Philosophical Theories of Consciousness, Contemporary Western Perspectives, Uriah Kriegel, in Cambridge Handbook of Consciousness (edited by M. Moscovitch, E. Thomspon, and P.D. Zelato).2006, Cambridge and New York: Cambridge UP, P 35- 66

النظريات الفلسفية للوعي

من منظور غربي معاصر

يوريا كرايجل

1. مقدمة: مفهوم الوعي

هذا الفصل سوف يقدم مسحا للمقاربات الحالية للوعي في الفلسفة التحليلية الأنجلو-أمريكية. سوف يركز الفصل على خمس مقاربات، سوف أشير إليها بالغموضية (mysterianism)، الثنائية (dualism)، التمثيلية (representationalism)، نظرية المراقبة على المستوى الأعلى (higher-order monitoring theory)، ونظرية التمثيل الذاتي (self-representation theory). مع كل مقاربة سوف أقدم على الترتيب 1) الصياغة الرائدة للوعي مع خطوطها العامة، 2) الدفاع عن هذه المقاربة، و 3) الاعتراضات عليها[1].

لن أصدر حكما نهائيا على أي مقاربة منها، على الرغم من أنه مع نهاية الفصل يجب أن يكون واضحا أين تقع وجهة نظري.

قبل البدء، دعنا نضع بعض التمييزات التي يمكن أن تساعدنا على أن تكون أفكارنا واضحة في المناقشات التي ستلي. المصطلح "الوعي" (consciousness) يطبق بمعان مختلفة لأنواع مختلفة من الأشياء. يطبق، بمعنى معين، للأنواع الحيوانية، كما هو الحال عندما نقول بأن "الغوريللات واعية بينما القواقع ليست كذلك:؛ في معنى آخر، بالنسبة للمخلوقات الحية الفردية، كما عندما نقول "جيم واعي، ولكن جيل في غيبوبة"؛ وبمعنى ثالث، بالنسبة لحالات، أحداث، وعمليات عقلية معينة، كما عندما نقول "أفكاري عن فيينا واعية، ولكن اعتقادات جيم بأن هناك طيور في الصين ليست كذلك". لكي نميز هذه المعاني المختلفة، يمكن لنا أن نسمي الأول، *وعي الأنواع الحيوانية*، والثاني، *وعي المخلوقات الحية*، والثالث، *وعي الحالة العقلية*.[2].

يبدو أن هناك روابط تصورية معينة بين هذه المعاني الثلاثة، بحيث يمكن تحليل أي منها من خلال الآخر. بشكل مقبول، وعي الأنواع الحيوانية يمكن تحليله على أساس وعي المخلوقات الحية: نوع حيواني س هو واع كنوع فقط في حالة ما إذا كانت عينة ممثلة من س تمتلك وعيا كمخلوق حي. ووعي المخلوقات، بدوره، يمكن تحليله على أساس حالات الوعي: مخلوق ص هو واع كمخلوق فقط في حالة ما إذا كان ص يمتلك (أو قادرا على أن يمتلك) حالات عقلية تمثل حالات واعية. وإذا كان ذلك، فإن وعي الحالة العقلية هو الفكرة الأكثر أساسية في الثلاثة.

وعي الحالة العقلية هو ذاته غامض ما بين عدة معاني. إذا كان جيم يؤمن بصمت بأن هناك طيور في الصين، ولكن بدون أن يشعر أبدا بشكل واع بهذا الاعتقاد، في حين أن جيل تتأمل عادة بوعي حقيقة وجود طيور في الصين، ولكنها لا تفعل ذلك الآن، فهناك معنى "للوعي" يمكن فيه لنا القول بأن اعتقاد جيم غير واع في حين أن اعتقاد جيل واع. دعنا نسمي هذا *الوعي المتاح* (availability consciousness)[3].

في المقابل، هناك معنى "للوعي" تكون فيه الحالة العقلية واعية عندما، وفقط عندما، يكون *هناك شيئا ما مثله* (something it is like) بالنسبة للموضوع – من الداخل – يمكن امتلاكه[4]. هكذا، عندما أتذوق ملعقة عسل، يكون هناك بشكل خاص جدا – حلو، ناعم، عسلي إذا شئت – شيئا ما بالنسبة لي حتى أحصل على هذه الخبرة الواعية الذوقية المحددة. دعنا نسمي هذا *الوعي الظاهري* (phenomenal consciousness).

بعض النظريات الرائدة في الوعي – مثل بارز (Baars) (1988، 1997) "نظرية مكان العمـل الكوني" (global workspace theory)، وكريك وكوش (Crick and Koch) (1990، 2003) "نظريـة الحيوية العصبونية" (neurobiological theory) المؤسسة على التزامن (synchrony) – تلقي الكثير من الضوء على الوعي المتاح والأفكار القريبة لها. ولكن هناك إحساس ملح أنها لا تفعل الكثيـر لتفـسير الوعي الظاهري. وأكثر من ذلك، هناك شعور واسع بأن هناك شيء أساسي بخصوص الطريقة التـي تفشل فيها في أن تفعل ذلك. أحد الطرق لإظهار هذا الشعور هو من خلال مفاهيم فلسفية مثـل الهـوة التفسيرية (explanatory gap) (ليفين 1983) أو المشكلة الصعبة (تشالمرز 1995). طبقا لتـشالمرز، على سبيل المثل، مشكلة تفسير الوظائف المعرفية المتعددة للوعي هـي "مـشكلات سـهلة" للـوعي، "المشكلة الصعبة" هي تلك الخاصة بإدراك لماذا يجب أن يكون هناك *شـيئا مـا مثلـه* لتنفيـذ هـذه الوظائف[5]. المعنى هو أن هناك حاجة إلى تبصرات لمنظومة مختلفة اختلافا تاما حتى يمكـن تكـوين نظريات علمية للوعي، وحقيقة للعلم نفسه، بالنسبة لكل ما يخص فهمنا للوعي الظاهري. هناك حاجـة لنوع ما من الاختراق، الذي يمكننا من تصور مشكلة الوعي بشكل جديد. وهنـا حيـث تظهـر فـي الصورة النظريات الفلسفية للوعي.[6,7]

2. Mysterianism

Some philosophers hold that science cannot and will not, in fact, help us understand consciousness. So-called mysterianists hold that the problem of consciousness – the problem of how there could be something like phenomenal consciousness in a purely natural world – is not a problem we are capable (even in principle) of solving. Thus consciousness is a genuine mystery, not merely a *prima facie* mystery which we may one day demystify.

We may introduce a conceptual distinction between two kinds of mysterianism – an ontological one and an epistemological one. According to ontological mysterianism, consciousness cannot be demystified because it is an inherently mysterious (perhaps super-natural) phenomenon.[8] According to epistemological mysterianism, consciousness is in no way inherently mysterious, and a greater mind could in principle demystify it – but it just so happens that we humans lack the cognitive capacities that would be required.

Epistemological mysterianism has actually been pursued by contemporary western philosophers. The most comprehensive development of the view is offered in Colin McGinn's (1989, 1995, 1999, 2004) writings. We now turn to an examination of his account.

2.1. McGinn's Mysterianism

McGinn's theory of consciousness has two central tenets. First, the phenomenon of consciousness is in itself perfectly natural and nowise mysterious. Second, the human mind's conceptual capacities are too poor to demystify consciousness. That is, McGinn is an epistemological mysterianist: he does not claim that the world contains, in and of itself, insoluble mysteries, but he does contend that *we* will never understand consciousness.

At the center of McGinn's theory is the concept of *cognitive closure*. McGinn (1989: 529) defines cognitive closure as follows: "A type of mind M is cognitively closed with respect to a property P (or a theory T) if and only if the concept-forming procedures at M's disposal cannot extend to a grasp of P (or an understanding of T)."[9] To be cognitively closed to X is thus to lack the procedure for concept formation that would allow one to form the concept of X.

To illustrate the soundness and applicability of the notion of cognitive closure, McGinn adduces the case of animal minds and *their* constitutive limitations. As James Joyce writes in *A Portrait of the Artist as a Young Man*, rats' minds do not understand trigonometry. Likewise, snails do not understand quantum physics and cats do not understand market

economics. Why should humans be spared this predicament? As a natural, evolved mechanism, the human mind must have its own limitations. One such limitation, McGinn suggests, may be presented by the phenomenon of consciousness.

Interestingly, McGinn does *not* claim that we are cognitively closed to consciousness itself. Rather, his claim is that we are cognitively closed to that property of the brain responsible for the *production* of consciousness. As someone who does not wish to portray consciousness as inherently mysterious, McGinn is happy to admit that the brain has the capacity to somehow produce conscious awareness. But *how* the brain does so is something he claims we cannot understand. Our concept-forming procedures do extend to a grasp of consciousness, but they do not extend to a grasp of the *causal basis* of consciousness in the brain.

2.2. The Master Argument for Mysterianism

A natural reaction to McGinn's view is that it may be based upon an overly pessimistic induction. From the fact that all the theories of consciousness we have come up with to date are hopelessly unsatisfactory, it should not be concluded that our future theories will be the same. It may well be that a thousand years hence we will look back with amusement at the days of our ignorance and self-doubt.

However, McGinn's main argument for his position is not the inductive argument just sketched. Rather, it is a deductive argument based on consideration of our cognitive constitution. The argument revolves around the claim that we do not have a single mechanism, or faculty, that can access *both* consciousness and the brain. Our access to consciousness is through the faculty of introspection. Our access to the brain is through the use of our senses, mainly vision. But unfortunately, the senses do not give us access to consciousness proper and introspection does not give us access to the brain proper. Thus, we cannot see with our eyes what it is like to taste chocolate. Nor can we taste with our buds what it is like to taste chocolate. We can, of course, taste chocolate. But we cannot taste the feeling of tasting chocolate. The feeling of tasting chocolate is something we encounter only through introspection. But alas, introspection fails to give us access to the brain. We cannot introspect neurons, and so could never introspect the neural correlates of consciousness.

Using the term "extrospective" to denote the access our senses give us to the world, McGinn's argument may be formulated as follows:

1) We can have introspective access to consciousness but not to the brain;
2) We can have extrospective access to the brain but not to consciousness;
3) We have no accessing method that is both introspective and extrospective; therefore,
4) We have no method that can give us access to both consciousness and the brain.

As we can see, the argument is based on considerations that are much more principled than a simple pessimistic induction from past theories. Dismayed as we may be by the prospects of mysterianism, we must not confuse McGinn's position for sheer despair. Instead, we must contend with the argument just formulated.

Some materialists would contest the first premise. Paul Churchland (1985) has repeatedly argued that we will one day be able to directly *introspect* the neurophysiological states of our brains. Perception and introspection are theory-laden, according to Churchland, and can therefore be fundamentally changed when the theory they are laden *with* is changed.[10] Currently, our introspective practice is laden with a broadly Cartesian theory of mind. But when we mature enough scientifically, and when the right neuroscientific theory of consciousness makes its way to our classroom and living room, this will change and we (or

rather our distant offspring) will start thinking about ourselves in purely neurophysiological categories.

Other materialists may deny the second premise of the argument. As long as brain states are considered to be merely *correlates* of conscious states, the claim that the conscious states cannot be perceived extrospectively is plausible. But according to materialists, conscious states will turn out to be *identical with* the brain states in question, rather than merely *correlated* therewith. If so, perceiving those brain states would just *be* perceiving the conscious states.[11] To assume that we cannot perceive the conscious states is to beg the question against the materialist.

2.3. The Case against Mysterianism

To repeat the last point, McGinn appears to assume that conscious states are *caused* by brain states. His argument does not go through if conscious states are simply *identical* to brain states. In other words, the argument does not go through unless any identity of conscious states with brain states is rejected.[12] But such rejection amounts to dualism. McGinn is thus committed to dualism.[13] On the view he presupposes, the conscious cannot be simply identified with the physical. Rather, there are two different kinds of state a person or organism may be in, brain states on the one hand and conscious states on the other.

Recall that McGinn's mysterianism is of the epistemological variety. The epistemological claim now appears to be conditional upon an ontological claim, namely dualism. So at the end of the day, as far as the *ontology* of consciousness is concerned, McGinn is a straightforward dualist. The plausibility of his (epistemological) mysterianism depends, to that extent, on the plausibility of (ontological) dualism. In the next section, we will consider the plausibility of dualism.

Before doing so, let us raise one more difficulty for mysterianism, and in particular the notion of cognitive closure. It is, of course, undeniable that rats do not understand trigonometry. But observe that trigonometric problems do not pose themselves to rats (Dennett 1995: 381-3). Indeed, it is precisely *because* rats do not understand trigonometry that trigonometric problems do not pose themselves to rats. For rats to grapple with trigonometric problems, they would have to understand quite a bit of trigonometry. Arguably, it is a mark of genuine cognitive closure that certain questions do not even pose themselves to the cognitively closed. The fact that certain questions about consciousness do pose themselves to humans may therefore indicate that humans are *not* cognitively closed to consciousness (or more accurately to the link between consciousness and the brain).[14]

3. Dualism

Traditionally, approaches to the ontology of mind and consciousness have divided into two main groups: monism and dualism. The former holds that there is one kind of stuff in the world, the latter that there are two.[15] Within monism, there is a further distinction between views that construe the single existing stuff as material and views that construe it as immaterial; the former are *materialist* views, the latter *idealist*.[16]

Descartes framed his dualism in terms of two different kinds of *substance* (where a substance is something that can in principle exist all by itself). One is the extended substance, or matter; the other is the thinking substance, or mind. A person, on this view, is a combination of two different *objects*: a body and a soul. A body and its corresponding soul "go together" for some stretch of time, but being two separate objects, their existence is independent and can therefore come apart.[17]

Modern dualism is usually of a more subtle sort, framed not in terms of *substances* (or *stuffs*), but rather in terms of *properties*. The idea is that even though there is only one kind of stuff, or substance, there are two kinds of properties, mental and physical, and neither can be

reduced to the other.[18] This is known as *property dualism*. A particularly cautious version of property dualism claims that while most mental properties are reducible to physical ones, conscious or phenomenal properties are irreducible.

3.1. Chalmers' Naturalistic Dualism

For many decades, dualistic arguments were treated mainly as a *challenge* to a physicalist worldview, not so much as a basis for a non-physicalist alternative. Thus dualism was not so much an explanation or account of consciousness, but rather the avoidance of one. This state of affairs has been rectified in the past decade or so, mainly through the work of David Chalmers (1995, 1996, 2002a).

Chalmers' theory of consciousness, which he calls *naturalistic dualism*, is stronger than ordinary dualism, in that it claims not only that phenomenal properties are not *identical* to physical properties, but also that they fail to *supervene* – at least with metaphysical or logical necessity[19] – on physical properties.[20] We tend to think, for instance, that biological properties necessarily supervene on physical properties, in the sense that two systems cannot possibly differ in their biological properties if all their physical properties are exactly similar. But according to Chalmers, phenomenal properties are different: two systems *can* be exactly the same physically but have different phenomenal properties.

At the same time, Chalmers does not take phenomenal properties to be accidental or random superpositions onto the physical world. On the contrary, he takes them to be causally grounded in physical laws. That is, instantiations of phenomenal properties are *caused* by instantiations of physical properties, and they are so caused in accordance with strict laws of nature.[21]

This means that phenomenal consciousness *can* be explained in physical terms. It is just that the explanation will not be a *reductive* explanation, but rather a *causal* explanation. To explain an event or phenomenon causally is to cite its cause, that is, to say what brought it about or gave rise to it.[22] According to Chalmers, one *could* in principle explain the instantiation of phenomenal properties by citing their physical causes.

A full theory of consciousness would uncover and enlist all the causal laws that govern the emergence of phenomenal properties from the physical realm. And a full description of nature and its behavior would have to include these causal laws on top of the causal laws obtained by "ultimate physics."[23]

Chalmers himself does not attempt to detail many of these laws. But he does propose a pair of principles to which we should expect such laws to conform. These are the "structural coherence" principle and the "organizational invariance" principle. The former concerns the sort of direct availability for global control that conscious states appear to exhibit, the latter the systematic correspondence between a system's functional organization and its phenomenal properties.[24]

3.2. The Case for Dualism

The best known arguments in favor of property dualism about consciousness are so-called "epistemic arguments." The two main ones are Frank Jackson's (1984) "Knowledge Argument" and Thomas Nagel's (1974) "what is it like" argument. Both follow a similar pattern. After painting forth a situation in which all physical facts about something are known, it is shown that some knowledge is still missing. It is then inferred that the missing knowledge must be knowledge of non-physical facts.

The Knowledge Argument proceeds as follows. Suppose a baby is kept in a black-and-white environment, so that she never has color experiences. But she grows to become an expert on color and color vision. Eventually, she knows all the physical facts about color and color vision. But when she sees red for the first time, she learns something new: she learns

what it is like to see red. That is, she acquires a new piece knowledge. Since she already knew all the physical facts, this new piece of knowledge cannot be knowledge of a physical fact. It is therefore knowledge of a non-physical fact. So, the fact thereby known (what it is like to see red) is a non-physical fact.

Nagel's argument, although more obscure in its original presentation, can be "formatted" along similar lines. We can know all the physical facts about bats without knowing what it is like to be a bat. It follows that the knowledge we are missing is not knowledge of a physical fact. Therefore, what it is like to be a bat is not a physical fact.

These arguments have struck many materialists as suspicious. After all, they infer an ontological conclusion from epistemological premises. This move is generally suspicious, but it is also vulnerable to a response that emphasizes what philosophers call the *intensionality* of epistemic contexts.[25] This has been the main response among materialists (Loar 1990, Tye 1986). The claim is that the Knowledge Argument's protagonist does not learn a new fact when she learns what it is like to see red, but rather learns an *old* fact in a new *way*; and similarly for the bat student.[26]

Consider knowledge that the evening star glows and knowledge that the morning star glows. These are clearly two different pieces of knowledge. But the fact thereby known is one and the same – the fact that Venus glows. Knowledge that *this* is what it is like to see red and knowledge that *this* is the neural assembly stimulated by the right wavelength may similarly constitute two separate pieces of knowledge that correspond to only one fact being known. So from the acquisition of a new piece of knowledge one cannot infer the existence of a new fact – and that is precisely the inference made in the above dualist arguments.[27,28]

A different argument for dualism that is widely discussed today is Chalmers' (1996) argument from the conceivability of zombies. Zombies are imaginary creatures which are physically indistinguishable from us but lack consciousness. We seem to be able to conceive of such creatures, and Chalmers wants to infer from this that materialism is false. The argument is often caricatured as follows:

1) Zombies are conceivable;
2) If A's are conceivable, then A's are (metaphysically) possible;[29] therefore,
3) Zombies are possible; but,
4) Materialism entails that zombies are not possible; therefore,
5) Materialism is false.

Or, more explicitly formulated:

1) For any physical property P, it is conceivable that P is instantiated but consciousness is not;
2) For any pair of properties F and G, if it is conceivable that F is instantiated when G is not, then it is (metaphysically) *possible* that F is instantiated when G is not; therefore,
3) For any physical property P, it is possible that P is instantiated and consciousness is not; but,
4) If a property F can be instantiated when property G is not, then F does not supervene on G;[30] therefore,
5) For any physical property P, consciousness does not supervene on P.

To this argument it is objected that the second premise is false, and the conceivability of something does not entail its possibility. Thus, we can conceive of water not being H_2O, but this is in fact impossible; Escher triangles are conceivable, but not possible.[31]

Chalmers' argument is more subtle than this, however. One way to get at the real argument is this.[32] Let us distinguish between the property of *being* water and the property of

appearing to be water, or being *apparent* water.[33] For a certain quantity of stuff to *be* water, it must be H_2O. But for it to *appear* to be water, it need only be clear, drinkable, liquid, etc. – or perhaps only *strike normal subjects* as clear, drinkable, liquid, etc. Now, although the unrestricted principle that conceivability entails possibility is implausible, a version of the principle restricted to what we may call *appearance properties* is quite plausible. Thus, if we can conceive of apparent water not being H_2O, then it is indeed possible that apparent water should not be H_2O.

Once the restricted principle is accepted, there are two ways a dualist may proceed. Chalmers' own argument seems to be more accurately captured as follows:[34]

1) For any physical property P, it is conceivable that P is instantiated but *apparent consciousness* is not;
2) For any pair of properties F and G, such that F is an *appearance* property, if it is conceivable that F is instantiated when G is not, then it is (metaphysically) possible that F is instantiated when G is not; therefore,
3) For any physical property P, it is possible that P is instantiated when apparent consciousness is not; but,
4) If a property F can be instantiated when property G is not, then F does not supervene on G; therefore,
5) For any physical property P, apparent consciousness does not supervene on P.

A materialist might want to reject this argument by denying Premise 2 (the restricted conceivability-possibility principle). Whether the restricted principle is true is something we cannot settle here. Note, however, that it is surely much more plausible than the corresponding unrestricted principle, and it is the only principle the argument for dualism really needs.

Another way the argument can be rejected is by denying the *existence* of such properties as *apparent water* and *apparent consciousness*.[35] More generally, perhaps, while "natural" properties such as being water or being conscious do exist, "unnatural" properties do not, and appearance properties are unnatural in the relevant sense.[36]

To avoid this latter objection, a dualist may proceed to develop the argument differently, claiming that in the case of consciousness, there is no distinction between appearance and reality (Kripke 1980). This would amount to the claim that the property of being conscious is identical to the property of appearing to be conscious. The conceivability argument then goes like this:

1) For any physical property P, it is conceivable that P is instantiated but apparent consciousness is not;
2) For any pair of properties F and G, such that F is an *appearance* property, if it is conceivable that F is instantiated when G is not, then it is (metaphysically) possible that F is instantiated when G is not; therefore,
3) For any physical property P, it is *possible* that P is instantiated when apparent consciousness is not; but,
4) If property F can be instantiated when property G is not, then F does not supervene on G; therefore,
5) For any physical property P, apparent consciousness does not supervene on P; but,
6) Consciousness = apparent consciousness; therefore,
7) For any physical property P, consciousness does not supervene on P.

Materialists may reject this argument by denying that there is no distinction between appearance and reality when it comes to consciousness (the sixth premise).

The debate over the plausibility of the various versions of the zombie argument continues. A full critical examination is impossible here. Let us move on, then, to consideration of the independent case against dualism.

3.3. The Case Against Dualism

The main motivation to avoid dualism continues to be the one succinctly worded by Smart (1959: 143) almost half a century ago: "It seems to me that science is increasingly giving us a viewpoint whereby organisms are able to be seen as physicochemical mechanisms: it seems that even the behavior of man himself will one day be explicable in mechanistic terms." It would be curious if consciousness stood out in nature as the only property that defied reductive explanation in microphysical terms. More principled arguments aside, this simple observation appears to be the chief motivating force behind naturalization projects that attempt to reductively explain consciousness and other recalcitrant phenomena.

As I noted above, against traditional dualists it was common to present the more methodological argument that they do not in fact propose any positive theory of consciousness, but instead rest content with arguing against existing materialist theories, and that this could not lead to real progress in the understanding of consciousness. This charge cannot be made against Chalmers, who does propose a positive theory of consciousness.

Chalmers' own theory is open to more substantial criticisms, however. In particular, it is arguably committed to epiphenomenalism about consciousness, the thesis that conscious states and events are *causally inert*. As Kim (1989a, 1989b, 1992) has pointed out, it is difficult to find causal work for non-supervenient properties. Assuming that the physical realm is *causally closed* (i.e., that every instantiation of a physical property has as its cause the instantiation of another physical property), non-supervenient properties must either (i) have no causal effect on the physical realm or (ii) causally overdetermine the instantiation of certain physical properties.[37] But since pervasive overdetermination can be ruled out as implausible, non-supervenient properties must be causally inert vis-à-vis the physical world. However, the notion that consciousness is causally inert, or *epiphenomenal*, is extremely counter-intuitive: we seem to ourselves to act on our conscious decisions all the time and at will.

In response to the threat of epiphenomenalism, Chalmers pursues a two-pronged approach.[38] The first prong is to claim that epiphenomenalism is *merely* counter-intuitive, but does not face serious *argumentative* challenges. This is not particularly satisfying, however: all arguments must come to an end, and in most of philosophy, the end is bound to be a certain intuition or intuitively compelling claim. As intuitions go, the intuition that consciousness is not epiphenomenal is very strong.

The second prong is more interesting. Chalmers notes that physics characterizes the properties to which it adverts in purely relational terms – essentially, in terms of the laws of nature into which they enter. The resulting picture is a network of interrelated nodes, but the intrinsic character of the thus-interrelated nodes remains opaque. It is a picture that gives us what Bertrand Russell once wittingly called "the causal skeleton of the world." Chalmers' suggestion is that phenomenal properties may constitute the intrinsic properties of the entities whose relational properties are mapped out by physics. At least this is the case with intrinsic properties of obviously conscious entities. As for apparently inanimate entities, their intrinsic properties may be crucially *similar* to the phenomenal properties of conscious entities. They may be, as Chalmers puts it, "protophenomenal" properties.

Although intriguing, this suggestion has its problems. It is not clear that physics indeed gives us only the causal skeleton of the world. It is true that physics *characterizes* mass in terms of its causal relations to other properties. But it does not follow that the

property thus characterized is nothing but a bundle of causal relations. More likely, the relational characterization of mass is what *fixes the reference* of the term "mass," but the referent itself is nonetheless an intrinsic property. The bundle of causal relations is the reference-fixer, not the referent. On this view of things, although physics characterizes mass in causal terms, it construes mass not as the *causing* of effects E, but rather as the *causer* (or just the *cause*) of E. It construes mass as the relatum, not the relation.

Furthermore, *if* physics did present us with the causal skeleton of the world, then *physical* properties would turn out to be epiphenomenal (or nearly so). As Block (1990b) argued, functional properties – properties of having certain causes and effects – are ultimately inert, because an effect is always caused by its cause, not by its causing. So if mass was the *causing* of E, rather than the *cause* of E, then E would not be caused by mass. It would be caused, rather, by the protophenomenal property that satisfies the relational characterization attached to mass in physics.[39] The upshot is that if mass *was* the causing of E, rather than the cause of E, mass would not have the causal powers we normally take it to have. More generally, if physical properties were nothing but bundles of causal relations, they would be themselves causally inert.[40]

Chalmers faces a dilemma, then: either he violates our strongly held intuitions regarding the causal efficacy of phenomenal properties, or he violates our strongly held intuitions regarding the causal efficacy of physical properties. Either way, half his world is epiphenomenal, as it were. In any event, as we saw above the claim that physical properties are merely bundles of causal relations – which therefore call for the postulation of phenomenal and protophenomenal properties as the putative causal relata – is implausible.

Problems concerning the causal efficacy of phenomenal properties will attach to any account that portrays them as non-supervenient upon, or even as non-reducible to, physical properties. These problems are less likely to rear their heads for reductive accounts of consciousness. Let us turn, then, to an examination of the main reductive accounts discussed in the current literature.

4. Representationalism

According to the Representational Theory of Consciousness – or for short, *representationalism* – the phenomenal properties of conscious experiences can be reductively explained in terms of the experiences' representational properties.[41] Thus, when I look up at the blue sky, what it is like for me to have my conscious experience of the sky is just a matter of my experience's representation of the blue sky. The phenomenal character of my experience can be identified with (or with an aspect of) its representational content.[42]

This would be a theoretically happy result, since we have a fairly good notion as to how mental representation may be itself reductively explained in terms of informational and/or teleological relations between neurophysiological states of the brain and physical states of the environment.[43] The reductive strategy here is two-stepped, then: first reduce phenomenal properties to representational properties, then reduce representational properties to informational and/or other physical properties of the brain.

4.1. Tye's PANIC Theory

Not every mental representation is conscious. For this reason, a representational account of consciousness must pin down more specifically the kind of representation that would qualify as conscious. The most worked out story in this genre is probably Michael Tye's (1992, 1995, 2000, 2002) "PANIC Theory."[44]

The acronym "PANIC" stands for Poised, Abstract, Non-conceptual, Intentional content. So for Tye, a mental representation qualifies as conscious when, and only when, its representational content is (a) intentional, (b) non-conceptual, (c) abstract, and (d) poised.

What all these qualifiers mean is not particularly important, though the properties of non-conceptuality and poise are worth pausing to explicate.[45]

The content of a conscious experience is non-conceptual in that the experience can represent properties for which the subject lacks the concept. My conscious experience of the sky represents the sky not simply as being blue, but as being a very specific shade of blue, say $blue_{17}$. And yet if I am presented a day later with two samples of very similar shades of blue, $blue_{17}$ and $blue_{18}$, I will be unable to recognize which shade of blue was the sky's. This suggests that I lack the concept of $blue_{17}$. If so, my experience's representation of $blue_{17}$ is non-conceptual.[46]

The property of poise is basically a functional role property: a content is poised when it is ready and available to make direct impact on the formation of beliefs and desires. Importantly, Tye takes this to distinguish conscious representation from, say, blindsighted representations. A square can be represented both consciously and blindsightedly. But only the conscious representation is poised to make a direct impact on the beliefs the subject subsequently forms.

PANIC theory is supposed to cover not only conscious *perceptual* experiences, but all manners of phenomenal experience: somatic, emotional, etc. Thus, a toothache experience represents tissue damage in the relevant tooth, and represents it intentionally, non-conceptually, abstractly, and with poise.[47]

4.2. The Master Argument for Representationalism

The main motivation for representationalism may seem purely theoretical: it holds the promise of reductive explanation of consciousness in well understood informational and/or teleological terms. Perhaps because of this, however, the argument that has been most influential in making representationalism popular is a non-theoretical argument, one that basically rests on a phenomenological observation. This is the observation of the so-called *transparency of experience*. It has been articulated in a particularly influential manner by Harman (1990), but goes back at least to Moore (1903).

Suppose you have a conscious experience of the blue sky. Your attention is focused on the sky. You then decide to turn your attention *away* from the sky and onto your *experience* of the sky. Now your attention is no longer focused on the sky, but rather on the experience thereof. What are you aware of? It seems that you are still aware of the blueness of the sky. Certainly you are not aware of some *second* blueness, which attaches to your experience rather than to the sky. You are not aware of any *intermediary* blue quality interposed between yourself and the sky.

It appears, then, that when you pay attention to your experience, the only thing you become aware of is which features of the external sky your experience *represents*. In other words, the only introspectively accessible properties of conscious experience are its representational properties.

The transparency of experience provides a straightforward argument for representationalism. The argument may be laid out as follows:

1) The only introspectively accessible properties of conscious experience are its representational properties;
2) The phenomenal properties of conscious experience are given by its introspectively accessible properties; therefore,
3) The phenomenal properties of conscious experience are given by its representational properties.

The first premise is the thesis of transparency, the second one is intended as a conceptual truth (about what we mean by "phenomenal"). The conclusion is representationalism.

Another version of the argument from transparency, one which Tye employs, centers on the idea that rejecting representationalism in the face of transparency would require one to commit to an "error theory."[48] This version may be formulated as follows:

1) The phenomenal properties of conscious experience *seem* to be representational properties;
2) It is unlikely that the phenomenal properties of conscious experience are radically different from what they seem to be; therefore,
3) It is likely that the phenomenal properties of conscious experience *are* representational properties.

Here the transparency thesis is again the first premise. The second premise is the claim that convicting experience of massive error is to be avoided. And the conclusion is representationalism.

4.3. The Case Against Representationalism

Most of the arguments that have been marshaled against representationalism are arguments by counter-example. Scenarios of varying degrees of fancifulness are adduced, in which allegedly (i) a conscious experience has no representational properties, or (ii) two possible experiences with different phenomenal properties have the same representational properties, or (iii) inversely, two possible experiences with the same phenomenal properties have different representational properties. For want of space, I will present only one representative scenario from each category.

Block (1996) argues that phosphene experiences are non-representational. These can be obtained by rubbing one's eyes long enough that when one opens them again, one "sees" various light bits floating about. Such experiences do not represent any external objects or features, according to Block.

In response, Tye (2000) claims that such experiences do represent – it is just that they *mis*represent. They misrepresent there to be small objects with phosphorescent surfaces floating around the subject's head.

A long debated case in which phenomenal difference is accompanied by representational sameness is due to Peacocke (1983). Suppose you stand in the middle of a mostly empty road. All you can see in front of you are two trees. The two trees, A and B, have the same size and shape, but A is twice as far from you as B. Peacocke claims that, being aware that the two trees are equisizable, you represent to yourself that they have the same properties. And yet B "takes up more of your visual field" than A, in a way that makes you experience the two trees differently. There is phenomenal difference without representational difference.

Various responses to this argument have been offered by representationalists. Perhaps the most popular is that although you represent the two trees to have the same size properties, you also represent them to have certain different properties – e.g., B is represented to portend a larger visual angle than A (DeBellis 1991, Harman 1990, Tye 2000). To be sure, you do not necessarily possess the concept of portending a visual angle. But recall that the content of experience can be construed as non-conceptual. So your experience can represent the two trees to portend different visual angles without employing the concept of portending a visual angle. Thus a representational difference is matched to the phenomenal difference.

Perhaps the most prominent alleged counter-example is Block's (1990a) *Inverted Earth* case. Inverted Earth is an imaginary planet just like earth, except that every object there has the color complimentary to the one it has here. We are to imagine that a subject is clothed with color-inverting lenses and shipped to Inverted Earth unbeknownst to her. The color inversions due to the lenses and to the world cancel each other out, so that her phenomenal

experiences remain the same. But externalism about representational content ensures that the representational content of her experiences eventually change.[49] Her bluish experiences now represent a yellow sky. When her sky experiences on Inverted Earth are compared to her earthly sky experience, it appears that the two groups are phenomenally the same but representationally different.

This case is still being debated in the literature, but there are two representationalist strategies for accommodating it. One is to argue that the phenomenal character also changes over time on Inverted Earth (Harman 1990); the other is to devise accounts of representational content that make the representational content of the subject's experiences remain the same on Inverted Earth, externalism notwithstanding (Tye 2000).[50]

There may be, however, a more principled difficulty for representationalism than the myriad counter-examples it faces.[51] Representationalism seems to construe the phenomenal character of conscious experiences purely in terms of the *sensuous qualities* they involve. But arguably there is more to phenomenal character than sensuous quality. In particular, there seems to be a certain mine-ness, or for-me-ness, to them.

One way to put it is as follows (Kriegel 2005a, Levine 2001, Smith 1986). When I have my conscious experience of the blue sky, there is a bluish way it is like for me to have my experience. A distinction can be drawn between two components of this "bluish way it is like for me": the bluish component, which we may call *qualitative character*, and the for-me component, which we may call *subjective character*. We may construe phenomenal character as the compresence of qualitative and subjective character. This subjective character, or for-me-ness, is certainly an elusive phenomenon, but it is present in every conscious experience. Indeed, its presence seems to be a condition of any phenomenality: it is hard to make sense of the idea of a conscious experience that does not have this for-me-ness to it. If it did not have this for-me-ness, it would be a mere sub-personal state, a state that takes place *in* me but is not *for* me in the relevant sense. Such a sub-personal state seems not to qualify as a conscious experience.

The centrality of subjective character (as construed here) to consciousness is something that has been belabored in the phenomenological tradition (see Thompson and Zahavi, this volume, and Zahavi 1999). The concept of *pre-reflective self-consciousness* – or a form of self-awareness that does not require focused and explicit awareness of oneself and one's current experience, but is rather built into that very experience – is one that figures centrally in almost all phenomenological accounts of consciousness.[52] But it has been somewhat neglected in analytic philosophy of mind.[53]

The relative popularity of representationalism attests to this neglect. While a representationalist account of sensuous qualities – what we have called qualitative character – may turn out to win the day (if the alleged counter-examples can be overcome), it would not provide us with any perspective on subjective character.[54] Therefore, even if representationalism turns out to be a satisfactory account of qualitative character, it is unlikely to be a satisfactory account of phenomenal consciousness proper.

5. Higher-Order Monitoring Theory

One theory of consciousness from analytic philosophy that *can* be interpreted as targeting subjective character is the higher-order monitoring theory. On this view, what makes a mental state conscious is the fact that the subject is *aware* of it in the right way. It is only when the subject is aware (in that way) of a mental state that the state becomes conscious.[55]

Higher-order monitoring theories tend to anchor consciousness in the operation of a monitoring device. This device monitors and scans internal states and events, and produces higher-order representations of some of them.[56] When a mental state is represented by such a higher-order representation, it is conscious. So a mental state M of a subject S is conscious when, and only when, S has another mental state, M*, such that M* is an appropriate

representation of M. The fact that M* represents M guarantees that there be something it is like *for S* to have M.[57]

Observe that on this view, what confers conscious status on M is something outside M, namely, M*. This is the reductive strategy of the higher-order monitoring theory. Neither M nor M* is conscious in and of itself, independently of the other state. It is their coming together in the right way that yields consciousness.[58]

Versions of the higher-order monitoring theory differ mainly in how they construe the monitoring device and/or the representations it produces. The most seriously worked out version is probably David Rosenthal's (1986, 1990, 2002a, 2002b). Let us take a closer look at his "higher-order thought" theory.

5.1. Rosenthal's Higher-Order Thought Theory

According to Rosenthal, a mental state is conscious when its subject has a suitable higher-order thought about it.[59] The higher-order state's being a *thought* is supposed to rule out, primarily, its being a *quasi-perceptual* state.

There is a long tradition, hailing from Locke, of construing the monitoring device as analogous in essential respects to a sense organ (hence as being a sort of "inner sense") and accordingly as producing mental states that are crucially similar to perceptual representations, and that may to that extent be called "quasi-perceptual." This sort of "higher-order perception theory" is championed today by Armstrong (1968, 1981) and Lycan (1987, 1996). Rosenthal believes that this is a mistake, and the higher-order states that confer consciousness are not analogous to perceptual representations.[60] Rather, they are intellectual, or cognitive, states – that is, thoughts.

Another characteristic of thoughts – besides being non-perceptual – is their being *assertoric*. An assertoric state is one that has a *thetic*, or mind-to-world, direction of fit.[61] This is to be contrasted with states (such as wanting, hoping, disapproving, etc.) that have primarily a *telic*, or world-to-mind, direction of fit.[62] A third characteristic of thoughts – at least the kind suitable for conferring consciousness – is that they are *occurrent* mental states.[63]

Crucially, a suitable higher-order thought would also have to be *non-inferential*, in that it could not be the result of a conscious inference from the lower-order state (or from any other state, for that matter).[64] To be sure, the thought is formed through some process of information processing, but that process must be automatic and unconscious. This is intended to reflect the *immediacy*, or at least *felt* immediacy, of our awareness of our conscious states.[65] The fact that my experience of the sky has for-me-ness entails that I am somehow aware of its occurrence; but not any sort of awareness would do – very mediated forms of awareness cannot confer conscious status on their objects.

One last characteristic Rosenthal ascribes to the "suitable" higher-order representation is that it represents the lower-order state as a state *of oneself*. Its content must be, as this is sometimes put, *de se* content.[66] So the content of the higher-order representation of my conscious experience of the sky is not simply something like "this bluish experience is taking place," but rather something like "I myself am having this bluish experience."[67]

It is worth noting that according to Rosenthal the second-order representation is normally an *unconscious* state. To be sure, it need not necessarily be: in the more introspective, or reflective, episodes of our conscious life, the second-order state becomes itself conscious. It is then accompanied by a *third*-order state, one that represents its occurrence in a suitable way. When I explicitly introspect and dwell on my conscious experience of the sky, there are three separate states I am in: the (first-order) experience, a (second-order) awareness of the experience, and a (third-order) representation of that awareness. When I stop introspecting and turn my attention back to the sky, however, the third-order state evaporates and consequently the second-order state becomes unconscious

again. In any event, at any one time the subject's highest-order state, the one that confers consciousness on the chain of lower-order states "below" it, is unconscious.[68]

In summary, Rosenthal's central thesis is that a mental state is conscious just in case the subject has a non-perceptual, non-inferential, assertoric, *de se*, occurrent representation of it. This account of consciousness is not intended as an account of introspective or reflective consciousness, but of regular, everyday consciousness.

5.2. The Master Argument for Higher-Order Monitoring Theory

The master argument for the higher-order monitoring approach to consciousness has been succinctly stated by Lycan (2001):

1) A mental state M of subject S is conscious when, and only when, S is aware of M in the appropriate way;
2) Awareness of X requires mental representation of X; therefore,
3) M is conscious when, and only when, S has a mental state M*, such that M* represents M in the appropriate way.

Although the second premise is by no means trivial, it is the first premise that has been the bone of contention in the philosophical literature (see, e.g., Dretske 1993).

One can defend the claim that conscious states are states we are aware of having simply as a piece of conceptual analysis – as a platitude reflecting the very meaning of the word "conscious" (Lycan 1996). To my ear, this sounds right: a mental state of which the subject is completely unaware is a sub-personal, and therefore unconscious, state.

To some, however, this seems plainly false. When I have an experience of the sky, I am attending to the sky, they stress, not to myself and my internal goings-on. By consequence, I am aware of the sky, not of my experience of the sky. I am aware *through* my experience, not *of* my experience.

This objection seems to rely, however, on an unwarranted assimilation of awareness and attention. There is a distinction to be made between attentive awareness and inattentive awareness. If S attends to X and not to Y, it follows that S is not attentively aware of Y, but it does not follow that S is completely unaware of Y. For S may still be inattentively aware of Y.

Consider straightforward visual awareness. The distinction between foveal vision and peripheral vision means that our visual awareness at any one time has a periphery as well as a focal center. Right now, I am (visually) focally aware of my laptop, but also (visually) peripherally aware of an ashtray at the far corner of my desk. A similar distinction applies to perceptual awareness in other modalities: I am now (auditorily) focally aware of Duke Ellington's voice and (auditorily) peripherally aware of the air-conditioner's hum overhead.

There is no reason to think that a similar distinction would *not* apply to higher-order awareness. In reflective moods I may be focally aware of my concurrent experiences and feelings, but on other occasions I am just peripherally aware of them. The former is an attentive form of second-order awareness, the latter an inattentive one. Again, from the fact that it is inattentive it would be fallacious to infer that it is no awareness at all.

When it is claimed that conscious states are states we are aware of, the claim is not that we are focally aware of every conscious state we are in. That is manifestly false: the focus of our attention is mostly on the outside world. The claim is rather that we are at least peripherally aware of every conscious state we are in.[69] As long as M is conscious, S is aware, however dimly and inattentively, of M. Once S's awareness of M is extinguished altogether, M drops into the realm of the unconscious. This seems highly plausible on both conceptual and phenomenological grounds.[70]

5.3. The Case Against Higher-Order Monitoring Theory

Several problems for the monitoring theory have been continuously debated in the philosophical literature. I will focus on what I take to be the main three.[71]

The first is the problem of animal and infant consciousness. It is intuitively plausible to suppose that cats, dogs, and human neonates are conscious, that is, have conscious states; but it appears empirically implausible that they should have second-order representations (Lurz 1999). The problem is particularly acute for Rosenthal's account, since it is unlikely that these creatures can have *thoughts*, and moreover of the complex form "I myself am enjoying this milk."

There are two ways to respond to this objection. One is to deny that having such higher-order representations requires a level of sophistication of an order unlikely to be found in (say) cats. Thus, Rosenthal (2002b) claims that whereas *adult human* higher-order thoughts tend to be conceptually structured and employ a rich concept of self, these are not *necessary* features of such thoughts. There could be higher-order thoughts that are conceptually simple and employ a rudimentary concept of self, one that consists merely in the ability to distinguish oneself from anything that is not oneself. It may well turn out that worms, woodpeckers, or even day-old humans lack even this level of conceptual sophistication – in which case we would be required to deny them consciousness – but it is unlikely that cats, dogs, and *year*-old humans lack them.

The second possible line of response is to dismiss the intuition that animals such as cats, dogs, and even monkeys, do in fact have conscious states. Thus, Carruthers (1998, 1999) claims that there is a significant amount of projection that takes place when we ascribe to, say, our pets conscious states. In reality there is very little evidence to suggest that they have not only perceptual and cognitive states, but conscious ones.

Both lines of response offer some hope to the defender of higher-order monitoring, but also implicate her theory in certain counter-intuitive and *prima facie* implausible claims. Whether these could somehow be neutralized, or accepted as outweighed by the theoretical benefits of higher-order monitoring theory, is something that is very much under debate.

Perhaps more disturbing is the problem of so-called "targetless" higher-order thoughts (or more generally, representations). When someone falsely believes that the almond tree in the backyard is blooming again, there are two ways she may get things wrong: (i) it may be that the backyard almond tree is *not* blooming, or (ii) it may be that there is no almond tree in the backyard (blooming or not). Let us call a false belief of type (ii) a *targetless* thought. The higher-order monitoring theory gets into trouble when a subject has a targetless higher-order thought (Byrne 1997).[72] Suppose at a time *t* subject S thinks (in the suitable way) that she has a throbbing toothache, when in reality she has no toothache at all (throbbing or not). According to higher-order monitoring theory, what it is like for S at *t* is the way it is like to have a throbbing toothache, even though S has no toothache at *t*. In other words, if S has an M* that represents M when in reality there is no M,[73] S will be under the impression that she is in a conscious state when in reality she is not. (She is not in a conscious state because M does not exist, and it is M that is supposed to bear the property of being conscious.) Moreover, on the assumption that a person is conscious at a time *t* only if she has at least one conscious state at *t*,[74] this would entail that when a subject harbors a targetless higher-order misrepresentation, she is not conscious, even though it feels to her as though she is. This is a highly counter-intuitive consequence: we want to say that a person cannot be under the impression that she is conscious when she is not.

There are several ways higher-order monitoring theorists may respond to this objection. Let us briefly consider three possible responses.

First, they may claim that when M* is targetless, the property of being conscious, although not instantiated by M, is instantiated by M*. But as we saw above, according to their view, M* is normally unconscious. So to say that M* instantiates the property of being

conscious would be to say that it is, in the normal case, both conscious and not conscious – which is incoherent.[75]

Second, they may claim that the property of being conscious is, in reality, not a property of the discrete state M, but rather attaches itself to the compound of M and M*.[76] But this will not work either, because higher-order monitoring theory would then face a dilemma: either the compound state M + M* is a state we are aware of having, or it is not; if it is not, then the higher-order monitoring theory is false, since it claims that conscious states are states we *are* aware of having; and if it is, then according to the theory it must be represented by a third-order mental state, M**, in which case the same problem would recur when M** is targetless.

Third, they may claim that there are no targetless higher-order representations. But even if this can be shown to be the actual case (and it is hard to imagine how this would be done), we can surely conceive of counterfactual situations in which targetless higher-order representations do occur.[77]

A third problem for the higher-order monitoring theory is its treatment of the *epistemology* of consciousness (Goldman 1993b, Kriegel 2003b). Our knowledge that we are in a conscious state is first-person knowledge, knowledge that is not based on inference from experimental, or theoretical, or third-personal evidence. But if the higher-order monitoring theory were correct, what would make our conscious states conscious is (normally) the occurrence of some unconscious state (i.e., the higher-order representation), so in order to know that we are in a conscious state we would need to know of the occurrence of that unconscious state. But knowledge of unconscious states is necessarily theoretical and third-personal, since we have no direct acquaintance with our unconscious states.

Another way to put the argument is this. How does the defender of higher-order monitoring theory know that conscious states are states we are aware of? It does not seem to be something she knows on the basis of experimentation and theorization. Rather, it seems to be intuitively compelling, something that she knows on the basis of first-person acquaintance with her conscious states. But if the higher-order monitoring theory were correct, it would seem that that knowledge would have to be purely theoretical and third-personal. So construed, this "epistemic argument" against higher-order monitoring theory (HOMT) may be formulated as follows:

1) If HOMT were correct, our awareness of our conscious states would normally be an unconscious state; that is,
2) We do not have non-theoretical, first-person knowledge of our unconscious states; therefore,
3) If HOMT were correct, we would not have non-theoretical, first-person knowledge of the fact that we are aware of our conscious states; but,
4) We do have non-theoretical, first-person knowledge of the fact that we are aware of our conscious states; therefore,
5) HOMT is incorrect.

The upshot of the argument is that the awareness of our conscious states must in the normal case be itself a conscious state. This is something that the higher-order monitoring theory cannot allow, however, since within its framework it would lead to infinite regress. The problem is to reconcile the claim that conscious states are states we are aware of having with the notion that we have non-theoretical knowledge of this fact.

6. The Self-Representational Theory of Consciousness

One approach to consciousness that has a venerable tradition behind it, but has only very recently regained a modest degree of popularity, is what we may call the "self-

representational theory." On this view, mental states are conscious when, and only when, they represent their own occurrence. Thus, my conscious experience of the blue sky represents both the sky and itself – and it is *in virtue* of representing itself that it *is* a conscious experience.

Historically, the most thorough development and elucidation of the self-representational theory is Brentano's (1874). Through his work, the view has had a significant influence in the phenomenological tradition. But apart from a couple of exceptions – Lehrer (1996, 1997) and Smith (1986, 1989) come to mind – the view had enjoyed virtually no traction in Anglo-American philosophy. Recently, however, versions of the view, and close variations on it, have been defended by a number of philosophers.[78]

Rather than focus on any one particular account of consciousness along these lines, I will now survey the central contributions to the understanding of consciousness in terms of self-representation.

6.1. Varieties of Self-Representational Theory

Brentano held that every conscious state is intentionally directed at two things. It is *primarily* directed at whatever object it is about, and it is *secondarily* directed at itself. My bluish sky experience is directed primarily at the sky and secondarily at itself. In more modern terminology, a conscious state has two representational contents: an other-directed (primary) content and a self-directed (secondary) content. Thus, if S consciously fears that *p*, S's fear has two contents: the primary content is *p*, the secondary content is itself, the fear that *p*. The distinction between primary intentionality and secondary intentionality is presumably intended to capture the difference (discussed above) between attentive or focal awareness and inattentive or peripheral awareness.[79]

Caston (2002) offers an interesting gloss on this idea in terms of the type/token distinction. For Caston, S's conscious fear that *p* is a single *token* state that falls under two separate state *types*: the fear-that-*p* type and the awareness-of-fear-that-*p* type. The state has two contents, arguably, precisely in virtue of falling under two types.

Brook and Raymont (forthcoming) stress that the self-representational content of the conscious state is not simply that the state occurs, but rather that it occurs *within oneself* – that it is one's own state. Just as Rosenthal construed the content of higher-order states as "I myself am having that state," so Brook and Raymont suggest that the full self-representational content of conscious states is something like "I myself am herewith having this very state."[80]

For Brentano and his followers, the self-directed element in conscious states is an aspect of their intentionality, or content. In David Woodruff Smith's (1986, 2004) "modal account," by contrast, the self-directed element is construed not as an aspect of the representational content, but rather as an aspect of the representational attitude (or mode). When S consciously fears that *p*, it is not in virtue of figuring in its own secondary content that the fear is conscious. Indeed, S's fear does not *have* a secondary content. Its only content is *p*. The "reflexive character" of the fear, as Smith puts it, is rather part of the *attitude* S takes toward *p*. Just as the attitudes toward *p* can vary from fear, hope, expectation, etc., so they can vary between self-directed or "reflexive" fear and un-self-directed or "irreflexive" fear. S's fear that *p* is conscious, on this view, because S takes the attitude of self-directed fear toward *p*.[81, 82]

One way in which the self-representational thesis can be relaxed to make a subtler claim is the following. Instead of claiming that a mental state M of a subject S is conscious just in case M represents itself, the thesis could be that M is conscious just in case S has an M* that is a representation of M and there is a *constitutive, non-contingent* relation between M and M*.[83] One constitutive relation is of course identity. So one version of this view would be that M is conscious just in case M is identical with M* – this is how Hossack (2002) formulates his thesis – and this seems to amount to the claim that M is conscious just in case

it represents itself (constitutes a representation of itself). But the point is that there are other, weaker constitutive relations that fall short of full identity.

One such relation is the part-whole relation. Accordingly, one version of the view, the one defended by Gennaro (1996, 2006), holds that M* is a *part of* M; another version, apparently put forth by Kobes (1995), holds that M is part of M*; and yet another version, Van Gulick's (2001, 2004, 2006), holds that M is conscious when it has two parts, one of which represents the other.

In Van Gulick's "higher-order global states theory," S's fear that *p* becomes conscious when the fear and S's awareness of the fear are somehow integrated into a single, unified state. This new state supersedes its original components, though, in a way that makes it a genuine unity, rather than a sum of two parts, one of which happens to represent the other. The result is a state that, if it does not represent itself, does something very close to representing itself.[84]

6.2. The Master Argument for the Self-Representational Theory

The basic argument for the self-representational approach to consciousness is that it is the only way to accommodate the notion that conscious states are states we are aware of without falling into the pitfalls of higher-order monitoring theory.

The argument can be organized, then, as a disjunctive syllogism that starts from the master argument for higher-order monitoring theory but then goes beyond it:

1) A mental state M of subject S is conscious when, and only when, S is aware of M;
2) Awareness of X requires mental representation of X; therefore,
3) M is conscious when, and only when, S has a mental state M*, such that M* represents M.
4) Either M* = M or M* ≠ M;
5) There are good reasons to think that it is not the case that M* ≠ M; therefore,
6) There are good reasons to think that it is the case that M* = M; therefore,
7) Plausibly, M is conscious when, and only when, M is self-representing.

The fourth premise could also be formulated as "either M* and M do not entertain a constitutive, non-contingent relation, or they do." The conclusion of the relevantly modified argument would then be the thesis that M is conscious when, and only when, S has a mental state M*, such that (i) M* represents M, and (ii) there is a constitutive, non-contingent relation between M and M*.

The fallacy in the master argument for higher-order monitoring theory is the supposition that if S is aware of M, then S must be so aware in virtue of being in a mental state that is *numerically different* from M. This supposition is brought to the fore and rejected in the argument just sketched.

The case for the fifth premise consists in all the reasons to be suspicious of the higher-order monitoring theory, as elaborated in §5.3 above, although it must also be shown that the same problems do not bedevil the self-representational theory as well.

Consider first the epistemic argument. We noted that higher-order monitoring theory fails to account for the non-theoretical, first-personal knowledge we have of the fact that we are aware of our conscious states. This is because it construes this awareness as normally an unconscious state. The self-representational theory, by contrast, construes this awareness as a conscious state, since it construes the awareness as the same state, or part of the state, of which one is thereby aware. So the self-representational theory, unlike the higher-order monitoring theory, *can* provide for the right epistemology of consciousness.

Consider next the problem of targetless higher-order representations. Recall, the problem ensues from the fact that M* could in principle misrepresent not only that M is F

when in reality M is not F, but also that M is F when in reality there is no M. The same problem does not arise for self-representing states, however: although M could in principle misrepresent itself to be F when in reality it is not F, it could not possibly misrepresent itself to be F when in reality it does not exist. For if it did not exist it could not represent anything, itself included. Thus the problem of targetless higher-order representations has no bite against the self-representational theory.

These are already two major problems that affect gravely the plausibility of higher-order monitoring theory but do not apply to the self-representational theory. They make a strong *prima facie* case for the fifth premise above. The fourth premise is a logical truism and the first and second ones were defended in §5.2 above. So the argument appears to go through.

6.3. Problems for the Self-Representational Theory

One problem that does persist for the self-representational theory is the problem of animal consciousness. The ability to have self-representing states presumably requires all the conceptual sophistication that the ability to have higher-order monitoring states does (since the self-representational content of a conscious state is the same as the representational content a higher-order monitoring state would have), perhaps even greater sophistication.[85]

Another problem is the elucidation and viability of the notion of self-representation. What does it mean for a mental state to represent itself, and what sort of mechanism could subserve the production of self-representing states? There is something at least initially mysterious about the notion of a self-representing state that needs to be confronted.

In fact, one might worry that there are principled reasons why self-representation is incompatible with any known naturalist account of mental representation. These accounts construe mental representation as some sort of natural relation between brain states and world states. Natural relations, as opposed to conceptual or logical ones, are based on causality and causal processes. But causality is an anti-reflexive relation, that is, a relation that nothing can bear to itself. Thus no state can bring about its own occurrence or give rise to itself. The argument can be formulated as follows:

(1) Mental representation involves a causal relation between the representation and the represented;
(2) The causal relation is anti-reflexive; therefore,
(3) No mental state can cause itself; and therefore,
(4) No mental state can represent itself.

The basic idea being that there is no conceivable naturalist account of mental representation that could allow for self-representing mental representations.

Even more fundamentally, one may worry whether the appeal to self-representation really *explains* consciousness. Perhaps self-representation is a *necessary* condition for consciousness, but why think it is also a *sufficient* condition? A sentence such as "this very sentence contains six words" is self-representing, but surely there is nothing it is like to be that sentence.[86]

It may be responded to this last point that what is required for consciousness is *intrinsic* or *original* self-representation, not *derivative* self-representation.[87] Sentences and linguistic expressions do not have any representational content in and of themselves, independently of being interpreted. But plausibly, mental states do.[88] The same goes for self-representational content: sentences and linguistic expressions may be derivatively self-representing, but only mental states can be non-derivatively self-representing. A more accurate statement of the self-representation theory is therefore this: A mental state M of a subject S is conscious when, and only when, M is non-derivatively self-representing.

Still, self-representing zombies are readily conceivable. It is quite easy to imagine unconscious mental states in our own cognitive system – say, states formed early on in visual processing – that represent themselves without thereby being conscious.[89] Furthermore, it is easy to imagine a creature with no conscious awareness whatsoever who harbors mental states that represent themselves. Thus Chalmers' zombie argument can be run in a particularized version directed specifically against the self-representational theory.[90]

Conclusion: Directions for Future Research

Much of the philosophical discourse on consciousness is focused on the issue of reducibility. As we just saw, the zombie argument and other dualist arguments can be tailored to target any particular reductive account of consciousness. This debate holds great intrinsic importance, but it is important to see that progress toward a scientific explanation of consciousness can be made without attending to it.

All three reductive approaches to consciousness we considered – the representational, higher-order monitoring, and self-representational theories – can readily be refashioned as accounts not of consciousness itself, but of the *emergence base* (or *causal basis*) of consciousness. Instead of claiming that consciousness *is* (or is *reducible to*) physical structure P, the claim would be that consciousness *emerges from* (or is *brought about by*) P. To make progress toward the scientific explanation of consciousness, we should focus mainly on what the right physical structure is – what P is. Whether P is consciousness itself or only the emergence base of consciousness is something we can set aside for the purposes of scientific explanation. If it turns out that P is consciousness itself (as the reductivist holds), then we will have obtained a *reductive* explanation of consciousness; if it turns out that P is only the emergence base of consciousness (as the dualist holds), then we will have obtained a *causal* explanation of consciousness. But both kinds of explanation are bona fide scientific explanations.

In other words, philosophers could usefully reorganize their work on consciousness around a distinction between two separate issues or tasks. The first task is to devise a positive account of the physical (or more broadly, natural) correlate of consciousness, without prejudging whether it will constitute a reduction base or merely an emergence base. Work along these lines will involve modifying and refining the representational, higher-order monitoring, and self-representational theories and/or devising altogether novel positive accounts. The second task is to examine the a priori and a posteriori cases for reducibility. Work here will probably focus on the issue of how much can be read off of conceivability claims, as well as periodic reconsideration of the plausibility of conceivability claims in light of newer and subtler positive accounts of consciousness.[91]

Another front along which progress can certainly be made is tightening the connection between the theoretical and experimental perspectives on consciousness. Ultimately, one hopes that experiments could be designed that would test well defined empirical consequences of philosophical (or more generally, purely theoretical) models of consciousness. This would require philosophers to be willing to put forth certain empirical speculations, as wild as these may seem, based on their theories of consciousness; and experimental scientists to take interest in the intricacies of philosophical theories in attempt to think up possible ways to test them.

All in all, progress in our understanding of consciousness and the outstanding methodological and substantive problems it presents has been quite impressive over the past two decades. The central philosophical issues are today framed with a clarity and precision that allows a corresponding level of clarity and precision in our thinking about consciousness. Even more happily, there is no reason to suppose that this progress will come to a halt or slow down in the near future.[92]

References

- Armstrong, D. M. 1968. *A Materialist Theory of the Mind*. New York: Humanities Press.
- Armstrong, D. M. 1978. *A Theory of Universals*, vol. 2. Cambridge: Cambridge UP.
- Armstrong, D. M. 1981. "What Is Consciousness?" In his *The Nature of Mind*. Reprinted in Block et al. (1997).
- Anscombe, G. E. M. 1957. *Intention*. Oxford: Blackwell.
- Baars, B. 1988. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge UP.
- Baars, B. 1997. *In the Theater of Consciousness: The Workspace of the Mind*. Oxford and New York: Oxford UP.
- Block, N. J. 1990a. "Inverted Earth." *Philosophical Perspective* 4: 52-79.
- Block, N. J. 1990b. "Can the Mind Change the World?" In G. Boolos (ed.), *Meaning and Method: Essays in Honor of Hilary Putnam*. Cambridge: Cambridge UP.
- Block, N. J. 1995. "On a Confusion About the Function of Consciousness." *Behavioral and Brain Sciences* 18: 227-247. Reprinted in Block et al. (1997).
- Block, N. J. 1996. "Mental Paint and Mental Latex." in *Philosophical Issues* 7: 19-50.
- Block, N. J., O. Flanagan, and G. Guzeldere (eds.) 1997. *The Nature of Consciousness: Philosophical Debates*. Cambridge MA: MIT Press.
- Brentano, F. 1874. *Psychology from Empirical Standpoint*. Edited by O. Kraus. English edition L. L. McAlister. Traslated by A. C. Rancurello, D. B. Terrell, and L. L. McAlister. London: Routledge and Kegan Paul, 1973.
- Brook, A. and P. Raymont. *A Unified Theory of Consciousness*. Book manuscript.
- Brueckner, A. and E. Berukhim 2003. "McGinn on Consciousness and the Mind-Body Problem." In Q. Smith and D. Jokic (eds.), *Consciousness: New Philosophical Perspectives*. Oxford and New York: Oxford UP.
- Byrne, D. 1997. "Some Like It HOT: Consciousness and Higher Order Thoughts." *Philosophical Studies* 86: 103-129.
- Byrne, A. 2001. "Intentionalism Defended." *Philosophical Review* 110: 199-240.
- Carruthers, P. 1989. "Brute Experience." *Journal of Philosophy* 85: 258-269.
- Carruthers, P. 1996. *Language, Thought, and Consciousness*. Cambridge: Cambridge UP.
- Carruthers, P. 1998. "Natural Theories of Consciousness." *European Journal of Philosophy* 6: 203-222.
- Carruthers, P. 1999. "Sympathy and Subjectivity." *Australasian Journal of Philosophy* 77: 465-482.
- Carruthers, P. 2000. *Phenomenal Consciousness*. Cambridge: Cambridge UP.
- Carruthers, P. 2006. "Conscious Experience Versus Conscious Thought." In U. Kriegel and K. Williford (eds.), *Consciousness and Self-Reference*. Cambridge MA: MIT Press.
- Castaˊeda, H.-N. 1966. "'He': A Study in the Logic of Self-Consciousness." *Ratio* 8: 130-157.
- Caston, V. 2002. "Aristotle on Consciousness." *Mind* 111: 751-815.
- Chalmers, D. J. 1995. "Facing Up to the Problem of Consciousness." *Journal of Consciousness Studies* 2: 200-219.
- Chalmers, D. J. 1996. *The Conscious Mind*. Oxford and New York: Oxford UP.
- Chalmers, D. J. 2002a. "Consciousness and Its Place in Nature." In D. J. Chalmers (ed.), *Philosophy of Mind*. Oxford and New York: Oxford UP.
- Chalmers, D. J. 2002b. "The Components of Content." In D. J. Chalmers (ed.), *Philosophy of Mind*. Oxford and New York: Oxford UP.
- Chisholm, R. 1966. *A Theory of Knowledge*. Englewood Cliffs, NJ: Prentice-Hall.
- Churchland, P. M. 1979. *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge UP.
- Churchland, P. M. 1985. "Reduction, Qualia, and the Direct Introspection of Brain States." *Journal of Philosophy* 82: 8-28.
- Crick, F. and C. Koch 1990. "Towards a Neurobiological Theory of Consciousness." *Seminars in the Neurosciences* 2: 263-275. Reprinted in Block et al. (1997).
- Crick, F. and C. Koch 2003. "A Framework for Consciousness." *Nature Neuroscience* 6: 119-126.
- Cummins, R. 1979. "Intention, Meaning, and Truth Conditions." *Philosophical Studies* 35: 345-360.
- DeBellis, M. 1991. "The Representational Content of Musical Experience." *Philosophy and Phenomenological Research* 51: 303-324.
- Dennett, D. C. 1969. *Consciousness and Content*. London: Routledge.
- Dennett, D. C. 1981. "Towards a Cognitive Theory of Consciousness." In his *Brainstorms*, Brighton: Harvester.
- Dennett, D. C. 1987. *The Intentional Stance*. Cambridge MA: MIT Press.
- Dennett, D. C. 1991. *Consciousness Explained*. Cambridge MA: MIT Press.
- Dennett, D. C. 1995. *Darwin's Dangerous Idea*. New York: Simon and Schuster.
- Dretske, F. I. 1981. *Knowledge and the Flow of Information*. Oxford: Clarendon.

- Dretske, F. I. 1988. *Explaining Behavior*. Cambridge MA: MIT Press.
- Dretske, F. I. 1993. "Conscious Experience." *Mind* 102: 263-283.
- Dretske, F. I. 1995. *Naturalizing the Mind*. Cambridge MA: MIT Press.
- Fodor, J. A. 1974. "Special Sciences." *Synthese* 28: 97-115.
- Foster, J. 1982. *The Case for Idealism*. London: Routledge.
- Gennaro, R. J. 1996. *Consciousness and Self-Consciousness*. Philadelphia/Amsterdam: John Benjamin Publishing Co.
- Gennaro, R. J. 2002. "Jean-Paul Sartre and the HOT Theory of Consciousness." *Canadian Journal of Philosophy* 32: 293-330.
- Gennaro, R. J. 2006. "Between Pure Self-Referentialism and (Extrinsic) HOT Theory." In Kriegel and Williford 2006.
- Goldman, A. 1993a. "The Psychology of Folk Psychology." *Behavioral and Brain Sciences* 16: 15-28.
- Goldman, A. 1993b. "Consciousness, Folk Psychology, and Cognitive Science." *Consciousness and Cognition* 2: 364-383.
- Grice, P. 1957. "Meaning." *Philosophical Review* 66: 377-388.
- Guzeldere, G. 1995. "Is Consciousness the Perception of What Passes in One's Own Mind?" In T. Metzinger (ed.), *Conscious Experience*. Padborn: Schoeningh-Verlag. Reprinted in Block et al. (1997).
- Harman, G. 1990. "The Intrinsic Quality of Experience." *Philosophical Perspectives* 4: 31-52. Reprinted in Block et al. (1997).
- Horgan, T. and J. Tienson 2002. "The Intentionality of Phenomenology and the Phenomenology of Intentionality." In D. J. Chalmers (ed.), *Philosophy of Mind*. Oxford and New York: Oxford UP.
- Hossack, K. 2002. "Self-Knowledge and Consciousness." *Proceedings of the Aristotelian Society* 2002: 163-181.
- Jackson, F. 1984. "Epiphenomenal Qualia." *Philosophical Quarterly* 34: 147-152.
- Kim, J. 1989a. "The Myth of Nonreductive Materialism." *Proceedings and Addresses of the American Philosophical Association* 63: 31-47.
- Kim, J. 1989b. "Mechanism, Purpose, and Explanatory Exclusion." *Philosophical Perspectives* 3: 77-108.
- Kim, J. 1992. "Multiple Realization and the Metaphysics of Reduction." *Philosophy and Phenomenological Research* 52: 1-26.
- Kobes, B. W. 1995. "Telic Higher-Order Thoughts and Moore's Paradox." *Philosophical Perspectives* 9: 291-312.
- Kriegel, U. 2002a. "Phenomenal Content." *Erkenntnis* 57: 175-198.
- Kriegel, U. 2002b. "Emotional Content." *Consciousness and Emotion* 3: 213-230.
- Kriegel, U. 2002c. "PANIC Theory and the Prospects for a Representational Theory of Phenomenal Consciousness." *Philosophical Psychology* 15: 55-64.
- Kriegel U. 2003a. "Consciousness, Higher-Order Content, and the Individuation of Vehicles." *Synthese* 134: 477-504.
- Kriegel, U. 2003b. "Consciousness as Intransitive Self-Consciousness: Two Views and an Argument." *Canadian Journal of Philosophy* 33: 103-132.
- Kriegel, U. 2004a. "The New Mysterianism and the Thesis of Cognitive Closure." *Acta Analytica* 18: 177-191.
- Kriegel, U. 2004b. "Consciousness and Self-Consciousness." *The Monist* 87: 185-209.
- Kriegel, U. 2005a. "Naturalizing Subjective Character." Forthcoming in *Philosophy and Phenomenological Research*.
- Kriegel, U. 2005b. "Review of Jeffrey Gray, *Consciousness: Creeping up on the Hard Problem*." *Mind* 114: 417-421.
- Kriegel, U. 2006a. "The Same-Order Monitoring Theory of Consciousness." In Kriegel and Williford 2006.
- Kriegel, U. 2006b. "The Concept of Consciousness in the Cognitive Sciences: Phenomenal Consciousness, Access Consciousness, and Scientific Practice." In P. Thagard (ed.), *Handbook of the Philosophy of Psychology and Cognitive Science*. Amsterdam: North-Holland.
- Kriegel, U. and K. Williford (eds.). *Self-Representational Approaches to Consciousness*. Cambridge MA: MIT Press.
- Kripke, S. 1980. "The Identity Thesis." In his *Naming and Necessity*. Reprinted in Block et al. (1997).
- Lehrer, K. 1996. "Skepticism, Lucid Content, and the Metamental Loop." In A. Clark, J. Ezquerro, and J. M. Larrazabal (eds.), *Philosophy and Cognitive Science*. Dordrecht: Kluwer.
- Lehrer, K. 1997. *Self-Trust: A Study of Reason, Knowledge, and Autonomy*. Oxford and New York: Oxford UP.
- Levine, J. 1983. "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* 64: 354-361.
- Levine, J. 2001. *Purple Haze: The Puzzle of Consciousness*. Oxford and New York: Oxford UP.
- Lewis, D. K. 1972. "Psychophysical and Theoretical Identifications." *Australasian Journal of Philosophy* 50: 249-258.

- Lewis, D. K. 1988. "What Experience Teaches." In W. G. Lycan (ed.), *Mind and Cognition*. Reprinted in Block et al. (1997).
- Lewis, D. K. 1993. "Causal Explanation." In D.-H. Ruben (ed.), *Explanation*. Oxford: Oxford UP.
- Libet, B. 1985. "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action." *Behavioral and Brain Sciences* 8: 529-566.
- Loar, B. 1990. "Phenomenal States." *Philosophical Perspectives* 4: 81-108. Reprinted in Block et al. (1997).
- Lurz, R. 1999. "Animal Consciousness." *Journal of Philosophical Research* 24: 149-168.
- Lurz, R. 2003. "Neither HOT nor COLD: An Alternative Account of Consciousness." *Psyche* 9.
- Lycan, W. G. 1987. *Consciousness*. Cambridge MA: MIT Press.
- Lycan, W. 1996. *Consciousness and Experience*. Cambridge MA: MIT Press.
- Lycan, W. G. 2001. "A Simple Argument for a Higher-Order Representation Theory of Consciousness." *Analysis* 61: 3-4.
- McGinn, C. 1989. "Can We Solve the Mind-Body Problem?" *Mind* 98: 349-366.
- McGinn, C. 1995. "Consciousness and Space." *Journal of Consciousness Studies* 2: 220-30.
- McGinn, C. 1999. *The Mysterious Flame*. Cambridge MA: MIT Press.
- McGinn, C. 2004. *Consciousness and Its Objects*. Oxford: Oxford UP.
- Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. New York: Penguin.
- Maloney, J. C. 1989. *The Mundane Matter of the Mental Language*. Cambridge: Cambridge UP.
- Mellor, D. H. 1978. "Conscious Belief." *Proceedings of the Aristotelian Society* 78: 87-101.
- Moore, G. E. 1903. "The Refutation of Idealism." In his *Philosophical Papers*. London: Routledge and Kegan Paul.
- Nagel, T. 1974. "What Is It Like to Be a Bat?" *Philosophical Review* 83: 435-450. Reprinted in Block et al. 1997.
- Natsoulas, T. 1993. "What Is Wrong with Appendage Theory of Consciousness?" *Philosophical Psychology* 6: 137-154.
- Natsoulas, T. 1996. "The Case for Intrinsic Theory: I. An Introduction." *Journal of Mind and Behavior* 17: 267-286.
- Neander, K. 1998. "The Division of Phenomenal Labor: A Problem for Representational Theories of Consciousness." *Philosophical Perspectives* 12: 411-434.
- Nemirow, L. 1990. "Physicalism and the Cognitive Role of Acquaintance." In W. G. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell.
- Peacocke, C. 1983. *Sense and Content*. Oxford: Clarendon.
- Putnam, H. 1967. "The Nature of Mental States." Originally published as "Psychological Predicates," in W. H. Capitan and D. D. Merrill (eds.), *Art, Mind, and Religion*. Reprinted in D. M. Rosenthal (ed.), *The Nature of Mind*. Oxford: Oxford UP.
- Putnam, H. 1975. "The Meaning of 'Meaning'." In his *Mind, Language, and Reality,* Cambridge: Cambridge UP.
- Rey, G. 1988. "A Question about Consciousness." In H. Otto and J. Tueidio (eds.), *Perspectives on Mind*. Norwell: Kluwer Academic Publishers. Reprinted in Block et al. (1997).
- Rey, G. 1998. "A Narrow Representationalist Account of Qualitative Experience." *Philosophical Perspectives* 12: 435-457.
- Rosenthal, D. M. 1986. "Two Concepts of Consciousness." *Philosophical Studies* 94: 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. 1993. "Thinking that One Thinks." In M. Davies and G. W. Humphreys (eds.), *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell.
- Rosenthal, D. M. 2000. "Consciousness and Metacognition." In D. Sperber (ed.), *Metarepresentation*. Oxford: Oxford UP.
- Rosenthal, D. M. 2002a. "Explaining Consciousness." In D. J. Chalmers (ed.), *Philosophy of Mind*. Oxford and New York: Oxford UP.
- Rosenthal, 2002b. "Consciousness and Higher-Order Thoughts." In L. Nadel (ed.), *Macmillan Encyclopedia of Cognitive Science*. New York: Macmillan Publishers.
- Seager, W. 1999. *Theories of Consciousness*. London: Routledge.
- Searle, J.R. 1992. *The Rediscovery of Mind*. Cambridge MA: MIT Press.
- Segal, G. 2000. *A Slim Book on Narrow Content*. Cambridge MA: MIT Press.
- Shoemaker, S. 1994a. "Phenomenal Character." *Nous* 28: 21-38.
- Shoemaker, S. 1994b. "Self-knowledge and 'Inner Sense.' Lecture III: The Phenomenal Character of Experience." *Philosophy and Phenomenological Research* 54: 291-314.
- Shoemaker, S. 1996. "Colors, Subjective Reactions, and Qualia." *Philosophical Issues* 7: 55-66.
- Shoemaker, S. 2002. "Introspection and Phenomenal Character." In D. J. Chalmers (ed.), *Philosophy of Mind*. Oxford and New York: Oxford UP.

- Siewert, C. P. 1998. *The Significance of Consciousness*. Princeton NJ: Princeton UP.
- Searle, J. R. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge UP.
- Smart, J. J. C. 1959. "Sensations and Brain Processes." *Philosophical Review* 68: 141-156.
- Smith, D. W. 1986. "The Structure of (Self-)Consciousness." *Topoi* 5: 149-156.
- Smith, D. W. 1989. *The Circle of Acquaintance*. Dordrecht: Kluwer Academic Publishers.
- Smith, D. W. 2004. "Return to Consciousness." In his *Mind World*. Cambridge: Cambridge UP.
- Thau, M. 2002. *Consciousness and Cognition*. Oxford and New York: Oxford UP.
- Thomasson, A. L. 2000. "After Brentano: A One-Level Theory of Consciousness." *European Journal of Philosophy* 8: 190-209.
- Thompson, E. and D. Zahavi. "Philosophical Issues: Continental Perspectives: Phenomenology." This volume.
- Tye, M. 1986. "The Subjective Qualities of Experience." *Mind* 95: 1-17.
- Tye, M. 1992. "Visual Qualia and Visual Content." In T. Crane (ed.), *The Contents of Experience*. Cambridge: Cambridge UP.
- Tye, M. 1995. *Ten Problems of Consciousness*. Cambridge MA: MIT Press.
- Tye, M. 2000. *Consciousness, Color, and Content*. Cambridge MA: MIT Press.
- Tye, M. 2002. "Visual Qualia and Visual Content Revisited." In D. J. Chalmers (ed.), *Philosophy of Mind*. Oxford and New York: Oxford UP.
- Van Gulick, R. 1993. "Understanding the Phenomenal Mind: Are We All Just Armadillos?" Reprinted in Block et al. (1997).
- Van Gulick, R. 2001. "Inward and Upward – Reflection, Introspection, and Self-Awareness." *Philosophical Topics* 28: 275-305.
- Velmans, M. 1992. "Is Human Information Processing Conscious?" *Behavioral and Brain Sciences* 14: 651-669.
- Wegner, D. M. 2002. *The Illusion of Conscious Will*. Cambridge MA: MIT Press.
- Williford, K. W. 2006. "The Self-Representational Structure of Consciousness." In Kriegel and Williford 2006.
- Zahavi, D. 1998. "Brentano and Husserl on Self-Awareness." *tudes Phénomènologiques* 27-8: 127-169.
- Zahavi, D. 1999. *Self-awareness and Alterity*. Evanston IL: Northwestern UP.
- Zahavi, D. 2004. "Back to Brentano?" *Journal of Consciousness Studies* 11.
- Zahavi, D. and J. Parnas 1998. "Phenomenal Consciousness and Self-Awareness: A Phenomenological Critique of Representational Theory." *Journal of Consciousness Studies* 5: 687-705.

---

1 More accurately, I will present *central aspects* of the main account, the case in favor, and the case against. Obviously, space and other limitations do not allow me to present the full story on each of these.

2 The distinction between creature consciousness and state consciousness is due to Rosenthal (1986).

3 Availability consciousness as construed here is very similar to the notion of *access consciousness* as defined by Block (1995). There are certain differences, however. Block defines access consciousness as the property a mental state has when it is poised for free use by the subject in her reasoning and action control. It may well be that a mental state is availability-conscious if and only if it is access-conscious. For a detailed discussion of the relation between phenomenal consciousness and access consciousness, see Kriegel 2006b.

4 It is debatable whether thoughts, beliefs, desires, and other cognitive states can at all be conscious in this sense. I will remain silent on this issue here. For arguments that they can, see Goldman 1993a, Horgan and Tienson 2002, and Siewert 1998.

5 The terms "easy problems" and "hard problem" are intended as mere label, not as descriptive. Thus it is not suggested here that understanding any of the functions of consciousness is at all easy in any significant sense. Any scientist who devoted time to the study of consciousness knows how outstanding the problems in this field are. These terms are just a terminological device designed to bring out the fact that the problem of why there is something it feels like to undergo a conscious experience appears to be of a different order than the problems of mapping out the cognitive functions of consciousness.

6 In the course of the discussion I avail myself of philosophical terminology that may not be familiar to the non-philosophically trained reader. However, I have tried to recognize all the relevant instances and such and include an endnote that provides a standard explication of the terminology in question.

7 This is so even if phenomenal consciousness does not turn out to have much of a functional significance in the ordinary cognitive life of a normal subject – as some (Libet 1985, Velmans 1992, Wegner 2002) have indeed argued.

8 No major philosopher holds this view, to my knowledge**.**

9 Many of the key texts discussed in this chapter are conveniently collected in Block et al. (1997). Here, and in the rest of the chapter, I refer to the reprint in that volume.

10 This is what Churchland often discusses under the heading of the "plasticity of mind" (see especially Churchland 1979).

11 It may not be perceiving those brain states *as* brain states. But it will nonetheless be a matter of perceiving the brain states.

12 The view – sometimes referred to as *emergentism* – that consciousness is caused by the brain, or causally emerges from brain activity, is often taken by scientists to be materialist enough. But philosophers, being interested in the *ontology* rather than *genealogy* of consciousness, commonly take it to be a form of dualism. If consciousness cannot be shown to be itself material, but only caused by matter, then consciousness is itself immaterial, as the dualist claims. At the same time, the position implicit in scientists' work is often that what is caused by physical causes in accordance with already known physical laws should be immediately considered physical. This position, which I have called elsewhere *inclusive materialism* (Kriegel 2005b), is not unreasonable. But the present chapters is dedicated to *philosophers'* theories of consciousness, so I will set it aside.

13 It should be noted that McGinn himself has repeatedly claimed that his position is not dualist. Nonetheless others have accused him of being committed to dualism (e.g., Bueckner and Berukhim 2003). There is no doubt that McGinn does not *intend* to commit to dualism. In a way, his position is precisely that due to our cognitive closure we cannot even know whether materialism is true or dualism. Yet it is a fair criticism to suggest that McGinn is committed to dualism despite himself because his argument for mysterianism would not go through unless dualism was true.

14 More generally, it is curious to hold, as McGinn does, that an organism's concept-forming procedures are powerful enough to *frame* a problem, without being powerful enough to *frame* the solution. To be sure, the wrong solution may be framed, but this would suggest not that the conceptual capabilities of the organism are at fault, but rather that the organism made the wrong turn somewhere in its reasoning. The natural thought is that if a conceptual scheme is powerful enough to frame a problem it should be powerful enough to frame the solution. Whether the correct solution will actually *be* framed is of course anyone's guess. But the problem cannot be a constitutive limitation on concept formation mechanisms. (For a more detailed development of this line of critique, see Kriegel 2004a.) There is a counter-example of this sort of claim, however. Certain problems that can be framed within the theory of rational numbers cannot be solved within it; the conceptual machinery of irrational numbers must be brought in to solve these problems. It might be claimed, however, that this sort of exception is limited to formal systems, and does not apply to theories of the natural world. Whether this claim is plausible is something we will not adjudicate here.

15 Monism divides into two sub-groups: materialist monism, according to which the only kind of stuff there is is matter, and idealist monism, according to which the stuff in question is some sort of mindstuff.

16 Idealism is not really considered a live option in current philosophical discussions, although it *is* defended by Foster (1982). I will not discuss it here.

17 Such coming-apart happens, for Descartes, upon death of the physical body. We should note that Cartesian Substance dualism drew much of its motivation from religious considerations, partly because it provided for the survival of the soul. The main difficulty historically associated with it is whether it can account for the causal interaction between the mind and the body.

18 So property dualism is compatible with substance monism. Unlike Descartes and other old-school dualists, modern dualists for the most part hold that there is only one kind of stuff, or substance, in the world – matter. But matter has two different kinds of *properties* – material and immaterial.

19 A kind of property F supervenes on a kind of property G with logical necessity – or for short logically supervenes on them – just in case two object's differing with respect to their F properties without differing with respect to their G properties would be in contravention of laws of logic. A kind of property F supervenes on a kind of property G with metaphysical necessity – or for short metaphysically supervenes on them – just in case it is impossible for two object's differing with respect to their F properties without differing with respect to their G properties. Philosophers debate whether there is a difference between the two (logical and metaphysical supervenience). That debate will not concern us here.

20 This stronger claim will require a stronger argument. The claim that phenomenal properties are not identical to physical properties could be established through the now familiar argument from multiple realizability (Putnam 1967). But multiple realizability does not entail failure of supervenience. To obtain the latter, Chalmers will have to appeal to a different argument, as we will see in the next sub-section.

21 As a consequence, phenomenal properties do supervene on physical properties with *nomological* necessity, even though they do not supervene with metaphysical or logical necessity. A kind of property F supervenes on a kind of property G with nomological (or natural) necessity – or for short nomologically supervenes on them – just in case two object's differing with respect to their F properties without differing with respect to their G properties would be in contravention of laws of nature.

22 So causal explanation is the sort of explanation one obtains by citing the cause of the explanandum. For discussions of the nature of causal explanation, see (e.g.) Lewis 1993.

23 The latter will govern only the causal interaction *among* physical events. They will not cover causal interaction between physical and phenomenal, non-physical events. These will have to be covered by a special and new set of laws.

24 In Baars' (1988, 1997) Global Workspace Theory, consciousness is *reductively* explained in terms of global availability. In a functionalist theory such as Dennett's (1981, 1991), consciousness is *reductively* explained in terms of functional organization. Chalmers' position is that neither can explain consciousness *reductively*, though both may figure as part of the *causal* explanation of it. These theories will not be discussed in the present chapter, since they are fundamentally psychological (rather than philosophical) theories of consciousness.

25 A linguistic context is intensional if it disallows certain inferences, in particular existential generalization (the inference from "a is F" to "there is an x, such that x is F") and substitution of co-referential terms *salva veritate* (the inference from "a is F" and "a = b" to "b is F"). Epistemic contexts – contexts involving the ascription of knowledge – are intensional in this sense.

26 Another popular materialist response to these arguments is that what is being gained is not new knowledge, but rather new abilities (Lewis 1988, Nemirow 1990). Upon being released from her room, the Knowledge Argument's protagonist does not acquire new knowledge, but rather a new set of abilities. And likewise what we lack with respect to what it is like to be a bat is not any particular knowledge, but a certain ability – the ability to imagine what it is like to be a bat. But from the acquisition of a new ability one can surely not infer the existence of a new fact.

27 Materialists reason that since what it is like to see red is identical to a neurophysiological fact about the brain, and *ex hypothesi* the Knowledge Argument's protagonist knows the latter fact, she already knows the former. So she knows the fact of what it is like to see red, but not *as* a fact about what it is like to see red. Instead, she knows the fact of what it is like to see red *as* a fact about the neurophysiology of the brain. What happens when she comes out of her room is that she comes to know the fact of what it is like to see red *as* a fact about what it is like to see red. That is, she learns in a new way a fact she already knew in another way. The same applies to knowledge of what it is like to be a bat: we may know all the facts about what it is like to see a bat, and still gain new knowledge about bats, but this new knowledge will present to us a fact we already know in a way we do not know it yet.

28 It could be responded by the dualist that some pieces of knowledge are so different that the fact known thereby could not possibly turn out to be the same. Knowledge that the evening star is glowing and knowledge that the morning star is glowing are not such. But consider knowledge that justice is good and knowledge that banana is good. The dualist could argue that these are such different pieces of knowledge that it is impossible that the facts thereby known should turn out to be one and the same. The concepts of evening star and morning star are not different enough to exclude the possibility that they pick out the same thing, but the concepts of

justice and banana are such that it cannot possibly be case that justice should turn out to be the same thing as bananas.

29 The kind of possibility we are concerned with here, and in the following presentation of variations on this argument, is not practical possibility, or even a matter of consistency with the laws of nature. Rather it possibility in the widest possible sense – that of consistency with the laws of logic and the very essence of things. This is what philosophers refer to as metaphysical possibility.

30 The modal force of this supervenience claim is concordant with that of the claim in Premise 2, that is, that of metaphysical necessity.

31 The reason it is impossible is that there is no such thing as *contingent identity*, according to the official doctrine hailing from Kripke. Since all identity is necessary, and necessity is cashed out as truth in all possible world, it follows that when a = b in the actual world a = b in all possible worlds, that is, a is *necessarily* identical to b.

32 The interpretation I will provide is based on certain key passages in Chalmers 1996: 131-134, but will cast the argument in terms that are mine, not Chalmers'.

33 I mean the property of apparent water to be more or less the same as the property philosophers often refer to as "watery stuff," i.e., the property of being superficially (or to the naked eye) the same as water (i.e., clear, drinkable, liquid, etc.).

34 Chalmers (1996: 132; my italics) writes: "…the primary intension [of "consciousness"] determines a *perfectly good property* of objects in possible worlds. The property of being watery stuff [or apparent water] is a perfectly reasonable property, even though it is not the same as the property of being H_2O. If we can show that there are possible worlds that are physically identical to ours but in which the properly introduced by the primary intension is lacking, then dualism will follow."

35 Our discussion so far has presupposed a "latitudinous" approach to properties, according to which there is a property that corresponds to every predicate we can come up with. (Thus, if we can come up with the predicate "is a six-headed space lizard or a flying cow," then there is the property of being a six-headed space lizard or a flying cow. This does not mean, however, that the property is actually instantiated by any actual object.) But on a *sparse* conception of property – one which rejects the latitudinous assumption – there may not be appearance properties at all.

36 The notion of a natural property is hard to pin down and is the subject of philosophical debate. The most straightforward way of understanding natural properties is as properties that figure in the ultimate laws of nature (Armstrong 1978, Fodor 1974).

37 That is, they would have to their causal efficacy restricted to bringing about physical events and property-instantiations that already have independent sufficient causes (and that would therefore take place anyway, regardless of the non-supervenient properties. (This is the second option of the dilemma.)

38 This is the strategy in Chalmers 1996. Later on, Chalmers (2002a) embraces a three-pronged approach, the third prong consisting in accepting causal overdetermination.

39 When a cause C causes an effect E, C's causing of E may have its own (mostly accidental) effects (e.g., it may surprise an observer who did not expect the causing to take place), but E is not one of them. This is because E is caused by C, not by C's causing of E. Dretske (1988) distinguished between *triggering* causes and *structuring* causes, the latter being causes of certain causal relations (such as C's causing of E), and offers an account of structuring causes. But this is an account of the *causes* of causal relations, not of their *effects*. To my knowledge, there is no account of the effects of causal relations, mainly because these seem to be chiefly accidental.

40 Or at least they would be nearly epiphenomenal, having no causal powers except perhaps to bring about some accidental effects of the sort pointed out in the previous endnote.

41 By "representational properties" it is meant properties that the experience has in virtue of what it represents – not, it is important to stress, properties the experience has in virtue of what does the representing. In terms of the distinction between vehicle and content, representational properties are to be understood as content properties rather than vehicular properties. We can also make a distinction between two kinds of vehicular properties: those

that are essential to the vehicling of the content, and those that are not. (Block's (1996) distinction between mental paint and mental latex (later, "mental oil") is supposed to capture this distinction.) There is a sense in which a view according to which phenomenal properties are reductively accountable for in terms of vehicular properties essential to the vehicling is representational, but the way the term "representationalism" is used in current discussions of consciousness, it does not qualify as representationalism. A view of this sort is defended, for instance, by Maloney (1989), but otherwise lacks a vast following. I will not discuss it here.

42 By the "phenomenal character" of a mental state at a time *t* I will mean the set of all phenomenal properties the state in question instantiates at *t*. By "representational content" I mean whatever the experience represents. (Experiences represent things, in that they have certain accuracy or veridicality conditions: conditions under which an experience would be said to get things right.)

43 See Dretske 1981, 1988 for the most thoroughly worked out reductive account of mental representation in informational and teleological terms. According to Dretske (1981), every event in the world generates a certain amount of information (in virtue of excluding the possibility that incompatible event take place). Some events also take place only when other events take place as well, and this is sometimes dictated by the laws of nature. Thus it may be a law of nature that dictates that an event type E_1 is betokened only when event type E_2 is betokened. When this is the case, E_1 is said to be *nomically dependent* upon E_2, and the tokening of E_1 carries the information that E_2 has been betokened. Or more accurately, the tokening of E_1 carries the information generated by the tokening of E_2. Some brain states bear this sort of relation to world states: the former come into being, as a matter of law, only when the latter do (i.e., the former are nomically dependent upon the latter). Thus, a certain type of brain state may be tokened only when it rains. This brain state type would thus carry the information that it rains. An informational account of mental representation is based on this idea: that a brain state can represent the fact that it rains by carrying information about it, which it does in virtue of nomically depending on it.

44 Other representational theories can be found in Byrne (2001), Dretske (1995), Lurz (2003), Shoemaker (1994a, 1994b, 1996, 2002) and Thau (2002). Some of these versions are importantly different from Tye's, not only in detail but also in spirit. This is particularly so with regard to Shoemaker's view (as well as Lurz's). For a limited defense and elaboration of Shoemaker's view, see Kriegel (2002a, 2002b). In what way this defense is limited will become evident at the end of this section.

45 The properties of intentionality and abstractness are fairly straightforward. The former is a matter of intensionality, that is, the disallowing of existential generalizations and truth-preserving substitutions of co-referential terms. The second is a matter of the features represented by experience not being concrete entities (this is intended to make sense of misrepresentation of the same features, in which case no concrete entity is being represented).

46 This line of thought can be resisted on a number of scores. First, it could be argued that I do have a *short-lived* concept of $blue_{17}$, which I possess more or less for the duration of my experience. Second, it could be claim that although I do not possess the *descriptive* concept "$blue_{17}$," I do possess the *indexical* concept "*this* shade of blue," and that it is the latter concept that is deployed in my experience's representational content. Be that as it may, the fact that conscious experiences can represent properties which the subject cannot recognize across relatively short stretches of time is significant enough. Even if we do not wish to treat them as non-conceptual, we must treat them at least as "sub-recognitional." Tye's modified claim would be that the representational content of experience is poised, abstract, sub-recognitional, intentional content.

47 To be sure, it does not represent the tissue damage *as* tissue damage, but it does represent the tissue damage. Since the representation is non-conceptual, it certainly cannot employ the concept of "tissue damage."

48 An error theory is a theory that ascribes a widespread error in commonsense beliefs. The term was coined by J. L. Mackie (1977). Mackie argued that values and value judgement are subjective. Oversimplifying the dialectic, a problem for this view is that such judgement as "murder is wrong" appear to be, and are commonly taken to be, objectively true. In response Mackie embraced what he termed an error theory: that the common view of moral and value judgements is simply one huge mistake.

49 Externalism about representational content, or "content externalism" for short, is the thesis that the representational content of experiences, thoughts, and even spoken statements is partially determined by objects outside the subject's head. Thus, if a person's interactions with watery stuff happen to be interactions with H_2O, and another person's interactions with watery stuff happen to be interactions with a superficially similar stuff that is not composed of H_2O, then even if the two persons cannot tell apart H_2O and the other stuff, and are

unaware of the differences in the molecular composition of the watery stuff in their environment, the representational contents of their respective water thoughts (as well as water pronouncements and water experiences) are different (Putnam 1975). Or so externalists claim.

50 Another option is to go internalist with respect to the representational content that determines the phenomenal properties of conscious experiences. With the recent advent of credible account of narrow content (Chalmers 2002b, Segal 2000), it is now a real option to claim that the phenomenal properties of experience are determined by experience's narrow content (Kriegel 2002a, Rey 1998). However, it may turn out that this version of representationalism will not be as well supported by the transparency of experience.

51 For one such line of criticism, on which I will not elaborate here, see Kriegel 2002c.

52 Elsewhere, I construe this form of pre-reflective self-consciousness as what I call *intransitive self-consciousness*. Intransitive self-consciousness is to be contrasted with transitive self-consciousness. The latter is ascribed in reports of the form "I am self-conscious of my thinking that *p*," whereas the former is ascribed in reports of the form "I am self-consciously thinking that *p*." For details see Kriegel 2003b, 2004b.

53 Part of this neglect is justified by the thesis that the for-me-ness of conscious experiences is an illusory phenomenon. For an argument for the psychological reality of it, see Kriegel 2004b.

54 There are versions of representationalism that may be better equipped to deal with the subjective character of experience. Thus, according to Shoemaker's (2002) version, a mental state is conscious when it represents a subject-relative feature, such as the disposition to bring about certain internal states in the subject. It is possible that some kind of for-me-ness can be accounted for in this manner. It should be noted, however, that this is not one of the considerations that motivates Shoemaker to develop his theory the way he does.

55 Rosenthal prefers to put this idea as follows: conscious states are states we are conscious of. He then draws a distinction between consciousness and consciousness of – intransitive and transitive consciousness (Rosenthal 1986, 1990). To avoid unnecessary confusion, I will state the same idea in terms of awareness-of rather than consciousness-of. But the idea is the same. It is what Rosenthal calls sometimes the "transitivity principle" (e.g., Rosenthal 2000): a mental state is intransitively conscious only if we are transitively conscious of it.

56 The representation is "higher-order" in the sense that it is a representation of a representation. In this sense, a first-order representation is a representation of something that is not itself a representation. Any other representation is higher-order.

57 More than that, according to Rosenthal (1990), for instance, the particular *way* it is like for S to have M is determined by the particular *way* M* represents M. Suppose S tastes an identical wine in 1980 and in 1990. During the eighties, however, S had become a wine connoisseur. Consequently, wines she could not distinguish at all in 1980 strike her in 1990 as worlds apart. That is, during the eighties she acquired a myriad of concepts for very specific and subtle wine tastes. It is plausible to claim that what it is like for S to taste the wine in 1990 is different from what it was like for her to taste it in 1980 – even though the wines' own flavors are identical. Arguably, the reason for the difference in what it is like to taste the wine is that the two wine-tasting experiences are accompanied by radically different higher-order representations *of* them. This suggests, then, that the higher-order representation not only determines *that* there is something it is like for S to have M, but also *what* it is like for S to have M.

58 I do not mean the term "yield" in a causal sense here. The higher-order monitoring theory does not claim that M*'s representing of M somehow *produces*, or *gives rise to*, M's being conscious. Rather, the claim is conceptual: M's being conscious *consists in*, or *is constituted by*, M*'s representing of M.

59 Other versions of the higher-order thought view can be found in Carruthers (1989, 1996), Dennett (1969, 1991), and Mellor (1978).

60 Rosenthal (1990: 739-40) claims that it is essential to a perceptual state that it has a sensory quality, but the second-order representations do not have sensory qualities and are therefore non-perceptual. Van Gulick (2001) details a longer and more thorough list of features that are characteristic of perceptual states and considers which of them is likely to be shared by the higher-order representations. His conclusion is that some are and some are not.

61 The notion of direction of fit has its origins in the work of Anscombe (1957), but has been developed in some detailed and put to extensive work mainly by Searle (1983). The idea is that mental states divide into two main groups, the cognitive ones (paradigmatically, belief) and the conative ones (paradigmatically, desire). The former are such that they are supposed to make the mind fit the way the world is (thus "getting the facts right"), whereas the latter are such that they are supposed to make the world fit the way the mind is (a change in the world is what would satisfy them).

62 Kobes (1995) suggests a version of higher-order monitoring theory in which the higher-order representation has essentially a telic direction of fit. But Rosenthal construes it as having only a thetic one.

63 Carruthers (1989, 1996, 2000), and probably also Dennett (1969, 1991), attempt to account for consciousness in terms of merely *tacit* or *dispositional* higher-order representations. But these would not do, according to Rosenthal. The reason for this is that a merely dispositional representation would not make the subject aware of her conscious state, but only *disposed* to being aware of it, whereas the central motivation behind the higher-order monitoring view is the fact that conscious states are state we are aware of having (Rosenthal 1990: 742).

64 Earlier on, Rosenthal (1990) required that the higher-order thought be not only non-inferential, but also non-observational. This latter requirement was later dropped (Rosenthal 1993).

65 A person may come to believe that she is ashamed about something on the strength of her therapist's evidence. And yet the shame state is not conscious. In terms of the terminology introduced in the introduction, the state may become availability-conscious, but not phenomenally conscious. This is why the immediacy of awareness is so crucial. Although the person's second-order belief constitutes an awareness of the shame state, it is not a non-inferential awareness, and therefore not immediate awareness.

66 *De se* content is content that is *of oneself*, or more precisely, of oneself *as oneself*. Castaˆeda (1966), who introduced this term, also claimed that *de se* content is irreducible to any other kind of content. This latter claim is debatable and is not part of the official higher-order thought theory.

67 Rosenthal's (1990: 742) argument for this requirement is the following. My awareness of my bluish experience is an awareness of *that particular* experience, not of the general type of experience it is. But it is impossible to represent a mental state as particular without representing in which subject it occurs. Therefore, the only way the higher-order thought could represent my experience in its particularity is if it represented it as occurring in me.

68 This is necessary to avert infinite regress. If the higher-order state was itself conscious, it would have to be itself represented by a yet higher-order state (according to the theory) and so the hierarchy of states would go to infinity. This is problematic on two scores. Firstly, it is empirically implausible, and perhaps impossible, that a subject should entertain an infinity of mental states whenever conscious. Secondly, if a mental state's being conscious is explained in terms of another conscious states, the explanation is "empty," inasmuch as it does not explain consciousness in terms of something other than consciousness.

69 This claim can be made on phenomenological grounds, instead of on the basis of conceptual analysis. For details, see Kriegel 2004b.

70 To repeat, the conceptual grounds are the fact that it seems to be a conceptual truth that conscious states are states we are aware of having. This seems to be somehow inherent in the very concept of consciousness.

71 There are other arguments that have been leveled against the higher-order monitoring theory, or specific versions thereof, into which I will not have the space to go. For arguments not discussed here, see Block 1995, Caston 2002, Dretske 1995, Guzeldere 1995, Kriegel 2006a, Levine 2001, Natsoulas 1993, Rey 1988, Seager 1999, Zahavi and Parnas 1998.

72 The argument has also been made by Caston (2002), Levine (2001), and Seager (1999). For a version of the argument directed at higher-order perception theory (and appealing to higher-order misperceptions), see Neander 1998.

73 Note that M* does not merely misrepresent M to be F when in reality M is not F, but misrepresents M to be F when in reality there is no M at all.

74 This would be a particular version of the supposition we made at the very beginning of this chapter, by way of analyzing creature consciousness in terms of state consciousness.

75 Furthermore, if M* were normally conscious, the same problem would arise with the third-order representation of *it* (and if the third-order representation were normally conscious, the problem would arise with the fourth-order state. To avert infinite regress, the higher-order monitoring theorist must somewhere posit an unconscious state, and when she does, she will be unable to claim that that state instantiates the property of being conscious when it misrepresents.

76 This appears to be Rosenthal's latest stance on the issue (in conversation).

77 There are surely other ways the higher-order monitoring theorist may try to handle the problem of targetless higher-order representations. But many of them are implausible and all of them complicate the theory considerably. One of the initial attractions of the theory is its clarity and relative simplicity. Once it is modified along any of the lines sketched above, it becomes significantly less clear and simple. To that extent, it is considerably less attractive than it initially appears.

78 See Brook and Raymont (2006), Caston (2002), Hossack (2002), Kriegel (2003b), and Williford (2006). For the close variation, see Carruthers (2000, 2006), Gennaro (1996, 2002, 2006), Kobes (1995), Kriegel (2003a, 2005, 2006a), and Van Gulick (2001, 2004).

79 For fuller discussion of Brentano's account, see Caston (2002), Kriegel (2003a), Smith (1986, 1989) Thomasson (2000), and Zahavi (1998, 2004).

80 So the self-representational content of conscious states is *de se* content. There are places where Brentano seems to hold something like this as well. See also Kriegel 2003a.

81 For more on the distinction between content and attitude (or mode), see Searle 1983. For a critique of Smith's view, see Kriegel 2005a.

82 A similar account would be that conscious states are not conscious in virtue of standing in a certain relation to themselves, but this is because their secondary intentionality should be given an adverbial analysis. This is not to say that all intentionality must be treated adverbially. It may well be that the primary intentionality of conscious states is a matter of their standing in a certain informational or teleological relation to their primary objects. Thus, it need not be the case that S's conscious fear that *p* involves S's fearing *p*-ly rather than S's standing in a fear relation to the fact that *p*. But it *is* the case that S's *awareness* of her fear that *p* involves being aware fear-that-*p*-ly rather than standing in an awareness relation to the fear that *p*. To my knowledge, nobody holds this view.

83 A constitutive, non-contingent relation is a relation that two things do not just happen to entertain, but rather they would not be the things they are if they did not entertain those relations. Thus A's relation to B is constitutive if bearing it to B is part of what constitutes A's being what it is. Such a relation is necessary rather than contingent, since there is no possible world in which A does not bear it to B – for in such a world it would no longer be A.

84 Elsewhere, I have defended a view similar in key respects to Van Gulick's – see Kriegel 2003a, 2005, 2006a.

85 Indeed, the problem may be even more pressing for a view such as the higher-order global states theory. For the latter requires not only the ability to generate higher-order contents, but also the ability to *integrate* those with the right lower-order contents.

86 For a more elaborate argument that self-representation may not be a sufficient condition for consciousness, one that could provide a reductive explanation of it, see Levine 2001 Ch. 6.

87 I am appealing here to a distinction defended, e.g., by Cummins (1979), Dretske (1988), and Searle (1992). Grice noted that some things which exhibit aboutness of meaningfulness, such as words, traffic signs, and arrows, do so only on the assumption that someone *interprets* them to have the sort of meaning they have. But these acts of interpretation are themselves contentful, or meaningful. So their own meaning must be either derived by further interpretative acts or be intrinsic to them and non-derivative. Grice's claim was that thoughts and other mental states have an aboutness all their own, independently of any interpretation.

88 This is denied by Dennett (1987), who claims that all intentionality is derivative.

89 One might claim that such states are less clearly conceivable when their self-representational content is fully specified. Thus, if the content is of the form "I myself am herewith having this very bluish experience," it is less clearly the case that one can conceive of the an unconscious state having this content.

90 The conceivability of unconscious self-representing states may not be *proof* of their possibility, but it is *evidence* of their possibility. It is therefore evidence against the self-representational theory.

91 The reductivist may claim that zombies with the same physical properties we have are conceivable only because we are not yet in a position to focus our mind on the right physical structure. As progress is made toward identification of the right physical structure, it will become harder and harder to conceive of a zombie exhibiting this structure but lacking all consciousness.

92 For comments on an earlier draft of this chapter, I would like to thank George Graham, David Jehle, Christopher Maloney, Amie Thomasson, and especially David Chalmers.