# What is it like to see a bat? A critique of Dretske's representationalist theory of qualia

**Andrew Bailey**
University of Guelph

**Abstract**
This paper critiques the representationalist account of qualia, focussing on the Representational Naturalism presented by Fred Dretske in *Naturalizing the Mind*. After laying out Dretske's theory of qualia and making clear its externalist consequences, I argue that Dretske's definition is either too liberal or runs into problems defending its requirements, in particular 'naturalness' and 'mentalness.' I go on to show that Dretske's account of qualia falls foul of the argument from misperception in such a way that Dretske must either admit that his kind of qualia have nothing at all to do with what mental life subjectively feels like, or that veridical perception involves qualia and misperception does not.

One of the main problems in the philosophy of mind is what is sometimes called the qualia problem. Qualia are the 'felt' or 'phenomenal' qualities associated with experiences, such as the viewing of a colour, the feeling of a pain, or the hearing of a sound. They are sometimes thought of as special properties of certain of our mental states that give those states a certain 'feel' — to know 'what it is like' to have an experience (in Thomas Nagel's phrase) is, traditionally, to know its qualia. The *problem* of qualia can be thought of as the attempt to reconcile such properties with a broadly scientific, physicalist outlook. Unless qualia can somehow be naturalized, it looks very much as if their existence is incompatible with the truth of physicalism. It is not at all clear how the painfulness of pain or the vivid redness of the visual sensation of a ripe apple are to be explained physically, for example. It is hard to see how those properties which make such experiences feel the way they do and not some other way, could be physical properties.

One modern response to the problem of qualia is to abandon physicalism and adopt, instead, a form of property dualism, admitting

that qualia are a special sort of property, in principle resistant to third-person empirical study. This is the solution adopted by David Chalmers, for example. Another response is simply to deny the existence of qualia, and assert, somehow, that there *is* no painfulness of pain or redness of red to be explained. This is a position which is sometimes attributed to Daniel Dennett.

The most interesting response, however, is the attempt to *naturalize* qualia: to show how the feature of the world that makes red visual sensations feel a certain way rather than another can be described and explained using the methods of science. Here there are essentially four theories on the table. First, qualia might be *type-identified* with some suitable third-person-observable property. Second, qualia might be said to *supervene* upon more fundamental physical properties but not be type-identical with them. Third, there is the Higher Order Thought theory, most associated with David Rosenthal (1990), Rocco Gennaro (1996) and Peter Carruthers (1996), which holds that *other* mental states are responsible for making certain mental states conscious, by representing them in certain ways. Finally, there is the representationalist approach, which is the newest of the four attempts to naturalize qualia, and which is perhaps generally held to be the most promising.

This type of view is prominently held by Michael Tye (e.g. 1995), Gilbert Harman (1990), John McDowell (1994), Georges Rey (1992) and William Lycan (e.g. 1996), but arguably its most well-developed defence appears in Fred Dretske's 1995 book *Naturalizing the Mind*. What all these representationalists have in common, roughly, is that they identify phenomenal consciousness not with the neural substrate of the brain's representation of the world, but with a subset of those representations themselves. To be a conscious state, on this view, just is to be a certain kind of representation, and no further appeal to intrinsic properties of the medium of representation is required.[1] As Dretske puts it, all mental facts are representational facts and all representational facts are facts about informational functions (1995, xiii). That is, all mental facts are entirely facts about what information some representation is designed to convey, and are never facts about the nature of that representation itself.

[1] One of the best general discussions of the representationalist approach to consciousness of which I am aware is contained in Seager 1999, chapters 6 and 7.

As an attempt to solve the qualia problem, this approach has a lot of *prima facie* plausibility. If successful it would open up qualia — and, more generally, the ways things perceptually seem to us and other sentient organisms — to study by the third-person methods of the natural sciences, allowing them to be fully naturalized. The subjects of study would simply be representations and their objects, both of which would presumably be classes of wholly physical entities, and we would have no need to worry about *extra* ghostly properties *of* those representations.

What I want to do in this paper is critically examine Representational Naturalism, as it is formulated in Dretske's seminal book, and argue that this approach has no chance of success. The fundamental flaw in the position, I shall try to show, is ironically what appears at first sight to be its greatest strength: it is that the representationalist *externalises* qualia, treating them as properties, not of representations, but of what is represented. Note, incidentally, that this is not merely the claim that the *content* of sensation is individuated by external factors, just as the externalist about belief will insist that two intrinsically identical mental representations can be different beliefs on Earth and Twin Earth. Rather, the representationalist claim about qualia is the stronger one that *qualia themselves* — those properties which traditionally constitute 'what it is like' to have sense experience — are external to the experiencer. We might call this latter kind of externalism 'qualia externalism' rather than 'content externalism,' and it is qualia externalism which fundamentally undermines representationalism.

The alternative for the naturalist, I shall conclude, must be to take qualia seriously as phenomenal properties of internal mental states. Our project, then, should be to understand how certain properties of certain brain states, which are essentially phenomenal from the first-person perspective, can also be described and explained from the third-person perspective.

I shall proceed by outlining some of the essentials of Dretske's theory, and then levelling four objections at it.

## 1. Dretske's account of sense experience

Since Dretske ultimately identifies qualia with the properties that the objects of experience are, in a particular way, represented to us as

having (1995, 65), he therefore approaches his theory of qualia not by directly examining these properties themselves but by building up an account of the representational relation involved. That is: not every representation picks out a quale — only certain types of mental representation do so — and it is these special kinds of representation that Dretske strives to characterize. He calls these representations *sense experiences*.[2]

Dretske's general account of representation is as follows. A system, S, represents a property, F, iff S has the function of providing information about the F of a certain domain of objects. S does so by occupying different discrete states corresponding to the different determinate values of F (1995, 2). Dretske's favourite example of a straightforward representational system is a speedometer: here F, the property represented, is speed; the domain contains only the vehicle to which the speedometer is attached; and the relevant discrete states of the system are positions of the needle on the dial.

Dretske's notion of a function has a strong teleological bent: what property a system represents depends *not* upon what properties it carries information about, but upon what property it is *intended* to represent (1995, 4). For example, a speedometer actually carries information about axle rotation and wheel revolutions per minute, as well as about vehicle speed; but, in virtue of its design, it only *represents* the property of speed. According to Dretske, the *telos* for a function can sometimes be 'natural,' rather than the product of deliberate design. In particular, we can say that some representations derive their 'intended function' from their evolutionary history or through individual learning, rather than by convention (1995, 7). It is only these *natural representations* that are candidates for being mental states, for Dretske. In fact, Dretske asserts that mental states form a proper subset of the natural representations (1995, 8), though — I think significantly — Dretske does not as far as I can tell provide a full characterization of what distinguishes a specifically mental representation from, say, a state of an organism's homeostatic or immune systems.

Not all mental representations are sense experiences of course. First, we must distinguish between what Dretske calls *sensory* and *conceptual* representations: between, for example, the experience of a

[2] 'Sense experience is the primary locus of consciousness. … [P]henomenal experience — the look, sound, taste and feel of things — dominates our mental lives. Remove it completely and one becomes … what? A zombie?' (1995, 1).

red colour patch, and the belief that one is seeing a strawberry. For Dretske this distinction corresponds to another: that between *systemic* and *acquired* indicator functions. In the former case some state of the system regularly corresponds to some property that it is intended to indicate, and so systemically indicates it. In the latter case, some state of the representing system is *assigned* a particular representational content, independently of what it may actually systemically indicate. Consider a simple speedometer mechanism that represents vehicle speed by measuring the rotation of the axle. Some state β of the speedometer dial *systemically* represents an axle rotation rate of N rpm. However in cars with different tire diameters the dial would have to be calibrated differently so that in one car state β might have the *acquired* function of indicating 50 kph and in another 60 kph (1995, 12–14).[3] Similarly, we might say that the human visual system has evolved to systemically indicate the determinate value of some objective determinable, which we call the property of colour, and is in fact 'designed' to distinguish between more than 16 million determinate shades of colour. However our visual systems are, we might say, *calibrated* to a greater or lesser degree of specificity among different individuals, and have the *acquired* function of indicating, presumably, the few hundred named hues, such as red, blue, maroon, aquamarine and buttercup yellow — roughly, those colours for which we have corresponding concepts.[4]

[3] Another of Dretske's examples involves a pressure gauge whose needle varies systemically with air pressure; such instruments are routinely calibrated to show altitude in, say, metres above sea level, and a given pointer position will then, Dretske says, be both a systemic representation of pressure and an acquired representation of altitude (1995, 20).

[4] It is tempting to object at this point that, since on Dretske's theory only natural, systemic functions pick out qualia, it follows that 'we are incapable of experiencing — of being *phenomenally* aware of — a variety of modern-day artefacts and properties. Since, e.g., there were no automobiles around when our perceptual systems were being 'designed' by natural selection, no ancestral perceptual system could have been selected for providing information about automobiles, and hence no natural representation could (presently) have the systemic function of indicating autos.' I think that this objection, when put this way, is mistaken: Dretske has in mind our natural, perceptual ability to discriminate between various different car-sized objects, and in this I think he is correct. A Cro-Magnon man could presumably perceptually discriminate a Toyota Tercel from a Chevrolet Cavalier even though, obviously, he would not have the slightest (conceptual) idea what they were. How-

One more distinction remains to be made to complete this sketch of Dretske's account of sense experience. For Dretske, not all natural systemic representations are experiences; experiences make up the proper subset of such representations that 'service the construction of' acquired representations that can be calibrated to 'more effectively service an organism's needs and desires' — that is, they 'are the states whose functions it is to supply information to a cognitive system for calibration and use in the control and regulation of behavior' (1995, 19). Dretske once again provides a speedometer analogy: in a speedometer, those states 'available for use in the control of behaviour' are the indicator states of the speedometer's dial. In the simple device already described, the system's analogue of experience is its representations of axle rotation speed; its belief-analogues are its representations of vehicle speed. In a more complex speedometer device, information about axle rotation is *combined* with information about the height of the axle above the road surface in order to determine the state of the dial. Here, the system's 'experiences' and 'beliefs' are both about the speed of the vehicle, and information about axle rotation is 'lost' — it is analogous to *non-conscious* information carried during the earlier stages of the processing of perception.

Sense experiences, then, are on Dretske's account systemic, natural representations that underlie the construction of behaviour-regulating acquired representations. *Qualia* — the 'raw feels' of experience — are no more and no less than the properties so represented, and thus are *distinct* from sense experiences and their properties. Dretske writes, for example, that '[t]he Representational Thesis identifies the qualities of experience — qualia — with the properties objects are [systemically] represented as having' (1995, 65).

## 2. What is it like, for Dretske, to see a bat (poodle)?

So, given this account, what can Dretske say about the phenomenal qualities — the 'what is it like-ness' — of conscious experience? For Dretske, what is it like to see a bat … or, to use his own example, a poodle? Dretske parses this into two questions:

ever, I do think worries of this kind conceal a deeper problem, that I attempt to bring out in my objection II of section 4.

a) When does something look like a poodle to someone? This question he interprets as meaning: What is it for some organism to have a specifically *poodle*-like experience?
b) How do things seem to someone experiencing a poodle representation?

His answer to the first question is that something looks (sensorially, rather than conceptually) like a poodle to S if it looks the way poodles normally look to S and if it looks different to S than some proper contrast class (e.g. other dogs). As Dretske puts it, S's 'experience of the dog represents the dog … as having … the manifest properties of poodles, those properties that make poodles look so much different from other dogs' (1995, 67).

The importance of this is that, using only the methods of science, we would now be in a position to definitively answer questions about qualia from the third person. For example, we might want to know whether other people actually experience qualia, and if so whether they are the *same* qualia as I experience. One reason this is a problem is that, generally speaking, we are not confident that other people's *discriminatory abilities* tell us all we need to know about their experiences: that Mary and Marvin are equally good (or bad) at discriminating poodles does not entail that poodles look the same to them. This kind of problem is sometimes known as the problem of inverted qualia: what I experience when I see a blue sky could be just like what you experience when you look at a yellow wall, but we both call the sky 'blue.'

Dretske asserts that his theory can solve the inverted qualia problem. For him, as we have seen, qualia are the properties that an experience has the *teleological function* of systemically representing something as having, whether it is actually performing that function or not. It follows, first of all, that these properties need not 'express themselves' in the behavioural dispositions of the system 'in which they exist' (1995, 72) — Dretske *agrees* that the nature of our experiences can come apart from our discriminatory capacities. Nevertheless, Dretske holds, the nature of our experience is still objectively determinable by the following set of identities:

i) qualia = experienced properties
ii) experienced properties = systemically (etc.) represented properties

> iii) systemically (etc.) represented properties = those properties about which the senses have the natural function of providing information (1995, 72).

Thus, according to Dretske, questions about qualia are really questions about the properties certain representational states of a system have the function of indicating, and as such are answerable using the third-person methods of science.

What about question b): How do things *seem* to perceivers? Dretske suggests that the way an experience represents an object is the way that object would be if the representational system were working correctly (1995, 73). Thus, if some organism is designed to differentiate between poodles and other objects, and that organism hallucinates that all the medium-sized objects around her are poodles, then everything seems to be a poodle to her. Once again, this is a properly naturalizing conclusion. Not only can I now tell *what* bats experience, but (according to Dretske) I can discover exactly how such experiences *feel* to bats simply by examining a bat's representational system in order to identify its proper functions, and then by studying those properties which the system has the function of representing.

Hence:

> Knowing what bats, fish, and neighbours experience is, in principle, no different from knowing how things 'seem' to a measuring instrument. In both cases it is a question of determining how a system is representing the world (1995, 81–82).

For example, Dretske imagines a mono-representational parasite that only has a thermal sense, with which it picks out receptive hosts that have a body temperature of exactly 18°C. To know what it is like for that parasite to sense a receptive host,

> [a]ll you have to know is what temperature is. If you know enough to know what it is to be at a temperature of 18°C, you know all there is to know about the quality of this parasite's experience. … For, if things are working right, what the host is — 18°C — is how things seem to the parasite. So if you want to know how things seem to the parasite, look at the host (1995, 83).

In addition to deriving this position from his general framework, Dretske gives what he takes to be an independent argument for this conclusion:

1) qualia are supposed to be the way things seem in the sense modality in question;
2) things sometimes *are* the way they seem; therefore
3) qualia are exactly the properties the object being perceived *has* when the perception is veridical;
4) the quale of the parasite is just like that it has when its perception is veridical; therefore
5) the quale of the parasite has to be exactly the property the object has, i.e. 18°C (1995, 83–84).

Furthermore, Dretske asserts, one can know what it is like to have a certain experience without being able to discriminate that property yourself. The familiar bat is one example; Mary the colour-blinded scientist in Frank Jackson's well known thought experiment (1986) another. Dretske even claims that one could know just what it is like to hear a musical change of key without being able to recognise one yourself (1995, 85–86).

Here, then, in summary is Dretske's account of qualia. Qualia are completely characterized as the objective, external properties that those mental representations which are our experiences have the natural, systemic function of indicating. Therefore qualia are just as objectively specifiable as are the systemic functions of any physical system.

## 3. Does Dretske have a theory of *qualia*?

Dretske suggests, at the beginning of his book, that his Representational Naturalism 'is the only approach to consciousness that has much to say about the baffling problems of phenomenal experience' (1995, xiii). And indeed, Dretske's account has quite a lot of initial plausibility. Unlike proponents of more traditional functional analyses he seems ready and willing to deal with the 'inverted spectra/earth' or 'knowledge argument' families of counter-examples,[5] and he makes

[5] On the former, see especially Shoemaker 1982 and Block 1990; on the latter see Jackson 1986.

moves intended to prevent his theory being too functionally 'liberal.'[6] Unlike many of the opponents of this kind of naturalistic analysis, Dretske (if his theory is correct) faces no so-called 'explanatory gap' — no serious problem in explaining how qualia can be related to, and explained in terms of, scientific theories of the physical world.[7] Nevertheless, having laid out Dretske's theory of qualia, I now want to raise four objections to it which, I think, show that the fundamental conundrum of qualia is not so easily dispatched.

First, a preliminary skirmish. We have seen that, along with other contemporary qualia representationalists, Dretske holds that qualia are properties of objects as represented by conscious beings; qualitative consciousness is the representing of an object as having qualia. This clearly has the consequence that Dretske is an externalist about qualia: qualia are the properties represented, not properties of the representation. And a consequence of *this* is that Dretske is using the word 'qualia' in a quite radically non-standard way. The standard definition, recall, is roughly that qualia are the qualities of conscious mental states which characterize 'what it is like' to experience things … and which constitute 'feeling like' anything at all.[8] Hence it is usual to take qualia to be putative properties of conscious mental states, perhaps even precisely those properties which make such states *conscious*. Dretske, of course, means no such thing when he uses the term. Secondly, one habitually speaks of qualia as being part of the mental life of the experiencer: organisms either 'have' or 'do not have' qualia, we say. Dretske must be committed to the position that qualia are *non*-mental, since he claims both that only representations are mental and that qualia are *not* representations but what is represented. Hence qualia are (typically) not properties of their experiencer[9] — it

[6] This term was popularised by Ned Block 1980.

[7] For various manifestations of this see Nagel 1974, Jackson 1993, Chalmers 1996, McGinn 1991.

[8] This same basic definition is used by most writers, with a wide range of sympathies: for example, Chalmers 1996, 4; Clark 1993, 1; Dennett 1988, 42–43. As originally conceived, they were properties of phenomenal individuals: see Lewis 1929 and Goodman 1977, 95 ff.

[9] One might want to speak of experiencers 'having' certain qualia associated with their experiences, or even constitutive of the content of those experiences, but

no longer makes any real sense to say that a conscious being 'has' qualia, on Dretske's account.[10]

Finally, qualia are usually supposed to be *phenomenal* properties: the point of talk about there being 'something it is like' to have qualia is that this distinguishes them from non-phenomenal properties such as being a cube, having a dial reading 37 kph, or being, on average 385,000 km from Earth's moon. Although such talk is ill-defined, the kernel of the idea is that one can be a cube and not *feel* or *be conscious* of one's cubehood; there is no sensation corresponding to, let alone constituting, being a certain distance from the moon. By contrast, to have a green quale just is to feel a certain way, to be conscious of a certain sensation. Dretske, obviously, does not restrict qualia to phenomenal properties: any property that some mental system has the systemic function of indicating is a quale, potentially including cubehood and distance to the near-side of the moon.

All of this raises the suspicion that Dretske, although he uses the word, is not really talking about *qualia* at all: that instead of giving a theory of those problematic properties usually picked out by the term, Dretske is simply changing the subject and talking about something else.[11] At this point, however, to simply accuse Dretske and the

this seems to me misleading in this context. After all, we do not speak of Oscar 'having' the property of being water! (See Putnam 1975, especially pp. 223–227.)

[10] Note that one consequence of this is that, for Dretske, sensation does not supervene upon the brain.

[11] In many ways Dretske's notion of *sense experience* is much closer to the standard definition of *qualia* than is his account of qualia, which invites the following response. It is possible, and perhaps more comfortable, to interpret Dretske as not making the identity claim about qualia that I claim he is making, but instead saying something like the following: to have qualia of type *T* is nothing more than to token states that systemically represent something as having property *T*; seeing red is merely having a sense experience of (i.e. one whose function it is to pick out) red, and nothing more. But where then, on this account, *are* the qualia — the phenomenal properties of experience — exactly? What are they properties *of* on this interpretation? There seem to be three options. Either Dretske is an eliminativist about qualia (which he denies, and which is anyway rather uninteresting); or qualia are additional, as yet unmentioned, properties of sense-experiences (in which case they have yet to be described, let alone theorised about); or they are properties of the objects of sense-experience (which is how Dretske explicitly describes them, and which takes us back to my interpretation).

other representationalists of missing the point would be to beg the question in favour of the existence of qualia more traditionally construed. After all, instead of seeing Dretske as changing the subject, we might view him as *re-focussing* the qualia debate in a more fruitful direction. The following four objections try to make the case that this is not so; they suggest that the difficult problems qualia present will not go away so easily.

## 4. Problems for Dretske

### *I. The problem of the demarcation of the mental*

Given Dretske's usage of *qualia*, as we have seen, virtually any property can in principle be a quale. The real work in Dretske's account is being done by the notion of a sense experience, since on his view qualia are merely the objects of such experiences. I shall argue, however, that Dretske's account of sense experience is too loose and thin to bear the added philosophical weight the concept must now carry. As we have seen, for Dretske, anything, *s*, that satisfies all the following conditions is a sense experience:

i. *s* is a discrete state of a (mental) representational system with the function of indicating some determinate value of an objective determinable;
ii. *s*'s indicator function is a natural one;
iii. *s*'s indicator function is a systemic one;

Another suggestion that has been made to me is that Dretske could (or should) be read as giving a 'contextualist' account of qualia, such that qualia are to be identified with worldly properties that are sensorialy presented to a subject, 'but only as they are experienced.' But I cannot see how to make this proposal work. If the idea is that external properties are only to count as qualia *while* they are being experienced, then it seems ad hoc; after all, our experiences do not typically *change* the properties of the things we perceive, and so these properties will *be the same property* whether or not we call them 'qualia.' Conversely, if this proposal is intended to shift attention from the external properties represented to *the way they are presented to us*, then the problem reverts to that of naturalistically accounting for these modes of presentation — i.e. features of the representation rather than of what is represented.

iv. *s* is 'cognitively accessible' in some incompletely specified sense that includes underlying a system of acquired representations which are used to control behaviour.

This definition is in danger of committing Dretske to the position that entities unanimously considered non-sentient experience qualia.

Consider a simple plant that continually sucks up water through its root system until specialized areas of its cell walls reach a certain state; these discrete states of the cell walls, we can suppose, are systemically and naturally linked to changes in pressure with the cells, and furthermore they have been 'intended' by evolution to have this function. At a certain point, as the pressure threatens to burst the cell walls, the pressure-indicators in the cells trigger a change in some more general system of the plant which has, let us say, three 'states': either it represents the pressure in its cells as inadequate, or as acceptable, or as too great. Once this system comes to indicate that the cell pressure is too high, the plant opens pores on its leaves which allow the fluid to transpire; when the pressure in its cells has fallen to a satisfactory level, the pores are closed again.

Such an organism apparently satisfies Dretske's conditions for full-blooded phenomenal experience. On his account, the pressure inside its cells constitutes a quale that is literally experienced in virtue of being connected in the right way to states which match Dretske's definition of natural, systemic representations (changes to the plant's cell walls) and which moreover underlie higher-level states of a homeostatic system that is 'calibrated' to either believe that the pressure is too high or that it is not, and adjusts its pore-opening behaviour appropriately.

Very similar stories could even be told for hypothetical relatively non-complex, non-*living*, prima facie non-conscious systems, such as pieces of computer code that 'evolve' in some kind of artificial environment. Assuming that we are unwilling to ascribe conscious sensory experience to simple plants and evolved computer viruses, it seems untenable to insist that Dretske's definition of *qualia* is in accord with our basic intuitions about the set of things likely to experience phenomenal mental states.

Dretske must presumably reject these counterexamples by showing that — despite what I have said here — plants and computer viruses fail to satisfy his four conditions for being the subject of experience (and therefore of qualia). The only way to do this, that I can

see, would be to assert that once the *mentality* condition is cashed out in more detail, it will exclude them. That is, much more is needed in a satisfactory account of condition iv. than the claim that supplying information for calibration and use in the control and regulation of behaviour is a hallmark of the mental. At best, this unclarity at the heart of Dretske's account is unfortunate — Dretske is, after all, engaged in the project of providing a theory of the mind. At worst, this is circular: only entities with a (conscious) mental life are candidates for possessing experience, where what it is to have a (conscious) mental life is classically understood as being an experiencing subject.

Perhaps it is best, therefore, to treat Dretske not as providing a theoretical treatment of our extant notion of qualia but as giving a principled *redefinition* of the term. However, if *this* is so, then it is far from clear how independently defensible Dretske's conditions might be. After all, now we can no longer say that Dretske's definition is justified simply by providing us with the correct extension for the term, and we are faced with the prospect of demanding a principled defence of each part of the definition. Why should it pick out *these* representations, and not others?

Apart from the incompleteness of Dretske's 'mentality' condition, it is difficult to see how the condition that the indicator function be a natural one can be independently motivated. Why should it make a difference exactly what roots the teleology of a function, as long as it has one? Suppose some advanced race were able to build or replicate living organisms by manipulating molecular raw materials, and imagine that, by chance, their designers hit upon a form identical in every relevant physical way with a human baby. Why should we say that this baby, when grown to adult-hood, does not experience perceptual sensations in the same way that we do? After all, *ex hypothesi*, it interacts with the world in exactly the same way we do, and precisely similar events take place in its brain.[12]

[12] A similar thought-experiment can be levelled against Dretske's claim that any teleology is involved at all. Suppose, *per impossibile*, that our human child were not designed, nor evolved, but created instantly — like Davidson's Swampman — from molecular raw materials by a freak lightening storm on some distant planet barren of sentient life (though not inhospitable). Again, the unfortunate baby would have the same causal connections to the world that we do, and would pass through brain states of the same type as ours — increased activity in striate cortex areas V1 and V2 after input from the Lateral Geniculate Nucleus (LGN), for example. Yet Dretske would have it that, because of the accident of its birth, this organism undergoes no

But if the 'naturalness' condition cannot be defended, then Dretske's class of beings able to experience qualia inflates even more drastically to include, for example, simple gauges like speedometers and thermometers as long as they are connected to appropriate 'cognitive' behaviour-control systems, such as the on-board computer on a late model Toyota. Would we be willing to say that a motor car has a mental life, or do we prefer to insist that 'having' qualia is separable from having a mental life? Neither position, unfortunately for Dretske, is very tenable.

## *II. The problem of individuating qualia*

We have seen that Dretske identifies qualia with the external, objective properties whose determinate values our experiences have the function of discriminating, and that the main advantage of this position is that it makes qualia themselves susceptible to empirical study. However, as Dretske admits, it is not always easy to identify those properties; for example, it is not always possible to determine the proper function of our experiences. Dretske takes this to be at base an empirical problem (1995, 88 ff.): thus, identifying the objective property of colour is, according to Dretske, straightforwardly a matter of discovering what property in the world our relevant visual apparatus evolved to indicate, although this may since have become confused by the phenomenon of metameric matching[13] and so on. However, there are reasons to believe that this uncertainty is actually a serious *conceptual* problem with Dretske's account.

First, though there *may* be a finitely describable set of determinate physical conditions which bring about every instance of an experience of the colour red — which is itself a rather dubious claim — these physical conditions *still* need not constitute an objective external physical property suitable for third-person examination. It is not implausible that our experiences of colour — and not just our colour judgements — are influenced by psychological factors, such as our expectations and other tacit beliefs, which affect *pre-conscious visual*

sense experience whatsoever; that all is dark inside its head, while ours — physically identical — is alight with visual, aural, tactile phenomena. This seems deeply implausible. However, since Dretske explicitly addresses this point in his book, and believes he has a reply to it (1995, 141 ff.), I shall not pursue this line here.

[13] That is, the phenomenon where a wide variety of objective circumstances can give rise to the same colour experience.

*processing*. Thus, for example, we experience the skin-colour of Caucasian people standing beneath the canopy of a spreading elm tree in midsummer as some shade of more or less pinkish brown; however, as a photograph in which the tree is not visible would reveal, their 'actual' skin-colour has a greenish tinge, due to the filtering of light through the leaves. One possible explanation for this phenomenon is that we have certain built-in expectations about the constancy of skin colour, and our brain 'filters out' the greening effect *before* we can become conscious of it.[14] Similarly, some of our colour experiences are probably 'distorted' by the physiological structures of our colour processing systems: optical illusions such as the Von Bezold spreading effect,[15] or apparent colour changes in the face of simultaneous chromatic contrast, are often explained in this way.[16]

The relevance of this to Dretske's account is that the evolution of our visual system may have taken this into account: we might have evolved to detect human-coloured things, *not* just through detecting some objective physical property of human skin, but also through taking into account whether the context makes them likely to be human, and abstracting away from the actual physical property to filter out conditions of illumination, for example. If this were so, then on Dretske's account the property our pinkish-brown colour experiences would have the natural, systemic function of indicating would be something like 'what might be expected to have human-skin col-

[14] I do not mean to claim that this is true only of Caucasians, still less that we are evolutionarily adapted to discriminate that skin colour before all others (!). The filtering out of the greening effect exists for any familiar object placed below the leaves, and white skin is simply one familiar example of the phenomenon.

[15] That is, the colour seen in a region of space is determined not only by the characteristics of the stimuli in that region, but also by those simultaneously present in surrounding regions. These effects can change the region in a direction opposite to the surround (a contrast effect), or in a direction toward that of the surround (an assimilation effect, more traditionally known as the von Bezold spreading effect). Von Bezold effects are especially common in when the coloured region and its surround are quite small, as, perhaps, in pointillist paintings.

[16] For example, C. L. Hardin writes that simultaneous chromatic contrast illusions are a function of 'the opponent systems tending to maximize visual differences while at the same time working toward an overall net chromatic balance' (1988, Plate 2).

our,' rather than any objective external property.[17] That is, qualia would *not* be simply the properties examined by the natural sciences (even if that set includes such contestable items as colour properties), but would be something else altogether, something much more observer-relative and even 'subjective.' This would not only require revision of Dretske's account, but would seriously undermine his central conclusions about the objectivity and third-person accessibility of qualia: in such a case, despite Dretske's best efforts and even if his account were otherwise acceptable, we might *still* never know what it is like to be a bat.

The second difficulty with individuating the objective properties that are supposed to be qualia is the problem of deciding how fine-grained these properties are. Dretske says that someone has a visual experience of a poodle only if their visual system *when it is functioning normally* has the function of demarcating between poodles and everything else — that is, if the representation has the function of indicating poodle-hood. But what is the relevant contrast class here? What *is* the property of looking like a poodle? Dretske admits that very good fakes, such as woolly robot poodles, do have that property, but insists that blurry medium-sized blobs do not — if all you can see are blurry blobs, then nothing looks like a poodle to you (1995, 66 ff.). But what about bichon frisés? These are small woolly dogs that, one is tempted to say, look like poodles. Suppose, because of some slight abnormality in your otherwise normal human perceptual system, you cannot perceptually distinguish between poodles and bichon frisés. Does this mean that you in fact do not have poodle sensations, but instead small-woolly-dog sensations? Presumably, if any normal human's visual system has the function of distinguishing poodles from bichon frisés, then yours does too: that, for Dretske, means you are capable of experiencing both poodle qualia and bichon frisé qualia. Which of the two qualia do you experience on this occasion, then … or is it some *third* quale altogether?[18]

[17] Information about the putative objective external property of pinkish-brownness would be 'lost' before the point of mental representation, analogously with representations of axle-rotation speed in Dretske's more complex speedometer system.

[18] Dretske mentions this very situation on page 69, but does not treat it as a problem.

The problem here is that, within Dretske's picture, there is sometimes *no principled way* of saying just which qualia some reasonably normal, functioning human beings are experiencing. Once again, if a goal of Dretske's theory is to naturalize qualia by making them, at least in principle, objectively and empirically identifiable, his account falls short. Furthermore this is not, it seems to me, merely a minor problem requiring a few further clarifications, or an issue tied closely to the details of Dretske's account in particular; nor is it merely a problem for *naturalised* representationalism, rather than representationalism more broadly construed. The trouble is that qualia are to be identified with the properties represented, rather than properties of the representations, yet — if the worry described above has weight — there sometimes just *are no* determinate properties represented; the *content* of our sense experience is sometimes vague or ambiguous in ways that qualia are not.

### *III. The problem of intentional inexistence*

Dretske holds that qualia are the external, objective properties that sense experiences have the function of representing. However, as Dretske notes, misrepresentation is possible: the world need not always be as it is experienced (1995, 4). Something can look blue but actually be some other colour altogether (or no colour at all). The difficulty is to say, in such cases, what entity *is* blue. On Dretske's picture, the quale *blueness* is just the same objective property as 'ordinary' blueness (whatever exactly that is) … it is a physical property held in common, let's suppose, by cornflowers, a clear sunlit sky and lapis lazuli. Yet it is surely possible to have what would traditionally be called blue qualia — or, more neutrally, a sense impression which involves the visual feel of blue — where there are no blue objects … where there is nothing that has the objective property of blueness. Suppose, for example, that one is gazing fixedly at a large orange screen after just looking at a bright blue light, and that a blue afterimage is swimming across one's gaze.

What can Dretske say about such cases? It would be incoherent to assert simultaneously that qualia are nothing more than properties of external objects (like skies and flowers), *and* that this is a case of misperception where there is no external object which has that property in the visual field, *and* that there is currently an instance of that

property — the quale — present.[19] The only consistent alternative for Dretske is to assert that, during cases of misrepresentation, no instance of the represented property is in fact present — there are no qualia. The peculiar conclusion follows that, while perception involves qualia, misperception does not. (Notice that it is not enough to retort that the *representation* of blueness is present in both cases — which is of course true — since qualia are, for the representationalist, neither representations nor properties of representations but the property *that is represented*, and *this* property *is* absent.)

Worse, this conclusion also leads us into an unpleasant dilemma. Either Dretske must bite the bullet and admit that qualia (in his sense) have nothing to do with what mental life *feels like*, in which case he still owes us an explanation of the subjective, phenomenal qualities of consciousness and has not dealt with the qualia problem at all; or he must assert that, by contrast with veridical perception, there is nothing it feels like to misperceive, since misperception does not involve qualia (or at least that it feels very different from veridical perception, since it involves *different* qualia). In short, either the absence of qualia makes a difference to subjective feel, or it does not, and either way Dretske's theory is unpalatable.

One *prima facie* plausible response to this argument is the following: could not the represented object and its properties be *merely intentional*, on Dretske's account? That is, we perceive a floating blue spot — that spot is the object which is represented in experience — but in reality there is no such entity. So, it seems, we have a perfectly straightforward account of misperception: we simply point out that *representing* a property does not require that the property be *instantiated*. To represent a blue spot in the right way, says Dretske, just is to have a perceptual experience of that after-image; we can do that perfectly well in cases of misperception; so where's the problem? The problem with this response is quite straightforward: we are currently interested in Dretske's account of *qualia* (rather than misperception *per se*), and qualia, for Dretske, are *not* elements of the *representation* of the world; they are qualities of *what is represented*. Thus, in cases where the represented object and its properties fail to actually exist, then

[19] Someone might want to respond that it only *seems* that there is a blue thing, but there really isn't; this misses the point of the objection. There is, *ex hypothesi*, a sense impression involving blueness — a blue quale, one wants to say — and it is *this* property that needs explaining.

neither do qualia. Merely intentional objects do not have real (token) colour properties, for example. Thus, again, it is a consequence of Dretske's account that either qualia have nothing to do with what it is like to undergo perceptual experience, or misperception feels completely different than veridical perception.[20]

### *IV. The conflation of representational vehicles with representational content*

I have already complained that Dretske pays inadequate attention to the status of qualia as *phenomenal* properties; now I wish to examine this issue head-on. It seems abundantly clear that there is no way of introducing a phenomenal element into Dretske's treatment of what he calls *qualia* themselves: qualia, for him, are just regular, everyday properties that happen to be the object of certain sorts of discriminations. Perhaps we would have more luck with Dretske's account of sense experiences? Unfortunately not: the way here is blocked by Dretske's three-way identification of *what it is like to have* certain experiences with the *content* of these representations with that *quale* they have the function of representing. The content and feel of an experience of an electrical field is, for Dretske, identical with the property of being an electric field itself — or rather, being a little more careful about it, a description of the quality and content of an experience is exhausted by statements to the effect that an objective determinable is one way rather than the other. On this account, as Dretske himself points out with satisfaction, one can learn about what it feels like to be a dogfish experiencing lines of electric charge in the surrounding water by discovering more about electric fields — how they work, how they are shaped, and so on (1995, 81 ff.).

[20] Here is one way one might try to escape this objection, suggested by some of Dretske's comments in 1999: perhaps one could argue that although there is no *blue object* in the offing, we nevertheless somehow perceive *uninstantiated blueness*. (After all, if we can perceive it at all, the *universal* blueness, unlike its token instances, is always 'available' to be the object of experience.) This response however strikes me is desperately implausible; what could it possibly *be* to perceive uninstantiated universals, if this amounts neither to merely falsely representing that the universal is instantiated — as I take Dretske's 1995 position to be — nor to tokening another, mental, property that is a mode of presentation of the universal, as a qualia realist might suppose?

As a general claim about representations, even as teleologically understood, this three-way identification does not ring true. Even if one identifies the content of a representation with what it is intended to represent, one cannot identify this construct with the way the representation is configured, or *looks* or *feels*. Consider the case of a water-colour painting of a landscape in the Lake District. Suppose, for the sake of argument, that the content of this painting is precisely what it is intended to represent: say, that particular section of scenery plus a certain melancholy emotional mood. It is far from clear that, even if we were to know everything there is to know about the topography of the Lake District and about melancholy emotions, we would know *what the painting looks like*. This is even more strikingly evident if we suppose the painting to be in some kind of post-modern neo-cubist style, or if we change the example slightly to a descriptive passage with the same content but written in Armenian.

The basic point here is simple enough. Various quite different states may represent some particular content *C*. And exhaustive knowledge about *C*'s intrinsic or (most of its) relational properties would not constitute — or even justify inference to — knowledge about how *C* may be represented.[21]

As a claim about conscious, phenomenal representations, there are strong reasons to think that Dretske's conflation of representation with represented is just as erroneous. Two representations could have the self-same function of representing some object *k* as bearing all and only the members of some particular set F of properties, and yet those two representations could *feel* different. Knowing everything that science has to say about all the members of F does not by itself licence an inference to what the sense-experience of *k* feels like for some other organism.

To illustrate this worry, let us reconsider Dretske's response to the problem of inverted qualia — the objection that, for some theoretical identification of qualia with some other set of properties *P*, the theory fails because qualia can be varied while the members of *P* are held constant. Because of Dretske's idiosyncratic usage of the term *qualia*, for the purposes of this section I shall label this selfsame objection the possibility of 'inverted raw feels,' where 'raw feel' is merely a place-

[21] And possibly vice versa: this is the problem of 'inference' to the external world.

holder term for the way a mental state feels to its experiencer, however this is eventually cashed out.

Now, suppose that two systems have a taste-of-red-(as-opposed-to-white)-wine-detecting-mechanism. As Dretske points out, they could function equally well, but one could experience the taste of every red wine as being like what I experience when I sample a fine Burgundy, and the other could taste like what I experience when I sip a poor Chianti (1995, 71). Dretske reconciles this possibility with the prohibition of inverted raw feels by insisting that only systems that have the function of detecting fine Burgundy can experience that quale/raw feel, and likewise with poor Chianti. Thus, one situation that could give rise to the situation described above would be if two finely discriminating red wine tasting machines both break down and their pointers get stuck at just one position in the red wine 'space,' as it were, with one stuck at fine Burgundy and the unlucky one stuck at bad Chianti. On this account, Dretske will say, there *is* a representational difference between the two systems, despite their functional similarity — it is not the case that raw feels have been altered without appropriate representational changes — and so his identity thesis survives. Furthermore, Dretske will say, it is a perfectly empirical matter to determine just 'what it is like' to be either system.

Nevertheless, despite all this, Dretske has *still* not eliminated the possibility of undetected raw feel change. The only way for him to do this would be for him to insist that two raw feels are *identical* whenever they are the same representation of some objective property by two systems which are identically calibrated. For imagine two wine-tasting systems that are both broken down in exactly the same way: to them both, all red wine is represented by an experience that has the function of indicating fine Burgundy. If it is still conceptually possible that these experiences could *feel* different to the two systems, then the possibility of inverted raw feels remains.

Dretske does hint that he might want to deny the possibility that two representations with identical functions can nevertheless differ (1995, 71, 75), but he never does so explicitly. And it is rather hard, at least on the face of it, to see how he could: surely, wine-tasting machines designed to identify exactly the same set of wine types with exactly the same degree of detail could represent their discriminatory conclusions in different ways: as chemical equations, points on a wine chart, identifying bundles of other properties (such as colour and viscosity) … or as particular taste sensations. And even if two such

systems represent Burgundy by a certain taste sensation, Dretske has still given us no reason to commit ourselves to the claim that it will be the *same* taste sensation: the possibility remains that one could experience the taste of every red wine as being like what I experience when I sample a fine Burgundy, and the other could taste like what I experience when I sip a poor Chianti.

Dretske might respond at this point as follows. On his view we have no first-person access at all to *how* something is represented in our brains — which is entirely a matter of 'brain writing' — but only to *what* is represented. Thus, *for the experiencer*, there *is* no difference between the various modes of representation; if we — or our wine-tasting machine — agree on what is represented, then we will feel exactly the same. That is, the individuation conditions of the experience are exactly the individuation conditions of the content of the representation; the individuation conditions of the content are purely external; awareness of one's own sensations is nothing more than awareness of its content — that is, of those *external* conditions.

What might be taken to motivate and justify such a stance? Primarily, for Dretske, it is the following argument (mentioned in my outline of Dretske's theory):

> 1) qualia are supposed to be the way things seem in the sense modality in question;
> 2) things sometimes *are* the way they seem; therefore
> 3) qualia are exactly the properties the object being perceived *has* when the perception is veridical. (1995, 83–84)

Here, premise 1) identifies qualia with the qualities of phenomenal mental experience — the taste of a strawberry, the experience of the colour of a ripe apple. Premise 2), in effect, says that mental content is sometimes veridical. From these two premises, however, it does not follow that the *qualities* of phenomenal mental experience — its raw feels — always correspond to the way the world would be if the experience were veridical.

Suppose Sally has a finely developed capacity to distinguish between strawberries and non-strawberries by tasting them. And suppose her perceptual system represents the content 'strawberry' by some particular taste-sensation *x*. Thus, for Sally, strawberries taste like *x*, and usually when she thinks she tastes strawberry she really *does* taste strawberry; the content 'strawberry' of her strawberry-related

perceptions is veridical. Nevertheless, it remains coherent and perhaps even plausible to say that the 'actual taste' of strawberries is *not x*: *X* might merely be the *way Sally represents* the content 'strawberry' to herself, a representational system which allows her to be highly accurate in her strawberry discriminations, but one where the quality *x* is not a property of the strawberries themselves. For example, it might be that there *is* no 'actual' strawberry taste that strawberries always have, and that is exactly similar to what we experience, but that our taste sensations are how we *represent* some other complex strawberry property, such as chemical composition, to ourselves.

Since all this is a coherent reading of the argument which makes its premises true and the conclusion false, Dretske's argument will only work if we *antecedently grant him* the assumption that representations themselves are completely 'transparent,' and we have no awareness of the *mode* of representation. Such a position does not follow from his argument, but must be assumed to make the argument work … and so clearly it is not supported by the argument.

## 5. Conclusion

Dretske, then, has reconstructed qualia as non-mental, non-phenomenal, external properties. One of Dretske's main motivations for this redefinition, he makes plain, is the desire to render qualia 'objective' and accessible to study from a third-person perspective; however, with other naturalist options still available, there is little reason to believe that Dretske's position is *entailed* by this desideratum. Furthermore, even on Dretske's account qualia turn out to be possessed of a regrettable slipperiness and observer-relativity.

The central weakness in the representationalist position on qualia, however, is not a problem with its naturalistic adequacy; it is its externalist commitment. This qualia externalism allows the theory to fall victim of the venerable[22] problem of the conspicuous absence of appropriate external qualia-holders during misperception. And it is the main reason that Dretske fails to capture the phenomenality at the heart of the notion of qualia. Dretske may *perhaps* have given a complete account of *mental content*; but he and his colleagues have conspicuously failed to give a complete account of *the contents of the mental*.

[22] Dating back to Ayer (1955), Austin (1964), and beyond.

The collapse of representationalism carries with it certain morals for the qualia naturalizer, and I shall conclude by briefly pointing these out. First, it now seems clear that any theory of qualia must be *internalist*: we must treat qualia as properties of states of the experiencer, rather than of what is experienced — most plausibly, in fact, as properties of some set of the experiencer's brain states. Second, we must take seriously the *phenomenality* of these properties if we are to explain the qualitative nature of our mental representations themselves, rather than merely the qualities of what they represent. So the physicalist must commit herself to something like the following claim: that there exist properties of brains which are instantiated by the neural substrate of the brain's representations of the world, which *feel a certain way* for that brain's owner (they feel red or painful, for instance), and which are *different* properties than those which our representations indicate. The quale of redness is a distinct property from actual redness, for example.

How, then, are these unusual properties to be studied by the third-person methods of science? The only plausible route for the physicalist at this point will be to attempt to *type-identify* qualia with properties which *are* describable from the third person — any other outcome will simply fail to be an account of qualia themselves, but at best will be an account of their supervenience base. Finally, such a set of identities could only be naturalistically *motivated* — rather than merely being brute correlations — if we had some theoretical story to tell which *explains* why these properties, considered from the third person, must feel one way rather than another from the first-person perspective of the brain-owner. This, then, is the project for the qualia naturalizer — this, and only this, is the proper route to the final solution of the qualia problem.[23]

Andrew Bailey
Department of Philosophy
The University of Guelph
Guelph, Ontario N1G 2W1, Canada
abailey@uoguelph.ca

[23] Thanks to Fred Dretske, John A. Baker, Bill Seager, an anonymous referee for this journal, and audiences at the Western Canadian Philosophical Association and the Philosophy Departments of the Universities of Calgary, Alberta and Guelph for their helpful comments on this paper.

## References

Austin, J.L. 1964. *Sense and Sensibilia.* New York: Oxford University Press.

Ayer, A.J. 1955. *The Foundations of Empirical Knowledge.* New York: St. Martin's Press.

Block, Ned. 1980. Troubles with Functionalism. In *Readings in the Philosophy of Psychology, Volume 1,* ed. by Ned Block. Cambridge, MA: Harvard University Press.

Block, Ned. 1990. Inverted Earth. In *Philosophical Perspectives 4*, ed. by James E. Tomberlin. Atascadero, CA: Ridgeview Publishing Company.

Carruthers, Peter. 1996. *Language, Thought and Consciousness.* Cambridge: Cambridge University Press.

Chalmers, David. 1996. *The Conscious Mind*. Oxford: Oxford University Press.

Clark, Austen. 1993. *Sensory Qualities*. Oxford: Clarendon Press.

Dennett, Daniel. 1988. Quining Qualia. In *Consciousness in Contemporary Science*, ed. by A. J. Marcel and E. Bisiach. Oxford: Clarendon Press.

Dretske, Fred. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Dretske, Fred. 1999. The Mind's Awareness of Itself. *Philosophical Studies* 95. 103–124.

Gennaro, Rocco. 1996. *Consciousness and Self-Consciousness.* Amsterdam: John Benjamins.

Goodman, Nelson. 1977. *The Structure of Appearance*. Dordrecht: D. Reidel.

Hardin, C.L. 1988. *Color for Philosophers*. Indianapolis: Hackett Publishing Company.

Harman, Gilbert. 1990. The Intrinsic Quality of Experience. In *Philosophical Perspectives 4,* ed. by James E. Tomberlin. Atascadero, CA: Ridgeview Publishing Company.

Jackson, Frank. 1986. What Mary Didn't Know. *The Journal of Philosophy* 83. 291–295.

Jackson, Frank. 1993. Armchair Metaphysics. In *Philosophy in Mind*, ed. by J. O'Leary-Hawthorne and M. Michael. Dordrecht: Kluwer.

Lewis, C.I. 1929. *Mind and the World Order: Outline of a Theory of Knowledge.* New York: Dover.

Lycan, William. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.

McDowell, John. 1994. The Content of Perceptual Experience. *Philosophical Quarterly* 44. 190–205.

McGinn, Colin. 1991. *The Problem of Consciousness.* Oxford: Basil Blackwell.

Nagel, Thomas. 1974. What Is It Like to Be a Bat? *Philosophical Review* 83. 435–450.

Putnam, Hilary. 1975. The Meaning of 'Meaning.' In *Mind, Language and Reality*. Cambridge: Cambridge University Press.

Rey, Georges. 1992. Sensational Sentences. In *Consciousness,* ed. by Davies and Humphreys. Oxford: Blackwell.

Rosenthal, David. 1990. *A Theory of Consciousness.* ZiF Technical Report. Bielefeld, Germany.

Seager, William. 1999. *Theories of Consciousness*. London: Routledge.

Shoemaker, Sydney. 1982. The Inverted Spectrum. *The Journal of Philosophy* 79. 357–81.

Tye, Michael. 1995. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind.* Cambridge, MA: MIT Press.