland, J. 1978. "The Nature and Plausibility of Cognitivism." *Behavioral and Brain Sciences* 1 (1978): 215-226.
- Henrich, D. 1966/1982. "Fichte's Original Insight." Translation D. R. Lachterman. *Contemporary German Philosophy* 1 (1982): 15-53.
- Husserl, E. 1928/1964. *Phenomenology of Internal Time-Consciousness*. Ed. M. Heidegger. Translation J. S. Churchill, 1964. Bloomington, IN: Indiana University Press.
- James, W. 1890. *The Principles of Psychology*, 2 vols. London: McMillan (second edition, 1918).

- Kriegel, U. 2002a. "PANIC Theory and the Prospects for a Representationalist Theory of Phenomenal Consciousness." *Philosophical Psychology* 15 (2002): 55-64.
- Kriegel, U. 2002b. "Consciousness, Permanent Self-Awareness, and Higher-Order Monitoring." *Dialogue* 41 (2002): 517-540.
- Kriegel, U. 2003a. "Consciousness, Higher-Order Content, and the Individuation of Vehicles." *Synthese* 134 (2003): 477-504.
- Kriegel, U. 2003b. "Consciousness as Intransitive Self-Consciousness: Two Views and an Argument." *Canadian Journal of Philosophy* 33 (2003): 103-132.
- Kriegel, U. 2003c. "Consciousness as Sensory Quality and as Implicit Self-Awareness." *Phenomenology and the Cognitive Sciences* 2 (2003): 1-26.
- Kriegel, U. 2006. "The Same-Order Monitoring Theory of Consciousness." In U. Kriegel and K. Williford (eds.), *Self-Representational Approaches to Consciousness*. Cambridge MA: MIT Press.
- Levine, J. 2001. *Purple Haze: The Puzzle of Consciousness*. Oxford and New York: Oxford University Press.
- Lewis, D. C. 1973. "Causation." *Journal of Philosophy* 70 (1973): 556-567.
- Lycan, W. G. 1990. "Consciousness as Internal Monitoring." *Philosophical Perspectives* 9 (1990): 1-14.
- Lycan, W. G. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press/Bradford Books.
- Lycan, W. G. 2001. "A Simple Argument for a Higher-Order Representation Theory of Consciousness." *Analysis* 61 (2001): 3-4.
- McCullagh, M. 2000. "Functionalism and Self-Consciousness." *Mind and Matter* 15: 481-499.
- McGinn, C. 1989a. *Mental Content*. Oxford: Blackwell.
- McGinn, C. 1989b. "Can We Solve the Mind-Body Problem?" *Mind* 98 (1989): 349-366.
- McGinn, C. 1999. *The Mysterious Flame*. Cambridge MA: MIT Press/Bradford Books.
- von der Malsburg, C. 1981. "The Correlation Theory of Brain Function." Technical Report 81-2, Max-Planck-Institute for Biophysical Chemistry, Göttingen.
- Mangan, B. 2001. "Sensation's Ghost: The Non-Sensory 'Fringe' of Consciousness." *Psyche* 7(18). <http://psyche.cs.monash.edu.au/v7/psyche-7-18-mangan.html>
- Metzinger, T. 1995. "Faster than Thought: Holism, Homogeneity, and Temporal Coding." In *Conscious Experience*, ed. T. Metzinger. Thorverton: Imprint Academic.

- Millikan, R. G. 1984. *Language, Thought, and Other Biological Categories*. Cambridge MA: MIT Press/Bradford Books.
- Moore, G. E. 1942. "A Reply to My Critics." In *The Philosophy of G. E. Moore*, ed. P. A. Schlipp. New York: Tudor Publishing Company.
- Moran, R. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton, NJ: Princeton University Press.
- Nagel, T. 1974. "What Is It Like to Be a Bat?" *Philosophical Review* 83 (1974): 435-450.
- Natsoulas, T. 1993. "What Is Wrong with Appendage Theory of Consciousness?" *Philosophical Psychology* 6 (1993): 137-154.
- Natsoulas, T. 1996. "The Case for Intrinsic Theory: II. An Examination of a Conception of Consciousness₄ as Intrinsic, Necessary, and Concomitant." *Journal of Mind and Behavior* 17 (1996): 369-390.
- Natsoulas, T. 1999. "The Case for Intrinsic Theory: IV. An Argument from How Conscious₄ Mental-Occurrence Instances Seem." *Journal of Mind and Behavior* 20 (1999): 257-276.
- Neander, K. 1998. "The Division of Phenomenal Labor: A Problem for Representational Theories of Consciousness." *Philosophical Perspectives* 12 (1998): 411-434.
- Papineau, D. 1993. *Philosophical Naturalism*. Oxford: Blackwell.
- Putnam, H. 1967, "Psychological Predicates." In *Art, Mind, and Religion*, ed. W. H. Capitan and D. D. Merrill. Pittsburgh, PA: University of Pittsburgh Press. Reprinted as "The Nature of Mental States" in *The Nature of Mind*, ed. D. M. Rosenthal, 1991. Oxford and New York: Oxford University Press.
- Putnam, H. 1975. "The Meaning of 'Meaning'." In his *Mind, Language, and Reality*. Cambridge: Cambridge University Press.
- Rey, G. 1988. "A Question about Consciousness." In *Perspectives on Mind*, ed. H. Otto and J. Tueidio. Norwell: Kluwer Academic Publishers.
- Rosenthal, D. M. 1986. "Two Concepts of Consciousness." *Philosophical Studies* 94 (1986): 329-359.
- Rosenthal, D. M. 1990. "A Theory of Consciousness." ZiF Technical Report 40, Bielfield, Germany. Reprinted in Block et al., 1997.
- Rosenthal, D. M. 2003. *Consciousness and Mind*. Forthcoming with Oxford University Press.
- Russell, B. 1922. *Our Knowledge of the External World*. London: George Allen & Unwin.
- Russell, B. 1948. *Human Knowledge: Its Scope and Limits*. London: George Allen & Unwin.
- Ryle, G. 1949. *The Concept of Mind*. London: Hutchinson and New York: Barnes and Noble.

- Sartre, J.-P. 1937. *La Transcendence de l'ego*. Paris: Vrin.
- Sartre, J.-P. 1943. *L'Être et le Néant*. Paris: Gallimard.
- Searle, J. R. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Shoemaker, S. 1994. "Self-knowledge and 'Inner Sense.' Lecture II: The Broad Perceptual Model." *Philosophy and Phenomenological Research* 54 (1994): 271-290.
- Shoemaker, S. 1996. *The First-Person Perspective and Other Essays*. Cambridge and New York: Cambridge University Press.
- Smith, D. W. 1986. "The Structure of (Self-)Consciousness." *Topoi* 5 (1986): 149-156.
- Smith, D. W. 1989. *The Circle of Acquaintance*. Dordrecht: Kluwer Academic Publishers.
- Smith, D. W. 1999. "Intentionality Naturalized?" In *Naturalizing Phenomenology*, ed. J. Petitot, F. J. Varela, B. Pachoud, and J.-M. Roy. Stanford, CA: Stanford University Press.
- Sober, E. 1985. "Putting the Function Back into Functionalism." *Synthese* 64 (1985): 165-193.
- Sokolowski, R. 1974. *Husserlian Meditations*. Evanston, IL: Northwestern University Press.
- Stalnaker, R. 1998. "What Might Nonconceptual Content Be?" *Philosophical Issues* 9 (1998): 339-352.
- Thomasson, A. L. 2000. "After Brentano: A One-Level Theory of Consciousness." *European Journal of Philosophy* 8 (2000): 190-209.
- Tully, M. 2001. "Functional Concepts, Referentially Opaque Contexts, Causal Relations, and the Definition of Theoretical Terms." *Philosophical Studies* 105: 251-279.
- Tye, M. 1995. *Ten Problems of Consciousness*. Cambridge MA: MIT Press/Bradford Books.
- Tye, M. 2000. *Consciousness, Color, and Content*. Cambridge MA: MIT Press/Bradford Books.
- Van Gulick, R. 1980. "Functionalism, Information, and Content." *Nature and System* 2 (1980): 139-162.
- Van Gulick, R. 2001. "Inward and Upward—Reflection, Introspection, and Self-Awareness." *Philosophical Topics* 28 (2001): 275-305.
- Van Gulick, R. 2004. "Higher-Order Global States (HOGS): An Alternative Higher-Order Model of Consciousness." In R. Gennaro (ed.), *Higher-Order Theories of Consciousness*. Amsterdam and Philadelphia: John Benjamin Publishing Co.
- Van Gulick, R. 2006. "Mirror Mirror—Is That All?" In U. Kriegel and K. Williford (eds.), *Self-Representational Approaches to Consciousness*. Cambridge MA: MIT Press.

- Wider, K. 1997. *The Bodily Nature of Consciousness: Sartre and Contemporary Philosophy of Mind*. Ithaca, NY: Cornell University Press.
- Williams, J. N. 1981. "Inconsistency and Contradiction." *Mind* 90 (1981): 600-602.
- Zahavi, D. 1998. "Brentano and Husserl on Self-Awareness." *Études Phénoménologiques* 27-8 (1998): 127-169.
- Zahavi, D. 1999. *Self-awareness and Alterity*. Evanston, IL: Northwestern University Press.