

THE TWO-DIMENSIONAL CONTENT OF CONSCIOUSNESS

Simon Prosser

sjp7@st-andrews.ac.uk

ABSTRACT: In this paper I put forward a representationalist theory of conscious experience based on Robert Stalnaker's version of two-dimensional modal semantics. According to this theory the phenomenal character of an experience correlates with a content equivalent to what Stalnaker calls the *diagonal proposition*. I show that the theory is closely related both to functionalist theories of consciousness and to higher-order representational theories. It is also more compatible with an anti-Cartesian view of the mind than standard representationalist theories.

1. Introduction

Over the last decade or so there has been a rapidly developing interest in the relation between phenomenal consciousness and intentionality, with a number of philosophers arguing for close connections between the two. Most conspicuous among the theories that have been put forward is *representationalism*. Let the *phenomenal character* of a conscious experience be defined such that two experiences differ in phenomenal character if and only if there is a difference in *what it is like* to have those experiences. Representationalism is the view that the phenomenal character of an experience depends entirely on its representational content; experience contains no non-representational features. This contrasts with the idea that there are *qualia* that may or may not act as intermediaries for represented contents but which also contribute intrinsic phenomenological features that do not represent anything. Representationalism has been notably defended by Michael Tye (1995, 2000,

2002, 2003) and Fred Dretske (1993), but similar views or variants have also been put forward by Sydney Shoemaker (2000), David Chalmers (2004b, 2006), and many others.¹ Another family of views combining phenomenal consciousness with intentionality are *higher-order representational (HOR) theories*. According to these theories phenomenal consciousness can be analysed in terms of higher-order representational states that represent first-order conscious states. HOR theories have been defended in various forms by David Rosenthal (1986, 1997, 2002), Peter Carruthers (1996, 2000), Daniel Dennett (1978, 1991), David Armstrong (1968, 1984) and William Lycan (1996), among others.

In this paper I shall put forward a version of representationalism that differs from standard versions. In brief, the view is that phenomenal character correlates with a content equivalent to the *diagonal proposition* in Robert Stalnaker's (1978, 2001, 2004) version of two-dimensional modal semantics for utterances. Consequently I call the view *diagonal representationalism*. Diagonal representationalism embodies connections between representationalism, HOR theories and functionalism that, as far as I am aware, have not previously been noticed. In effect it combines elements of all three doctrines within a single theory.

The structure of the paper is as follows. I start by describing some motivations for representationalism, some varieties of representationalism and some standard objections. I accept that some of these objections pose genuine problems for most existing versions of representationalism. I then show, moreover, that some minimal functionalist assumptions accepted by many philosophers (including even some property dualists) are incompatible with standard versions of representationalism.

I then outline some key features of Stalnaker's two-dimensional semantics before describing diagonal representationalism in detail. I then show how diagonal representationalism combines representationalism, higher-order representation and a broadly functionalist theory of consciousness. In the final sections I show that standard versions of representationalism are committed to

a questionable view of the mind often known as the *Cartesian theatre*. Diagonal representationalism, by contrast, is compatible with a rejection of the Cartesian Theatre because there is just the right kind of indeterminacy about the diagonal representational content to account for there being no fact of the matter about certain aspects of phenomenal character. Different yet equally good interpretation schemes can be given for both.

2. Motivations and standard problems for representationalism

G. E. Moore (1903) argued that experience was *diaphanous*. When one sees a ripe tomato one can focus one's attention on the tomato, or on its redness, but one cannot focus one's attention on the experience itself without just focusing once again on the tomato or its redness. Representationalists take this to suggest that all that is encountered in an experience is its content: the ripe red tomato. There are debates over whether this is a good argument but in any case I shall treat representationalism as a hypothesis, to be rejected only if it is shown to be false. Hence after briefly distinguishing some different versions of representationalism I shall examine the main objections.

For materialists, representationalism offers the possibility of *reducing* phenomenal character to representational content and thus reducing the problem of accounting for phenomenal consciousness to the more naturalistically tractable problem of accounting for representation. Advocates of reductive representationalism include Tye and Dretske. Tye's theory concerns only *abstract* content; content that excludes particular objects or surfaces. This restriction is needed because hallucinatory experiences and experiences of different qualitatively identical objects can all be subjectively indistinguishable while differing in the particulars represented.

Those who reject the reducibility of the phenomenal to the physical on independent grounds can, however, accept *non-reductive* representationalism. Within the non-reductive camp we can distinguish two sorts of view. Firstly,

dualist representationalism: phenomenal conscious involves non-physical properties with no metaphysically necessary connection to the physical world. According to this view there is a nomological correlation between phenomenal character and representational content. Chalmers defends a theory of this kind, but there is also a disagreement between Chalmers and Tye and Dretske over *which* representational contents correlate with phenomenal character. In principle, however, a dualist representationalist could agree with Tye and Dretske about the *correlation* between phenomenal character and representational content in the actual world, disagreeing only with regard to worlds differing nomologically from actuality.

Secondly, *supervenient representationalism*: phenomenal character *supervenes* on representational content. According to this view it is metaphysically necessary that no two experiences differ in phenomenal character without differing in representational content, but the converse would not hold. A dualist could make a weaker, nomological claim. Each kind of representationalist claim could also be restricted to a population, or to an individual. In principle both reductionists and dualists could accept restricted views, though each would have to explain the ground of the difference between different populations or individuals. I lean strongly toward reductive representationalism but for much of the paper I shall present my theory as a general framework consistent with some of the non-reductive variants discussed above.

By *standard representationalism* I mean the view that phenomenal character depends entirely on externalistically individuated direct representational content, not including modes of presentation or indirectly represented contents of any kind. I shall call this kind of content *Russellian*.² So, for example, when one has a visual experience as of a black cat against a green background, and nothing else, the Russellian content of one's experience is made up entirely from the objects and properties in the world external to the subject, i.e. *black, cat, green*, etc. According to standard representationalism it is Russellian content alone that determines phenomenal character. It is natural to think of this as analogous to Russell's view that denotation alone determines meaning. It strikes me that

many moves that have been made in the debate about representationalism correspond to moves that were made in the debate over linguistic meaning post-Russell, substituting linguistic meaning for phenomenal character. Non-reductive versions of representationalism also count as versions of standard representationalism insofar as they deal only with Russellian content.

Now to objections. If inverted spectra are possible then standard representationalism is immediately in trouble. On the face of it, to defend a view that was not restricted to individuals one would have to say that when two individuals' visual spectra were inverted relative to one another only one of them could have veridical perceptions. This is counterintuitive; if the two individuals were otherwise identical then it is hard to see what could make one perception veridical rather than the other.

The possibility of inverted spectra that make no functional difference is controversial; materialists (myself included) typically deny that they are possible. The inverted spectrum objection does not, however, require inverted spectra that make no functional difference. This is illustrated by Ned Block's (1990, 1995) *Inverted Earth* thought experiment. Inverted Earth is identical to Earth in every detail except that each Inverted Earth object has the complementary colour to its Earthly duplicate. On Inverted Earth the sky is yellow, ripe tomatoes are green, ripe bananas are blue, and so on. Now, suppose you were to travel to Inverted Earth but while you travelled colour-inverting prisms were placed in your eyes, or your visual system was rewired such that each colour experience became the complementary colour experience.³ Your own bodily colours were also exchanged for their complements. When you arrived on Inverted Earth your experiences would match the experiences you would have had in the corresponding spatiotemporal location on Earth. If you had slept through the journey you would not even be aware of being in a new environment. If this happened then it would seem natural to say that upon waking your perceptions would be illusory; the sky would look blue to you when really it was yellow. But if you remained in your inverted state on Inverted Earth for long enough then according to Block it would no longer seem plausible to say

that you perceived incorrectly. The content of your experiences would shift just as, arguably, the meaning of a word eventually shifts when it is consistently used in relation to a new denotation, albeit unbeknown to the speaker. Consequently there can be experiences that differ in Russellian representational content but not in phenomenal character. Hence phenomenal character at most supervenes on Russellian representational content.

Worse still, however, there can be experiences differing in phenomenal character that do not differ in Russellian representational content. Suppose that on Earth in your original state you saw an unripe, green tomato. Years later on Inverted Earth, where ripe tomatoes are green, you saw a ripe tomato. In both cases your experience correctly represented the colour green. But the phenomenal character differed between the two experiences because of the colour inversion of your visual apparatus in the intervening period. Hence Russellian representational content and phenomenal character can vary independently of one another, so all versions of standard representationalism are false.⁴

Not all objections to standard representationalism are as effective as Inverted Earth. Tye (2002, 2003) shows how a number of objections regarding blurry images, after images, hallucinations, aspect switches and other phenomena can be dealt with. In my opinion these responses are largely convincing. But Inverted Earth strikes me as fatal to standard representationalism.

Block (1995, p. 25) has put forward a further objection to which even most non-standard versions of representationalism are vulnerable. This involves a thought experiment in which there is phenomenal character yet no representational content whatsoever. The scenario combines Davidson's (1987) swampman with the brain in a vat whose entire existence has been spent in the vat. Davidson held that the swampman, created instantaneously by chance from stray molecules, lacked representational content at the instant of creation because of its lack of causal-historical relations to its environment. But, as Block notes, this would only be correct given a causal chain theory of content; many other theories would allow that the swampman had representational content

because the swampman's internal states could track states of its environment counterfactually.

Brains in vats can also have representational contents. Many theories of content allow that the internal states of a brain in a vat represent features of the computer environment to which it is connected. But Block invites us to consider a brain created by chance, *à la* swampman, in a bathtub and not connected to a computer or anything else. Assuming that phenomenal states at least nomologically supervene on brain states the brain should have *some* kind of experience for a few fleeting moments. Yet since it is not and has never been connected to any perceptual apparatus its internal states do not track states of its environment and consequently represent nothing.⁵

3. Functionalism and representation

I shall now show why anyone who accepts even a minimally functionalist theory of consciousness should reject standard representationalism. By a *minimally* functionalist theory I mean any view according to which phenomenal character and functional role correlate, at least within an individual, across all worlds nomologically equivalent to the actual world. Even most property dualists are functionalists in this minimal sense. I accept a much stronger functionalist claim about conscious experience, but the present argument does not require it. I make no claim here about propositional attitudes.

Consider a phenomenal state. The functional correlate of this state is a member, x_i , of the n -tuple of internal states $\langle x_1 \dots x_n \rangle$ that jointly satisfy the Ramsified psychological theory relating mental states to one another and to inputs and outputs (Lewis 1972). Very crudely speaking, we can think of a functional role as involving three components: *in*, *out* and *sideways*. The *in* component consists of the sensory inputs that typically cause the mental state and the *out* component concerns the motor outputs typically caused by it. The

sideways component consists of the typical causal interactions between x_i and the other members of $\langle x_1 \dots x_n \rangle$.

This is an oversimplification because the three components are not independent of one another; in particular the motor outputs to which an internal state contributes typically result from interaction between several internal states. Nonetheless, some internal states can be thought of as standing at the end of an afferent pathway from a perceptual organ. In such cases the typical cause of the state can be specified without reference to the interaction between the state and other internal states. It is not plausible that our brains are so radically Quinean as to make the causation of each internal state by external influences depend on interactions with every other part of the system.⁶ The behavioural effects of x_i must normally coordinate with sensory inputs, but the manner in which this is achieved need not be specified when describing the afferent pathway leading to the perceptual state x_i .

Now consider the representational content of x_i . In what follows I shall assume that representation has close connections with the notion of *information* associated with Claude Shannon's (1948) work and brought to the attention of philosophers by Fred Dretske (1981). Crudely speaking, events of type A carry information about events of type B if and only if the occurrence of events of type A covaries with the occurrence of events of type B (this typically involves events of type B *causing* events of type A, though it does not have to). So, for example, the sound of a smoke alarm carries information about, and thus represents the presence of smoke because the alarm sounds when, and only when, there is smoke in the vicinity. This may be a rather crude assumption but it will suffice for present purposes. While attempts to define representation *purely* in terms of covariance are problematic it is nonetheless common ground among many theories of representation that covariance is an important factor in determining what a state represents. Even theories that deny this, however, such as causal chain theories, are generally compatible with most of what I shall claim, though I shall not pursue this here.⁷

I shall assume, then, that each member of $\langle x_1 \dots x_n \rangle$ represents whatever it covaries with. Insofar as x_i covaries with external states of affairs it is the *in* component of x_i 's causal role that determines this. If x_i is typically caused by input I_j then x_i represents the state of affairs associated with I_j , at least provided that the subject's behavioural outputs match up adequately with perceptions. So, for example, if x_i is typically caused by the presence of a ripe tomato in front of the subject in normal lighting conditions then this state of affairs is the Russellian content of x_i , at least provided that the subject's resulting actions are directed, where appropriate, toward that state of affairs. But this fixes only part of the functional role of x_i . The *sideways* component of the functional role, and part of the *out* component, could vary independently of the *in* component. A great deal is left open regarding the interactions of x_i with other internal states. Two individuals could thus have internal states that differed in functional role yet represented the same states of affairs. Presumably phenomenal character does not correlate with just the *in* component of functional role; consequently phenomenal character can vary independently of Russellian representational content. Hence standard representationalism is false, given minimal functionalist assumptions.

Block's Inverted Earth illustrates much the same point. Cross-wiring in the optic nerve could make it the case that the Earthling and Inverted Earthling each represented the colour red but the internal representational states differed in *sideways* functional role in such a way that where one subject experienced phenomenal red the other experienced phenomenal green, and vice versa.

4. Stalnaker's two-dimensional framework

I must now digress in order to explain some relevant features of Robert Stalnaker's (1978, 2001, 2004) two-dimensional modal semantics. In this section I outline some important features of Stalnaker's theory as applied to linguistic utterances and beliefs. Then in the next section I shall apply the same apparatus

to conscious experience. The resulting theory is what I call diagonal representationalism.

Let W_1 be the actual world, in which Le Corbusier was an architect. Let W_2 be a world in which words mean what they do in the actual world but Le Corbusier became a comedian in a double act with Stan Laurel. Let W_3 be just like the actual world except that in W_3 ‘Le Corbusier’ names Stan Laurel and ‘architect’ means comedian. For simplicity, let those be all the worlds. The *horizontal proposition* expressed by an utterance is the set of worlds in which the utterance would be true *given* the meaning it has in the world in which it is uttered. This can be thought of as *what is said* by the utterance. So, for example, when ‘Le Corbusier was an architect’ is uttered in W_1 it is true. When uttered in W_1 but evaluated relative to W_2 it is false. When uttered in W_1 but evaluated relative to W_3 it is true because in W_3 the very same person, Le Corbusier (or a modal counterpart if you prefer), was an architect despite not being named ‘Le Corbusier’. The horizontal proposition expressed by the utterance in W_1 is therefore $\{W_1, W_3\}$, as represented by the two ‘T’ (true) symbols in the first horizontal line of table 1. When the utterance occurs in W_2 the same horizontal proposition is expressed but the utterance is false relative to its own world. When the utterance occurs in W_3 , however, it expresses the same horizontal proposition that would be expressed by an utterance of ‘Stan Laurel was a comedian’ in W_1 . In our example this is the necessary proposition $\{W_1, W_2, W_3\}$, though the necessity comes about because we have considered only three worlds.

		World of evaluation		
		W_1	W_2	W_3
World of utterance	W_1	T	F	T
	W_2	T	F	T
	W_3	T	T	T

Table 1

One of Stalnaker's key insights was to recognise the importance of the *diagonal proposition*. This is the set of worlds we get if we look down the diagonal of the matrix from top-left to bottom-right and include all the worlds against which there is a 'T'. The diagonal proposition for our example is therefore $\{W_1, W_3\}$. This is the set of ways that the world could turn out to be, given only the assumption that the utterance in that world is true. Given a larger set of worlds this would normally differ from any of the horizontal propositions associated with the utterance. Suppose that you hear an utterance of 'Le Corbusier was an architect' but do not know who Le Corbusier was or what an architect is. There are many ways the world might be, for all you know. But if you know that the utterance is true then you narrow down the possibilities. You cannot be in a world in which 'Le Corbusier' refers to Le Corbusier, 'architect' means comedian but Le Corbusier was not a comedian, for example. Since we are not omniscient the diagonal proposition captures a psychologically important component of the content of our utterances or beliefs. I shall suggest below that it also determines the phenomenal character of conscious experiences.

Now consider utterances of 'Le Corbusier was an architect' and 'Charles-Edouard Jeanneret was an architect'. In actuality 'Charles-Edouard Jeanneret' was Le Corbusier's real name. What is said by both utterances is thus the same; they express the same horizontal proposition. But because there are worlds in which the two names refer to different people the corresponding diagonal propositions differ. The set of ways the world might be if 'Le Corbusier was an architect' is true differs from the set of ways the world might be if 'Charles-Edouard Jeanneret was an architect' is true. This explains why the two utterances can differ in cognitive significance. The difference between ascribing architect-hood to someone *qua* Le Corbusier and ascribing it to them *qua* Charles-Edouard Jeanneret makes no difference to the horizontal proposition but it does constitute a difference in the way in which one narrows down one's overall epistemic possibilities.

Although we sometimes consciously entertain metalinguistic thoughts the explanation of differences of cognitive significance just given does not require

that we always do. Consider token beliefs, to which the same account applies. If one believes 'Le Corbusier was an architect' the corresponding diagonal proposition is that one's belief is true. One's tacit acceptance of the diagonal proposition (or rejection, if one disbelieves the horizontal proposition) is evidenced by one's actions being appropriate to its truth, but no conscious higher-order belief is required for this.

Another key feature of Stalnaker's framework is that in individuating an utterance across worlds no restriction need be placed on the way in which the truth value of the utterance, or the references of the terms that comprise it, are determined within a given world. Suppose, for example, that in some worlds the extension of 'architect' is designers of buildings whereas in other worlds its extension is humorous entertainers. Stalnaker's account allows that utterances of 'architect' in all these worlds can count as utterances of the same term. Only the word itself is held constant. This 'metasemantic' account, as Stalnaker describes it, differs from the 'semantic' accounts of other two-dimensionalists including Chalmers.⁸

Although Stalnaker says relatively little about empty terms his theory provides some useful apparatus for dealing with them.⁹ This will be important when we come to consider Block's bathtub-brain. Let us use the symbol '-' to mean 'no truth value'. Consider another set of worlds W_1 , W_2 , and W_3 , in which the same utterance occurs. Suppose that in W_1 a component of the utterance lacks a reference. When the utterance occurs in W_1 the horizontal proposition expressed is the empty set. The set of worlds in which the utterance's negation is true is also the empty set; this distinguishes the utterance from one that expresses a necessary falsehood. The matrix of truth values is shown in table 2.

World of evaluation

		W ₁	W ₂	W ₃
World of utterance	W ₁	-	-	-
	W ₂	T	T	F
	W ₃	F	T	F

Table 2

Although the utterance in W₁ expresses no horizontal content there is, however, a non-empty diagonal proposition provided that the utterance *could* have been true (its terms could have referred). This is again illustrated in table 2; the diagonal proposition is {W₂}. When, in W₁, a subject acts on the belief state expressible by the utterance the subject acts, albeit mistakenly, in a way that would be appropriate in any of the worlds in the diagonal proposition.

There are, however, some subtleties to be noted, especially when the two-dimensional framework is applied to thoughts or, as I intend, conscious experiences. It may appear that the diagonal proposition is a kind of narrow content, available regardless of the layout of the external world, in contrast to the externalistically individuated horizontal proposition. Narrow content is sometimes defined as content determined purely by what is ‘in the head’. This way of defining it can, however, be misleading. I shall claim below that certain diagonal propositions are about neural states and are thus determined purely by what is in the head. But all externalist theories accept that contents *concerning* the subject’s own neural states are fixed by what is in the subject’s head. This is quite unlike the kind of narrow content to which externalists normally object, wherein the narrow content of a thought that putatively concerns an *external* state of affairs is determined by what is inside the subject’s head, perhaps only by the vehicle itself. Unlike the diagonal proposition, narrow content of this latter kind is not a set of possible worlds; it is not really content at all.¹⁰

Nonetheless, certain radically externalist philosophers reject the assumption that content that is available regardless of the existence or non-existence of external objects could be adequate for psychological explanation. Some hold that psychological explanation makes reference to object-dependent *de re* senses that comprise object-dependent Fregean thoughts (Evans 1981, 1982, McDowell, 1977, 1984). Similarly some also hold a disjunctive theory of perception that rejects what McDowell calls the 'highest common factor' conception of thought or experience, according to which the mental or perceptual state of a hallucinating subject is the same as that of a veridically perceiving subject in a subjectively indistinguishable state (Hinton 1973, Snowdon 1980, McDowell 1982). It might appear that the diagonal proposition for a mental state is just such a highest common factor.

The two-dimensional framework outlined above is, however, general and flexible enough to accommodate a variety of theories that individuate content in different ways for different purposes. It was assumed above, following Stalnaker, that we can talk about the *same* utterance (or its counterparts if you prefer) existing in different worlds regardless of the reference of its terms. But this assumption need not be made; nothing in the two-dimensional framework requires it. Suppose, for example, one followed David Kaplan (1990) in holding that words should be individuated as spatiotemporal continuants rather than by written or spoken form. Words with different references (or even the same reference but a different causal history) would count as different words even if they were written or pronounced in exactly the same way. Consequently one might hold that a given utterance would not have existed if any of its terms differed from their actual reference. The same move could be made regarding thoughts.

Where there is no utterance the question of truth or falsity does not arise. Consequently no horizontal proposition is expressed. Suppose that in W_1 'Le Corbusier' refers to Le Corbusier, in W_2 it refers to Stan Laurel and in W_3 it refers to no one. Then if words were object-dependent there would be different words in each of the worlds (or perhaps no word in W_3 , depending on the details of the

account). In that case each utterance of the form ‘Le Corbusier was an architect’ would have its own two-dimensional matrix as shown in table 3 (the truth values have been changed from the examples above).

<table style="margin: auto; border-collapse: collapse;"> <tr> <td style="padding: 0 10px;">W_1</td> <td style="padding: 0 10px;">W_2</td> <td style="padding: 0 10px;">W_3</td> </tr> <tr> <td style="padding: 0 10px;">W_1</td> <td style="padding: 0 10px;">T</td> <td style="padding: 0 10px;">F</td> </tr> <tr> <td style="padding: 0 10px;">W_2</td> <td style="padding: 0 10px;">-</td> <td style="padding: 0 10px;">-</td> </tr> <tr> <td style="padding: 0 10px;">W_3</td> <td style="padding: 0 10px;">-</td> <td style="padding: 0 10px;">-</td> </tr> </table>	W_1	W_2	W_3	W_1	T	F	W_2	-	-	W_3	-	-	<table style="margin: auto; border-collapse: collapse;"> <tr> <td style="padding: 0 10px;">W_1</td> <td style="padding: 0 10px;">W_2</td> <td style="padding: 0 10px;">W_3</td> </tr> <tr> <td style="padding: 0 10px;">W_1</td> <td style="padding: 0 10px;">-</td> <td style="padding: 0 10px;">-</td> </tr> <tr> <td style="padding: 0 10px;">W_2</td> <td style="padding: 0 10px;">T</td> <td style="padding: 0 10px;">T</td> </tr> <tr> <td style="padding: 0 10px;">W_3</td> <td style="padding: 0 10px;">-</td> <td style="padding: 0 10px;">-</td> </tr> </table>	W_1	W_2	W_3	W_1	-	-	W_2	T	T	W_3	-	-	<table style="margin: auto; border-collapse: collapse;"> <tr> <td style="padding: 0 10px;">W_1</td> <td style="padding: 0 10px;">W_2</td> <td style="padding: 0 10px;">W_3</td> </tr> <tr> <td style="padding: 0 10px;">W_1</td> <td style="padding: 0 10px;">T</td> <td style="padding: 0 10px;">F</td> </tr> <tr> <td style="padding: 0 10px;">W_2</td> <td style="padding: 0 10px;">T</td> <td style="padding: 0 10px;">T</td> </tr> <tr> <td style="padding: 0 10px;">W_3</td> <td style="padding: 0 10px;">-</td> <td style="padding: 0 10px;">-</td> </tr> </table>	W_1	W_2	W_3	W_1	T	F	W_2	T	T	W_3	-	-
W_1	W_2	W_3																																				
W_1	T	F																																				
W_2	-	-																																				
W_3	-	-																																				
W_1	W_2	W_3																																				
W_1	-	-																																				
W_2	T	T																																				
W_3	-	-																																				
W_1	W_2	W_3																																				
W_1	T	F																																				
W_2	T	T																																				
W_3	-	-																																				
Utterance in W_1	Utterance in W_2	Utterance in W_3																																				

Table 3

Thus when ‘Le Corbusier was an architect’ is uttered in W_1 it expresses the horizontal proposition $\{W_1, W_3\}$ and the diagonal proposition $\{W_1\}$. Similarly when a formally but not numerically identical utterance occurs in W_2 it expresses the horizontal proposition $\{W_1, W_2\}$ and the diagonal proposition $\{W_2\}$. A little more explanation is needed for W_3 . The kind of theory now under consideration rejects the idea that there is a highest common factor of content between empty and referring cases. To accommodate such a theory it is therefore desirable that the diagonal proposition in empty cases should differ from those in veridical cases. If it is assumed that there is no utterance in W_3 , merely a placeholder for an utterance, then one might think that the whole matrix should be blank because one cannot talk of the same utterance occurring in other possible worlds. There would be no content of any kind in W_3 and hence no possibility of explaining psychological similarities between empty and non-empty cases in terms of content.

Different diagonal propositions may, however, be individuated for the purposes of different kinds of explanation by considering different entities as bearers of content. Consider the inscription or sound under consideration in W_3 . It is not an utterance, given our current (temporary) supposition of an object-

dependence theory of utterances; but it is a *putative* utterance. If matters external to it had been different it would have been an utterance. Hence it still makes sense to talk about the horizontal proposition that it would have expressed, had it occurred in W_1 or W_2 . Consequently it has an associated diagonal proposition: $\{W_1, W_2\}$ in the example given in table 3. This is the proposition that the inscription or sound expresses a truth. Since the same inscription or sound occurs in W_1 and W_2 the same diagonal proposition is associated with the utterances in those worlds, as well as the different diagonal proposition that the *utterance* is true. So for W_1 and W_2 there is a choice of different diagonal propositions suitable for different theoretical purposes. Someone with theoretical reasons for individuating utterances, Fregean thoughts or other such entities in an object-dependent fashion may have reason to pay attention to the corresponding diagonal propositions. But such differences are not apparent to the subject; there is no subjectively discriminable difference between referring and empty utterances or thoughts, or veridical and hallucinatory perceptions in the cases that are of interest. When we wish to capture the subjective states that these subjects have in common we can individuate diagonal propositions in such a way that the same utterance can occur whatever the references of its terms. I shall follow the corresponding strategy for conscious experiences.¹¹

5. Diagonal Representationalism

It is now straightforward to apply the two-dimensional framework to conscious experiences. Utterances are true, false or neither. Similarly experiences are veridical, false or neither (I avoid the word 'illusory' because its normal use incorporates both false and empty cases). For any given experience there is a set of worlds that fit the experience. This is the *horizontal content* of the experience. But each experience also has a *diagonal content* equivalent to the proposition that the experience is veridical.

The diagonal proposition captures the cognitive significance of an utterance. Similarly the diagonal content of an experience captures its phenomenal character. Consider again an Earthling and an Inverted Earthling both observing the same red object but differing in the phenomenal character of their experiences. Their experiences share the same horizontal content because they represent the same external state of affairs but they differ in their diagonal contents because they are different phenomenal types.

Conversely, consider the Earthling and the Inverted Earthling each in their respective home environments observing a ripe tomato. Ripe tomatoes are red on Earth but green on Inverted Earth yet the two subjects have experiences with identical phenomenal character. The horizontal contents of the two experiences differ; one experience represents a red object while the other represents a green object. Should we say that the sameness of phenomenal character of the two experiences reflects sameness of diagonal content? A positive answer would be required for versions of representationalism that entail a one to one correlation between representational content and phenomenal character, but a negative answer could still be consistent with supervenient representationalism. Strictly speaking the experiences of the two subjects do *not* share the same diagonal content as this has been described so far. This is because the experiences are numerically distinct tokens. One could have been veridical while the other was false, had the world been different. Hence only supervenient representationalism can be defended, as things stand. Relative to an individual subject, however, a stronger claim could be defended: for a given subject, diagonal contents and phenomenal characters match one to one. This would make a reductionist theory possible provided the reduction was restricted to an individual subject.

It may nevertheless seem unsatisfactory to have no account of what different subjects' phenomenally identical experiences have in common in terms of representational contents. Fortunately there is a way to give such an account. Consider by analogy two utterances of 'Godzilla is nearby'. Because both utterances contain the indexical term 'nearby' there is no guarantee that they

share the same truth-value. Consequently their diagonal propositions differ. Now, conscious experiences are analogous to indexical utterances because they are had by a person at a place at a time. We can however, capture what separate utterances of the same indexical sentence have in common in terms of Quine's notion of *centred worlds*.¹² A centred world consists of a world and a marked space-time point (the 'centre') at which the subject is located. Whereas the truth value of a non-indexical utterance depends only on the world in which it is uttered, the truth value of an indexical utterance depends upon the centred world in which it is uttered. We can define *centred horizontal propositions* and *centred diagonal propositions* in terms of centred worlds in a way that exactly corresponds to the definitions of ordinary horizontal and diagonal propositions in terms of worlds. Two utterances of 'Godzilla is nearby' thus share the same centred diagonal proposition but express different centred horizontal propositions because they occur in different centred worlds (i.e. at different spatiotemporal locations). A corresponding account can be applied to conscious experiences, yielding the conclusion that what the Earthling and Inverted-Earthling's experiences have in common is that they share the same centred diagonal content even though they differ in horizontal centred content because they occur in different centred worlds. Every experience with the same phenomenal character therefore shares the same centred diagonal content no matter which world or subject it belongs to.

Diagonal representationalism accommodates Block's bathtub-brain by analogy with utterances containing empty terms. Although the experiences of the bathtub-brain have no horizontal content in the actual world there are other worlds in which the experience is veridical because the brain is connected to functioning perceptual apparatus. Consequently the experiences have a diagonal content with which their phenomenal character correlates. Block does consider the possibility of ascribing a kind of counterfactual content to the bathtub-brain on the grounds that it could have been able to perceive and track objects had it had perceptual apparatus. He rejects this on the grounds that 'to say that the bathtub-brain has the capacity to track is like saying that men have the capacity

to become pregnant if only they were given wombs' (1995, p. 26). It is not clear that Block has two-dimensional semantics in mind in this passage, but in any case his objection could not defeat diagonal representationalism. It is not necessary for the bathtub-brain to have a *capacity* to perceive. Suppose the bathtub-brain were scanned so that its internal processes could be observed. If one could sufficiently interpret its attempts at producing a behavioural output one might have reason to say that it was behaving as though its internal states were veridical. This does not require us to say that it has a *capacity* to perceive any more than we are required to say that a man who acts as if pregnant has a capacity to become pregnant.

6. Diagonal representationalism and higher-order representational theories

Higher-order representational (HOR) theories say that a mental state is conscious only when accompanied by, or disposed to be accompanied by a higher-order state that represents its occurrence. The higher-order state is not itself conscious unless accompanied by a further higher-order state that represents its occurrence, and so on. These theories sub-divide into higher-order thought (HOT) theories and inner-sense theories.¹³ HOT theories, according to which the higher-order state is the belief that one is in the first-order mental state, further sub-divide into non-dispositionalist and dispositionalist HOT theories. The former, whose chief advocate is David Rosenthal (1986, 1997, 2002), claim that a higher-order belief must be present in order for a state to be conscious. The latter, whose advocates include Daniel Dennett (1978, 1991) and Peter Carruthers (1996, 2000), make only the weaker claim that for a state to be conscious it must be *available* to higher-order thought. Inner-sense theories, whose advocates include David Armstrong (1968, 1984) and William Lycan (1996), take the higher-order state to be a kind of inner-perceptual or

introspective state produced by an inner-monitoring mechanism, rather than a belief.

Diagonal representationalism can be thought of as a kind of HOR theory. The diagonal content associated with an experience is the proposition that the experience is veridical. This higher-order content entails the higher-order proposition that the experience occurs. Higher-order content is thus essential to conscious experiences in at least the sense that phenomenal character correlates with a certain higher-order content. I shall argue, moreover, that in order for an experience to be conscious its diagonal content must be *represented*, and that the vehicle for this representation consists in the set of brain states normally caused by the experience. This in turn makes diagonal representationalism a version of functionalism as well as a HOR theory.

Standard examples such as 'pain is/is realised by a firing of C-fibres' could easily give the impression that whenever the relevant internal state obtains the subject is *ipso facto* in the corresponding conscious state. This is misleading. There would be no conscious experience if an internal state occurred without causing effects on the rest of the brain. Consider a sharp pain. Could it be felt yet produce no other internal events whatsoever? I assume not. A neuronal event is conscious only by virtue of the effects it produces in the rest of the brain. A functionalist account (of the stronger variety) allows for this, however, by defining the internal states $\langle x_1 \dots x_n \rangle$ in terms of an overall causal role that includes interaction with other internal states.

This provides a clue to the way in which diagonal content is represented in the brain. It was mentioned above that notwithstanding a degree of oversimplification it can be useful to think of the causal role of a mental state as resolvable into *in*, *out* and *sideways* components. As suggested above, we can think of horizontal content as determined by the *in* component alone. But, I shall argue, it is the *sideways* component of the causal role that determines diagonal content. This fits well with functionalism, for it is the sideways component of causal role that should predominantly determine phenomenal character.

When a firing of neurons occurs there may be many effects on other parts of the brain. Some of these are normally caused by a state of that type, while others occur as one-off effects due to the state that the brain is in at the time. When one feels a sudden intense pain in one's left ear, for example, a number of modifications to one's brain state are standardly caused by the pain state. Changes occur in parts of one's brain controlling facial muscles, making one disposed to wince, and so on. Other neural events could have the same effects; one may decide to wince voluntarily, for example. For each first-order state, however, there is presumably a unique set of neuronal effects that covaries with it. If one happens to be thinking about mathematics when one's ear hurts one's train of thought is likely to be interrupted. But the corresponding neuronal changes do not belong to the covariant set because one would not always be thinking about mathematics when one's ear hurts. They are one-off effects.

If a state, or set of states, represents what it covaries with then the set of covarying states together acts as a vehicle representing the occurrence of a first-order state. This need not take the form of a belief, though it might contribute to the formation of a higher-order belief. I shall not take a stand here on whether higher-order beliefs or self-monitoring mechanisms are necessary for consciousness. Instead I suggest only that the kind of higher-order representation implied by diagonal representationalism is necessary for consciousness. I remain neutral on the sufficient conditions.

So far we have a reason to believe that there are higher-order representations of the *occurrence* of experiential states. But why should their *veridicality* be represented? It must be *possible* for other brain states to represent a first-order state as veridical, otherwise it would not be possible to believe that one's first-order states were veridical. Now, in a naturally evolved organism the function of a perceptual system is to represent the environment correctly. Consequently the information that it produces is treated as veridical by default. What it means to say that the information is *treated* as veridical is that the actions caused by it are normally those that would be appropriate to the interests of the organism given that the content of the experience was veridical. Thus the normal covarying

effects of an experience represent it as veridical, as shown by the actions to which they give rise.

What about an organism whose actions were out of step with its perceptions? Suppose for example that someone was cross-wired in such a way that their actions were inverted left to right from those that would be appropriate to their perceptions. A human in this position would soon realise that their actions were not those that were intended. A process of feedback would eventually result in the actions being corrected. The experiences would not be treated as false. This provides a further reason for thinking that first-order experiential states are represented as veridical by the states they cause. This also applies to Block's bathtub-brain; although it cannot perform actions in any normal sense it has a mechanism that would coordinate its actions with perceptions if genuine perceptions and actions became possible. Hence even the bathtub-brain has a representation of the veridicality of its experiential states.

What about a creature whose actions were uncoordinated with its perceptions and who had no internal mechanism to coordinate them? What one should say would depend on the details of the case, but any uncertainty about the diagonal content represented would be matched by an uncertainty about the phenomenal character of its experiences. Perhaps there is nothing it is like to be such a creature at all.

7. Representationalism and the Cartesian theatre

The literature on representationalism is full of reference to *the* phenomenal character of an experience, as though items with determinate phenomenal properties appeared before the mind like actors on a stage. This picture of consciousness, dubbed the *Cartesian theatre*, has been heavily criticised, perhaps most notably Daniel Dennett (1988, 1991). Those of us who find these criticisms convincing therefore have a further reason to be suspicious of standard representationalism.

Consider Dennett's (1988) thought experiment involving the expert coffee tasters Chase and Sanborn. When Chase and Sanborn started their job they both enjoyed drinking the coffee manufactured by their employer. But now, a few years later, although the coffee itself has not changed Chase and Sanborn find that they no longer enjoy drinking it. Although both their perceptual systems have changed in the same way, however, they make different claims about the changes in their experiences.¹⁴ Chase claims still to experience the same *taste* but no longer enjoys it; the phenomenal character is the same as before but the reaction is different. Sanborn, by contrast, claims that although the original taste would still be just as enjoyable as before it has been replaced by a different, less enjoyable taste. So the reactive dispositions are unchanged but the phenomenal character is different.

Dennett claims that there is no fact of the matter about who is right. He thus denies the existence of *qualia* – experiential items with *intrinsic* phenomenal characters that are constituted independently from the subject's dispositions to react to them. Qualia, thus construed, are the actors on the stage in the Cartesian theatre. To reject them is to reject the Cartesian theatre.

Consider what a standard representationalist like Tye must say about Chase and Sanborn. There seems no reason to say that either of them misperceives the coffee at any time. Their job, as coffee tasters, is to make sure that the taste of the coffee remains the same and they remain just as good at this as ever. The Russellian content of their experiences therefore remains unchanged. But if, as Tye (2002, p. 448) and other standard representationalists claim, no two experiences can be alike in Russellian content yet differ in phenomenal character then the phenomenal character of Chase and Sanborn's coffee experiences must have remained constant. Chase must be correct and Sanborn must be mistaken. Standard representationalism is thus committed to a determinate fact of the matter about Chase and Sanborn's experiences.

Now consider what diagonal representationalism can say. The borderline between states that belong to the covarying set of effects of a first-order state and those that do not is likely to be vague and open to interpretation by the

subject. There may be states that are caused sometimes but not always, or there may be states that are caused regularly for some time but which later cease to be caused. This corresponds to an indeterminacy concerning where to draw the boundary between the phenomenal character of a state and reactions to it.

In my view, as in Dennett's, the phenomenal character of a state is a matter of the subject's reactive dispositions when in that state. When Chase and Sanborn taste the coffee a complex neural state representing certain properties of the coffee is produced.¹⁵ The occurrence of this state causes a set of secondary states, a subset of which covaries with the first-order state. Over time, however, due to physiological changes in Chase and Sanborn the set of secondary states changes. While a substantial subset remains the same some of the original pleasurable secondary states are replaced by different, less pleasant ones. In effect Chase and Sanborn then differ in the interpretation they give regarding membership of the covarying set of states, though obviously they do not think about the matter under that description. Chase interprets the set of covarying states as containing only those that did not change over time. The states that did change are regarded as relatively one-off effects. On this interpretation the same diagonal proposition is represented at all times, corresponding to the unchanged phenomenal character.

By contrast, Sanborn in effect includes some states within the covarying set that Chase excludes. Consequently the vehicle for the diagonal content changes over time. Moreover, the first-order state itself changes because its causal role changes. Consequently the diagonal content changes (a *different* state is now represented as veridical) and so does the phenomenal character. It might seem obligatory to regard the causal role of the first-order state as having changed because it produces different effects from those it produced when Chase and Sanborn enjoyed the coffee. But it can be interpreted (implicitly, by Chase) as having an unchanged causal role because of the abstraction from the physical details of implementation inherent in functionalist theories. The difference between Chase and Sanborn thus consists in a difference in what they take to be essential to the causal role of the first-order state. There is no fact of the matter

about which set of states *really* covaries with the first-order state, just as there is no fact of the matter about whether it is the taste or the reaction to the taste that changes. There are different semantic interpretation schemes that explain what happens equally well, and correspondingly there are different phenomenal interpretation schemes that explain what happens equally well. There is no need to look for a fact of the matter about which is correct; no such fact could be found.

8. Indirect realism?

Bertrand Russell once wrote: ‘what the physiologist sees when he looks at a brain is part of his own brain, not part of the brain he is examining’ (1927, p. 383). This is a vivid expression of a materialistic indirect realist theory of perception. Is diagonal representationalism committed to indirect realism?

No. The conscious states are the first-order states, not the higher-order ‘diagonal’ states. When one is in such a state one is aware of the external world, not a brain state. What it is like to be in that state depends on which other states are caused but these further states need not constitute an explicit *awareness* of the first-order state. One can of course distinguish mediation by perceptual objects (with which Russell was concerned) from mediation by psychological states.¹⁶ As explained above, however, the general framework developed here allows diagonal content to be individuated in such a way as to yield a disjunctive theory of perception with no ‘highest common factor’ psychological state. Diagonal representationalism is thus consistent with the view that perception is diaphanous, but diaphanousness can now be seen to be compatible with differently constituted subjects having different perceptual experiences when surveying the same scene.

9. Conclusions

The importance of diagonal representationalism lies mainly in three claims: firstly, it says that phenomenal character correlates with representational content, even though the content in question is not horizontal content. This should offer encouragement to those who seek a naturalistic reduction of phenomenal consciousness to representational content. Secondly, it shows that there is something correct about HOR theories, though it is left open which, if any, of the different versions is correct. And, thirdly, it shows how to bring representationalism and HOR theories together with a broadly functionalist account of consciousness.¹⁷

NOTES

¹ A full list of those who have defended representationalism or related views would be very long, but the following is a representative sample: Byrne 2001, Clark 2000, Crane 2003, Egan (forthcoming), Harman 1990, Kriegel 2002, Levine 2003, Thau 2002. Not all of these defend *reductive* claims (see below).

² Cf. Chalmers 2004b, p. 167.

³ In later versions of his argument Block (e.g. 1995) has changed some details in order to avoid certain objections. Since I find Block's argument convincing with regard to standard representationalism I shall gloss over this. I make some small changes for ease of exposition.

⁴ Those prepared to assume that phenomenal character correlates with physical state can reach the same conclusion by supposing instead that Inverted-Earthlings are congenitally colour-inverted relative to Earthlings. One can then compare Earthling and Inverted-Earthling experiences. This avoids any controversy regarding shifts of content over time. Tye (2000, pp. 136-40) resists Block's objections by trying to make it plausible that the transported Earthling continues to misperceive. I do not find this plausible; but Tye would also have to say that congenitally colour-inverted Inverted Earthlings would misperceive colours throughout their lives. This seems *really* implausible.

⁵ Non-standard versions of representationalism that I take to be refuted by Block's argument include Sydney Shoemaker's (2000) theory according to which an experience represents a *phenomenal property* of an object, a disposition to produce an experience of a certain kind in a certain kind of subject. While avoiding certain problems faced by Shoemaker's theory Andy Egan's (forthcoming) related theory seems to me also to be vulnerable to this objection. The bathtub-brain cannot represent Shoemaker's phenomenal properties or Egan's appearance properties because it has no perceptual apparatus and thus nothing determines which such property is represented by a given experience. Chalmers' (2004b) *Fregean Representationalism*, which shares some key features with diagonal representationalism, avoids the problem. Unfortunately I do not have space to discuss the relation between Fregean and diagonal representation here.

⁶ Consider for example the pathway from the eyes to the occipital cortex and beyond (see below on indeterminacy regarding where the pathway ends). For the use of 'Quinean' above, and for arguments in support of the claims made above see Fodor 1983.

⁷ Notoriously a full account of representation in terms of information involves more than simple correlation. Some theories appeal to teleology (Millikan 1989, Dretske 1994). I hope, however, that I shall be permitted to side step these issues in what follows. Insofar as content is conceptual

there may be further requirements; I make some general remarks below about how the general framework to be developed can incorporate theories on these matters.

⁸ See Stalnaker 2001, 2004 for the ‘semantic/metasemantic’ distinction. A similar point is made in Stalnaker 1999, pp. 14-16.

⁹ Stalnaker (1999, pp. 92-5) does briefly discuss empty names in a discussion of negative existentials. Negative existential statements may, however, be a special case. It might be held, for example, that negative existential statements containing empty names are true (Stalnaker appears to hold this view). But it is not clear that other statements containing empty names would thereby have truth values.

¹⁰ Cf. Evans’s (1982, pp. 201-2) objections to Fodor’s (1980) notion of narrow content.

¹¹ Where the diagonal proposition is the same across all worlds regardless of the reference of its terms it is similar to what Chalmers (2004a, pp. 169-70) calls the *orthographic contextual intension*. If, on the other hand, we assume that words with different references are different words then the resulting diagonal proposition is similar to what Chalmers calls the *linguistic contextual intension* (pp. 170-1).

¹² For discussion of centred worlds in relation to the semantics of indexicals see Quine 1969 and Lewis 1979. Stalnaker (1981) raises some objections to Lewis’s account but I do not think these objections affect the use being made of centred worlds here.

¹³ I am following Peter Carruthers’ (2001) terminology. Inner-sense theories are sometimes known as higher-order experience or higher-order perception theories.

¹⁴ Dennett also considers scenarios in which Chase and Sanborn undergo slightly different physiological changes, but we need not consider this here.

¹⁵ For an illustration of what kind of complex neural state might be involved see Paul Churchland’s (1995) discussion of *taste space*.

¹⁶ See John Foster 2000.

¹⁷ I would like to express my thanks to David Chalmers and Katherine Hawley for their comments on written versions of this paper and to audiences in St Andrews and Durham for their feedback.

REFERENCES

- Armstrong, D. (1968): *A Materialist Theory of the Mind*. London: Routledge.
- Armstrong, D. (1984): Consciousness and causality. In D. Armstrong and N. Malcolm (eds.) *Consciousness and Causality*. Oxford: Blackwell.
- Block, N. (1990): Inverted Earth. In James Tomberlin (ed.) *Philosophical Perspectives 4: Action Theory and Philosophy of Mind*. Atascadero: Ridgeview, 53-79.
- Block, N. (1995): Mental Paint and Mental Latex. *Philosophical Issues*, 7, 19-49.
- Byrne, A. (2001): Intentionalism Defended. *Philosophical Review*, 110, 199-239.
- Carruthers, P. (1996): *Language, Thought and Consciousness*. Cambridge: Cambridge University Press
- Carruthers, P. (2000): *Phenomenal Consciousness: a naturalistic theory*. Cambridge: Cambridge University Press.
- Carruthers, P. (2001): Higher-Order Theories of Consciousness. *The Stanford Encyclopedia of Philosophy (Summer 2001 Edition)*, edited by Edward N. Zalta. URL = <<http://plato.stanford.edu/archives/sum2001/entries/consciousness-higher/>>.
- Chalmers, D. J. (2004a): Epistemic Two-Dimensional Semantics. *Philosophical Studies*, 118, 153-226.

-
- Chalmers, D. J. (2004b): The Representational Character of Experience. In Brian Leiter (ed.) *The Future for Philosophy*. Oxford: Oxford University Press.
- Chalmers, D. J. (2006): Perception and the Fall from Eden. In Tamar Gendler and John Hawthorne (eds.) *Perceptual Experience*. Oxford: Oxford University Press.
- Churchland, P. M. (1995): *The Engine of Reason, The Seat of the Soul: A Philosophical Journey into the Brain*. Cambridge, MA: MIT Press.
- Clark, A. (2000): *A Theory of Sentience*. Oxford: Oxford University Press.
- Crane, T. (2003): The Intentional Structure of Consciousness. In Q. Smith and A. Jokic (eds.) *Consciousness: New Philosophical Essays*. Oxford: Oxford University Press.
- Davidson, D. (1987): Knowing One's Own Mind. *The Proceedings and Addresses of the American Philosophical Association*, 60, 441-58.
- Dennett, D. C. (1978): Toward a cognitive theory of consciousness. In C. W. Savage (ed.), *Minnesota Studies in the Philosophy of Science*, 9. Reprinted in D. C. Dennett, *Brainstorms*. Cambridge, MA: MIT Press, 1978.
- Dennett, D. C. (1988): Quining Qualia. In A. Marcel and E. Bisiach (eds.) *Consciousness in Contemporary Science*. Oxford: Oxford University Press, 42-77.
- Dennett, D. C. (1991): *Consciousness Explained*. London: Penguin.
- Dretske, F. (1981): *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press. Reprinted Stanford: CSLI Publications, 1999.
- Dretske, F. (1993): Conscious Experience. *Mind*, 102, 263-83.
- Dretske, F. (1994): If You Can't Make One, You Don't Know How it Works. In P. French, T. Uehling and H. Wettstein (eds.) *Midwest Studies in Philosophy*, 19, 468-82.
- Egan, A. (forthcoming): Appearance Properties? *Noûs*.
- Evans, G. 1981: Understanding Demonstratives. In Herman Parret and Jacques Bouveresse (eds.) *Meaning and Understanding*. Berlin/New York: De Gruyter, 280-303. Reprinted in Evans's *Collected Papers*. Oxford: Clarendon Press, 1985.
- Evans, G. 1982: *The Varieties of Reference*. Edited by John McDowell. Oxford: Clarendon Press.
- Fodor, J. A. 1980: Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology. *The Behavioural and Brain Sciences*, 3, 63-73.
- Fodor, J. A. 1983: *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.
- Foster, J. 2000: *The Nature of Perception*. Oxford: Oxford University Press.
- Harman, G. 1990: The Intrinsic Quality of Experience. In J. E. Tomberlin (ed.) *Philosophical Perspectives Vol. 12: Action Theory and Philosophy of Mind*. Atascadero, CA: Ridgeview Publishing.
- Hinton, J. M. 1973: *Experiences*. Oxford: Oxford University Press.
- Kaplan, D. 1990: Words. *Aristotelian Society Supplementary Volume*, 64, 93-119.
- Kriegel, U. 2002: Phenomenal Content. *Erkenntnis*, 57, 175-98.
- Levine, J. 2003: Experience and Representation. In Q. Smith and A. Jokic (eds.) *Consciousness: New Philosophical Essays*. Oxford: Oxford University Press.
- Lewis, D. 1972: Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy*, 50, 249-58.
- Lewis, D. 1979: Attitudes *De Dicto* and *De Se*. *Philosophical Review*, 88, 513-43.
- Lycan, W. 1996: *Consciousness and Experience*. Cambridge, MA: MIT Press.
- McDowell, J. 1977: On the Sense and Reference of a Proper Name. *Mind*, 86, 159-185. Reprinted in McDowell 1998.
- McDowell, J. 1982: Criteria, Defeasibility, and Knowledge. *Proceedings of the British Academy*, 68, 455-79. Reprinted in McDowell 1998.

-
- McDowell, J. 1984: *De Re Senses*. *The Philosophical Quarterly*, 34, 283-294. Reprinted in McDowell 1998.
- McDowell, J. 1998: *Meaning, Knowledge, and Reality*. Cambridge, MA/London: Harvard University Press.
- Millikan, R. G. 1989: Biosemantics. *Journal of Philosophy*, 86, 281-97.
- Moore, G. E. 1903: The Refutation of Idealism. In Moore's *Philosophical Papers*. London: Routledge and Kegan Paul.
- Quine, W. V. O. 1969: Propositional Objects. In W. V. O. Quine, *Ontological Relativity and Other Essays*. New York: Columbia University Press, 139-160.
- Rosenthal, D. M. 1986: Two Concepts of Consciousness. *Philosophical Studies*, 94, 329-59.
- Rosenthal, D. M. 1997: A Theory of Consciousness. In N. Block, O. Flanagan and G. Güzeldere (eds.) *The Nature of Consciousness*. Cambridge, MA: MIT Press, 729-53.
- Rosenthal, D. M. 2002: Explaining Consciousness. In D. J. Chalmers (ed.) *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press, 406-421.
- Russell, B. 1927: *The Analysis of Matter*. London: George Allen & Unwin.
- Shannon, C. 1948: The Mathematical Theory of Communication. *Bell System Technical Journal*. Reprinted under the same title (with introductory essay by Warren Weaver), University of Illinois Press, 1949.
- Shoemaker, S. 2000: Introspection and Phenomenal Character. *Philosophical Topics*, 28, 247-273. Abridged version reprinted in D. J. Chalmers (ed.) *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press 2002.
- Snowdon, P. 1980: Experience, Vision and Causation. *Proceedings of the Aristotelian Society*, 81, 175-192. Reprinted in J. Dancy (ed.) *Perceptual Knowledge*. Oxford: Oxford University Press, 1988.
- Stalnaker, R. C. 1978: Assertion. In P. Cole (ed.), *Syntax and Semantics, Volume 9: Pragmatics*. New York: Academic Press, 315-332. Reprinted in Stalnaker 1999.
- Stalnaker, R. C. 1981: Indexical Belief. *Synthese* 49, 129-151. Reprinted in Stalnaker 1999.
- Stalnaker, R. C. 1999: *Context and Content*. Oxford: Oxford University Press.
- Stalnaker, R. C. 2001: On considering a possible world as actual. *Aristotelian Society Supplementary Volume*, 75, 141-156.
- Stalnaker, R. C. 2004: Assertion Revisited: On the Interpretation of Two-Dimensional Modal Semantics. *Philosophical Studies*, 118, 299-322.
- Thau, M. 2002: *Consciousness and Cognition*. Oxford: Oxford University Press.
- Tye, M. 1995: *Ten Problems of Consciousness*. Cambridge, MA: MIT Press/A Bradford Book.
- Tye, M. 2000: *Consciousness, Color, and Content*. Cambridge, MA: MIT Press/A Bradford Book.
- Tye, M. 2002: Visual Qualia and Visual Content Revisited. In D. J. Chalmers (ed.) *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press.
- Tye, M. 2003: Blurry Images, Double Vision and Other Oddities: New Problems for Representationalism? In Q. Smith and A. Jokic (eds.) *Consciousness: New Philosophical Essays*. Oxford: Oxford University Press.