DRETSKE'S REPRESENTATIONAL THESIS

A thesis

submitted in partial fulfilment

of the requirements for the Degree

of

Master of Arts in Philosophy

in the

University of Canterbury

by

K. S. Moyst

University of Canterbury

1999

CONTENTS

ABSTRACT

The aim of this thesis is to defend the negative thesis that it is not the case that all mental facts are facts about indicator functions. That which I am arguing against is Fred Dretske's Representational Thesis. The Representational Thesis constitutes an attempt to naturalise features of the mind which have proved elusive to materialist theories of mind. I argue that such features prove equally elusive to the Representational Thesis.

Essentially, my main contention is that there are mental facts which are not facts about indicator functions. To support this position, I appeal (in chapter III) to a version of the knowledge argument against physicalism, which I modify for direct use against the Representational Thesis. This approach has the preliminary result of yielding two types of mental facts that *may not* be facts about indicator functions. Upon closer investigation, one of these two types of facts is found to be more likely to be readily amenable to the Representational Thesis than the other. The other type of facts relate more directly to features of our experiences that the Representational Thesis cannot adequately deal with.

I explore what these features are more fully in chapter IV. I do this by defining them more carefully and then examining certain imaginary situations in which different creatures seem to differ in these features in ways that are not satisfactorily explained by the Representational Thesis. It is because these creatures differ in such features, that we find that there are mental facts about them which are not explicable solely in terms of indicator functions. Consequently, the Representational Thesis is false.

CHAPTER I

INTRODUCTION

Naturalistic theories of mind are usually viewed as having serious difficulties in accounting for certain features of mental states, or events. In particular, accounting for the qualitative features of perceptual and other experiential events (like bodily sensations, moods, and emotions), is generally considered as the Achilles heel of such theories. In *Naturalizing the Mind*, Fred Dretske develops and defends a specific naturalistic thesis about mental events that he believes to be able to adequately deal with such problems. Dretske calls this thesis the Representational Thesis (1995: pxiii). It is in two parts:

(1) All mental facts are representational facts, and
(2) All representational facts are facts about informational functions.

Note that if 1 and 2 are both true, then the following conclusion must also be true:

(3) All mental facts are facts about informational functions.

While I have added 3 myself, it is important to see that the truth of 3 follows logically from the truth of 1 and 2. It is important to see this because it is 3 which I will attempt to refute. If 3 is wrong then either 1 or 2 (at least) is wrong. As *the Representational Thesis (RT) simply is 1 and 2*, then RT would be refuted.

Dretske uses the term "informational functions" in setting out RT, but he generally uses the term "indicator functions" when discussing RT. I will

explain what indicator functions are in chapter II.3. By "representational facts", Dretske means what can be called "intentional facts" (1995: p3). Intentional facts are facts about those external (or internal -in the case of bodily sensations, etc.) objects and properties which a representational act (or event) is *about.* Perhaps, on a different reading of what representational facts are, I could support 1. This is because *I believe that, even if we fully embrace the (very plausible) idea that all of our qualitative states are simply mental events that represent and misrepresent various external and internal states of affairs, there are still some crucial facts about qualitative events that seem to be facts about representational acts themselves, rather than about what they represent.* Lycan is one philosopher who suggests that there may be more to representation than intentionality alone but I do not know if, by this, he means that a full account of representation would include facts about representational acts, or something else (Lycan: 1996, p11). In any case, I could only support 1 if it allowed for certain facts about representational acts, rather than just facts about what such acts are about (i.e. that which the representational event represents to be the case), to be considered as representational facts. RT does not allow for this possibility.

Because Dretske's representational facts are intentional facts and intentionality may not be sufficient for representation, I will reformulate RT in terms of "intentional facts" rather "representational facts" in order to remove this possible source of ambiguity between representation and intentionality. As I will later explain (chapter II.4), however, Dretske's representational facts are not fully coextensive with the set of intentional facts. While representational facts just are intentional facts, the set of intentional facts is larger than the set of representational facts. As such, I will attach an asterisk to "intentional facts".

(thus: intentional facts*), so that 2* does not immediately come out as false. We now have the following (RT*):

(1*) All mental facts are intentional facts*, and
(2*) All intentional facts* are facts about informational functions.

I will drop the use of the asterisk from now on, as it should be clear that the intentional facts that I will speak of are Dretske's representational facts. In fact, I will largely use Dretske's terminology anyway.

Note that 3 (unmodified) follows from 1* and 2*, just as it does from 1 and 2. Because the representational facts of RT simply are intentional facts, RT and RT* are equivalent. I do not accept either 1* or 2*; I think that they are both wrong. Against 1*, I believe that there are certain crucial facts about qualitative mental events themselves which cannot just be construed of as intentional facts, even when experiences are fully conceived of as representational events. Against 2*, I believe that intentionality (even in the slightly lessened role of representational facts) cannot be adequately treated in terms of indicator functions. I will not, however, be concerned with arguing against 2*. In fact, even if we assume that 2* (and therefore 2) is correct, my arguments against 3 will still work. This does not mean that my arguments against 3 are actually arguments against 1* (and therefore 1), however. If I succeed in refuting 3, then I will have succeeded in refuting RT as 3 is a consequence of RT. The nature of my arguments against 3 are such that I believe that they *could be reformulated* directly against 1 (certainly, at least, if intentional facts are taken as being fully external to mental events), but as they stand they are arguments against 3.

I have divided this paper into three main parts (chapters II-IV). The first of these parts (chapter II) will be primarily concerned with the exposition of Dretske's Representational Thesis. In chapter III, I will develop my argument against 3 above, which I will begin to formulate in chapter II.8. Chapter IV will deal explicitly with the issue of qualia which will be seen to be at the heart of my argument against 3. In order to refute 3, I will argue that there are facts about our experiences (=qualitative mental states or events) that are not facts about informational functions. I will do this by appealing to our intuitions about various real and imaginary situations in which there are facts about experiences that do not appear to be captured by facts about informational functions.

The types of experiences that I will be concerned with in this paper are perceptual experiences. This is because Dretske focuses most of his discussion on perceptual experiences. He does this, understandably enough, for the simple fact that perceptual experiences have a commonly accepted representational element to them. Therefore, if he can show that these types of events are entirely representational (in the above mentioned sense), he can then move on to deal with the less obviously representational qualitative events in the same way. In fact, he has already started on this project (Dretske: forthcoming, pp13-17). Furthermore, most of my examples will relate to experiences of colours. Secondary qualities such as colours are far more problematic than are primary qualities for representationalist (=intentionalist) theories such as Dretske's.

CHAPTER II

THE REPRESENTATIONAL THESIS

In this chapter, I will be largely concerned with discussing the details of Dretske's Representational Thesis. I will, however, begin with an important note on mental acts and mental objects. By the end of chapter II, I will have introduced what I believe to be the central problem with RT as an account of the phenomenal mind.

1. MENTAL ACTS AND MENTAL OBJECTS

Just as Dretske does, I will assume that there are no mental objects in the sense of phenomenal objects such as sense-data. Our perceptual experiences are not experiences of mental objects. If it were the case that we experienced mental objects, then RT and materialism in general would be falsified by this fact alone. While some philosophers still do argue that we experience phenomenal objects and that materialism is, therefore, false, most do not. Instead, the general consensus is that *there are no mental objects -only mental acts* that take physical objects as their objects. This is Dretske's position as well as my own. I will not argue for this position; I will just assume it.

In line with this stand, I will tend to refer to experiences as events, or representational acts, and representings, rather than as states and representations. These latter terms are more easily read objectually than the former ones even though they are coextensive. States of a representational system

just are events; and the word "representation" simply nominalises the act of representing (Browne: 1996, p5). While Dretske tends to use the terminology of states and representations more often than I do, this should not be taken as signalling any fundamental difference in our ontologies of the mental. Also, the fact that I will sometimes talk of states and sometimes talk of events should not be taken as a sign of inconsistency on my behalf.

2. REPRESENTATION AND REPRESENTATIONAL SYSTEMS

The notions of representation and representational systems are fundamental to Dretske's RT. Representational systems include people and other animals and artefacts like computers and speedometers and thermometers. In fact, some representational systems are made up of a whole host of such systems. Perceiving and thinking animals like people for instance have a number of sensory and conceptual representational systems like vision, hearing, and speech-producing and comprehending systems. *Representational systems are systems that have the function of providing specific types of information.* The requirement that representational systems have the *function* of providing specific types of information is what supplies the essential normative element of representation in Dretske's representational account of the mind. This same element can be captured by the words "*supposed to*" or "*meant to*" or "*intended to*". Thus, if a representational system has the function of providing a certain type of information, then it is supposed to, or meant to, or intended to, provide this type of information.

This normative element is required to account for misrepresentation and is, therefore, required for a full account of the intentional features of representation. Something s which *cannot* misrepresent cannot be properly said to represent at all, because if s is supposed to represent something k then s is supposed to say something *about* k (e.g. that k has features F and G, or that k is present), and if this is the case then s must (at least in principle) be able to get this wrong. Furthermore, because intentionality is a necessary (even if not a sufficient) condition for representation, the possibility of error is necessary for representation proper as opposed to what I will term "quasi-intentional" phenomena. Quasi-intentional phenomena are phenomena which are not fully intentional and, therefore, not fully representational. Obviously, if there is more to representation than intentionality alone, then even fully intentional phenomena are not fully representational phenomena, but they are closer to being fully representational than quasi-intentional phenomena are -intentionality being necessary for representation. I will soon (in the next section) show how Dretske frames this distinction that I have made between quasi-intentional and fully intentional phenomena in terms of his distinction between indicators and indicator functions. First, however, I will demonstrate how the distinction is made by way of an example.

If I do a sketch which is supposed to constitute a plan of the ground floor of the University of Canterbury Library, but my sketch gets the whole layout of this floor wrong, then this means that my sketch misrepresents the actual layout. It does not mean that my sketch actually constitutes a plan of some other floor in another building, even if it can be used much better in this way. Likewise, if my sketch s was not supposed to constitute a plan of anything

(it was just a doodle consisting of an arbitrary arrangement of lines), and yet it was found to be able to accurately represent the ground floor of the library, then this does not mean that I produced a representation (i.e. a plan) of this floor or of anything at all. If s is *not meant to* represent something else L (e.g. the ground floor of the library), then it simply *does not represent or misrepresent* that thing or any other thing that it is not supposed to represent. Of course, if a person P discovers that s can be used to represent L (i.e. that s is *able to* be used to represent L), then P might use it as such; P might assign this use (or function) to s. In such a case, s can be said to misrepresent some feature of L. For example, a certain sized gap in a line, that was taken to represent the position of a door (because all other such gaps are found to correspond to door locations), might be said by P to misrepresent the actual state of affairs when P finds that there is no door in the corresponding part of the library. In Dretske's terminology, P assigned s the function of representing L, and s therefore misrepresented a feature of L.

If P had found s on a table in the library and had merely investigated whether s was a representation of L (i.e. because P had a hunch that it might be), then when P discovered that there was no door where P had thought that a door may have been represented by s as being, s *could not* be said to misrepresent this state of affairs. It could not misrepresent this because it was *never intended to* represent L by its maker (remember that it was just a doodle), and because it was not being used by P (and, therefore, it was not intended by P) to represent L. Conversely, s could not represent L if it could not (in principle) misrepresent L (or a feature of L): something s that cannot possibly misrepresent *cannot*

represent at all. S is merely *quasi*-intentional not *fully* intentional; it can only represent (and, therefore, misrepresent) if it is given this use.

As was seen above, s became fully intentional when it was assigned a function by P. Obviously, if all quasi-intentional phenomena became fully intentional only after they had been assigned functions by fully intentional representational systems like P, then this would initiate a regress that could only be satisfactorily halted by positing an agent (e.g. God) whose own intentionality did not depend upon another pre-existing fully intentional source. As such, Dretske does not give an account of how functions are determined that relies only upon intentional agents assigning functions to quasi-intentional phenomena. Instead, he appeals to nonintentional evolutionary forces that work in such ways as to bring functions into a world in which they were previously absent.

Later (chapter II.6), I will explain how Dretske believes that functions are determined. Next, however, I will look at an extremely important distinction for Dretske's Representational Thesis (RT).

3. INDICATORS AND INDICATOR FUNCTIONS

The distinction between what information a representational system *is able to* receive (a quasi-intentional fact) and what information it *is supposed to* receive (a fully intentional fact) can also be put in terms of Dretske's distinction between indicators and indicator functions (= the informational functions of part 2 of RT). An indicator is something that *is able to* carry information about something else; while an indicator function is something that an indicator can have that makes it the case that it *is supposed to* be about something else.

Something can be an indicator of something else without having the function of carrying information about that which it indicates. Likewise, something can have the function of indicating something that it does not actually indicate. In the former case, we do not have a case of representation proper, while in the latter case we have an instance of misrepresentation. It is important to Dretske's account of representation in terms of indicator (information-carrying) functions to note that there can be no misindication, only misrepresentation. An A token can be an indicator of B iff (if and only if) it *does* indicate B. If an A token does not indicate B on a given occasion, even though all other known A-type tokens have indicated B, then this token is simply not an indicator of B. It can, however, have the function of indicating B even if it does not indicate B -in which case A misrepresents, but does not misindicate, B.

Examples of indicators are tree rings, tracks in snow, and dark clouds. These (respectively) carry information about how old a tree is; the type of creature or artefact that was at a certain location within a certain period of time and the direction it was moving in; and the fact that it is going to rain. However, if (for example) a tree's rings suggest that it is 12 years old when it is actually 14 years old, then these rings simply do not (mis)indicate that the tree was 12 years old. If a person P who was counting the tree's rings assigned these particular rings the function of indicating the age of the tree, then we can say that the rings did indeed misrepresent the age of the tree; but this is only relative to the belief of P that age is what was indicated and, therefore, the indicator function that P assigned to the rings. In effect, P's belief about what these particular rings indicate misrepresented the age of the tree; the rings themselves did not. The same sort of story can be told for tracks in snow, dark clouds and any other

indicators. Token indicators never misindicate but they can misrepresent in the sense that they can be assigned the function (usually on the basis that they belong to an indicator-type which can indicate B) of indicating that which they do not actually indicate (B). In such cases they are not, strictly speaking, B indicators (because they do not indicate B) but they do have the function of being indicators of B. Token indicators can also be assigned the function of indicating that which tokens of their type cannot ever indicate at all (e.g. the presence of black clouds could be assigned the function of indicating that trees in the area are now indestructible), but this could only ever result in the misrepresentation of this state of affairs. It would never result in veridical representation and, so, there would be little use in assigning such functions.

4. REPRESENTATIONAL FACTS AND FACTS ABOUT REPRESENTATIONAL SYSTEMS

Dretske differentiates between representational facts and facts about representational systems (Dretske: 1995, p3). Facts about indicators can be facts about the representational systems that employ them, but facts about indicator-functions are representational facts. As I said in the introduction, representational facts are intentional facts. They relate to what a representational act is about. The normative element of representation, which Dretske seeks to capture with his notion of functions is, as I have shown (in section 2), important in determining what a representational act is about. Such acts are about that which they are supposed to be about, they are not just about that which they can be (are able to be) about. The fact that a representational act

can misrepresent that which it is about *is* an intentional fact; that object which is misrepresented is what the act says something wrong *about*. When there is no actual object that is misrepresented, the act is still *about* something -though in this case, the "something" stands in for a failed indexical (Dretske: 1995, p27). Strictly speaking, however, RT includes the properties that a representational act represents as part of what the act is *about*. As such, when a representational system is representing something that is not there the "something" is not (solely) a failed indexical, it also designates the properties that the act misrepresents a (nonexistent) object as having (Dretske: 1995, n20, p172).

In distinguishing between representational facts and facts about representational systems, Dretske writes that "The fact that the speedometer has a speed indicating function, and the fact that pointing at "37" means 37mph are representational facts about the instrument" (1995: p2). They are representational facts because they are facts about the indicator function of the instrument. They are facts about what the instrument was designed to do and, therefore, what it is supposed to do. To be more precise, they are facts about the instrument that relate to what states of the instrument are *about*. On the other hand, facts about the representational system, such as its colour, shape, length of its cables, and its material constitution are not representational facts. They do not tell us about the information that they are supposed to supply, or even whether they are *supposed to* supply information at all (1995: p3). Some of these facts can help us to determine what information the system is *able to* receive, but unless we know what information the system is *supposed to* receive we do not, according to Dretske, know what is being represented.

There is one important intentional fact, which Dretske does not take to be a representational fact -the fact that my experience is an experience of *this* object (e.g. *this* identical twin) rather than another object of the same type. It is this sort of fact that Wittgenstein has pointed to as containing "the whole problem of representation" (Wittgenstein: 1994, p69). Put in terms of thoughts rather than experiences, for example, two people can both have the veridical thought that MacBeth killed Duncan and became King of Scotland. However, one of these people could be thinking of the historical MacBeth who killed Duncan in battle and gained the throne of Scotland in line with the then prevailing Gaelic laws of succession, while the other person's thought is about Shakespeare's Macbeth (the small b being an anglicisation) who killed Duncan while asleep in bed and usurped the throne by breaking what amount to English laws of succession. In both cases, the simple thought that MacBeth killed Duncan and became King of Scotland is the same, but they pick out different people -a real person and a fictional character, respectively. How exactly this is done is the problem that Wittgenstein's comment alludes to.

Of course, my MacBeth example assumes that Shakespeare's Macbeth *is* a fictional character and not the same person as the historical character who has simply been slandered and vilified. In any case, most people's MacBeth thoughts are about a fictional character of whom it is true to say such things as that he killed the sleeping Duncan and usurped the throne. As such, I think we can still legitimately make a distinction between two MacBeths, both of whom killed someone called Duncan and became King of a country called Scotland. Furthermore, my saying that the thought about Shakespeare's Macbeth picks out a different *person* from the historical MacBeth does presuppose a realist and

anti-reductionist semantics of fiction (Proudfoot: 1998, p1). While such a semantics is very problematic, I will just assume its truth for the purpose of this example; nothing else will depend upon this.

The fact that Dretske does not include this sort of intentional fact (e.g. the fact that I am looking at, and therefore, having a visual experience of identical twin A rather than twin B) among his representational facts means, as I noted in the introduction, that Dretske's representational facts are members of a set of facts that is smaller than the full set of intentional facts. The former is a subset of the latter.

According to Dretske, the fact that a token state of a system represents a particular object k is not a (pure) representational fact about the system (Dretske: 1995, pp24-26). It is a hybrid representational fact; it is only partly a representational fact. What object a token state of a system represents is not determined by which of its properties are represented but by "a certain external *causal or contextual relation*" that he designates as *C*. The veridicality of such representational states is partly determined by C. However, this is not a fact about what a representational system S has the function of indicating. If S has the function of indicating the instantiation of property F, then saying that S is representing object k is the same as saying that S represents (or misrepresents) the F of k. "That S represents k, therefore, implies a representational fact -that, for some F, S represents the F of k. But it also implies something that is not a representational fact -viz., that k stands in relation C to S" (1995: p26).

5. REPRESENTATIONAL CONTENTS AND REPRESENTATIONAL VEHICLES

Dretske intends the distinction between representational facts and facts about representational systems to apply to differences between facts about the mind and facts about the brain (1995: p3). The analogy of the differences between representational facts and facts about representational systems, and the differences between facts about the mind and facts about the brain, is developed in terms of the distinction between representational contents and representational vehicles. If we use Dretske's example of a story, we can make out the distinction between content and vehicle in terms of the story that the words tell (the content) and the words that are used to tell the story (the vehicle) (1995: p34). Dretske makes the point that story vehicles are in books, but what happens in the story (the content) is not. The people and events that are represented in the story are not to be found between the covers of the book. To make the analogy with the mind and brain, Dretske says that thoughts and experiences are in the head in the same way that stories are in books -as vehicles (physical states of the brain that have a representational content) (1995: p35). But what makes a thought or experience the thought or experience that it is what it is a thought or experience about (the content), and this is, according to Dretske, the external property or properties that a token brain state has the function of indicating (1995: p37).

6. MENTAL ACTS OF PHENOMENAL REPRESENTATION

Experiences, like thoughts, are mental acts. Dretske employs the terms "phenomenal representations" and "conceptual representations" to distinguish

between these types of mental acts. This terminology is used to express the sharp distinction that, Dretske notes, is acknowledged by most philosophers and cognitive psychologists (1995: n8, p170). In line with what I have said earlier (section 1), I will use terms like "representings" and "acts of..." rather than just "representations".

To have an experience of something is to be in a phenomenal state, which is an event that is characterised by an act of phenomenal representation. To have a thought (i.e. to form a belief or judgment) about something is to be in a conceptual state, which is an event that is characterised by an act of conceptual representation. The fact that we are often (or even largely) in both sorts of states at the same time does not undermine this distinction. Having an experience of something does not require any sort of understanding of that which is being experienced, while having a thought about something does. Pre-linguistic infants and non-linguistic animals can have visual experiences of cars and aural experiences of the sound of a car's horn without possessing the concepts of CAR and CAR HORN which are required for beliefs about these things. Just as adults can have experiences of objects that they know nothing about. But one cannot have beliefs about something without having some sort of understanding, however partial or even mistaken. Such an understanding involves possessing the necessary concepts for the particular belief (Dretske 1995:pp8-12). As such, phenomenal representings (or acts of phenomenal representation) can also be classed as *nonconceptual* representings.

Dretske identifies phenomenal events (though he uses the word "states" rather than "events") with *mental events that have systemic representational properties*, which are properties that are derived from the system that produces

them (1995:p15). He identifies conceptual events with mental events that have acquired representational properties, which are properties that are acquired via a process of learning. The representational properties in question are indicator functions. This means that *token phenomenal events represent (have the function of indicating) those properties that the representational systems (e.g. visual and auditory systems) which produce them have the function of indicating.* The indicator function/s of a token phenomenal event is/are given by those properties whose instantiation, within a certain range of distances from the appropriate representational system, it is the function of the system to indicate. What it means to say that a token event has the function of indicating a property is that that event is supposed to indicate an instantiation of that property. This property constitutes the representational content of the event. Consequently, if the event indicates that the property is instantiated when this is not actually the case, misrepresentation occurs.

According to Dretske, the functions of mental events are acquired naturally (1995: p7). This makes them natural representations. Natural functions and natural representations are contrasted with conventional functions and conventional representations. Conventional functions and conventional representations are determined by the purposes and intentions of designers, builders, and users. According to Dretske *biological representational systems such as the human senses, "have information-providing functions, biological functions, they derive from their evolutionary history"* (1995: p7). As a result, the senses produce representations of those properties that their evolutionary history has given them the function of informing about. Dretske does note, however, that his account of the senses as having natural functions is not dependent upon any

particular theory of natural functions (1995: n4, pp169-170). Even if evolutionary history is not the only source or the major source of natural functions, his account of the senses as representing properties that it is their function to indicate should not be affected as long as there are natural functions. That it is not universally agreed that there are natural functions is noted by Dretske, though he offers no argument in support of their existence (1995: p7).

In making the distinction between phenomenal and conceptual representings, Dretske states that both of these types of mental acts are representings of purported fact (1995: p9). How phenomenal representings are representings of purported fact is, I believe, best understood in terms of Christopher Peacocke's discussion of the nonconceptual content of sense experience (Peacocke: 1992, p61). On his account, this sort of content (or, at least, what he calls "scenario content" which is one of the two types of nonconceptual perceptual content that he distinguishes between) can be individuated by specifying the ways in which the space around the perceiver could be filled out, while being consistent with the correctness of the representational content. Consequently, we could say that the facts that phenomenal representings purport to be true are facts about the instantiation of things in the world around the perceiver. Acts of phenomenal representation misrepresent when they represent the space around the perceiver as being filled out by things that are not actually as they are represented to be (e.g. things that are actually square-shaped that are represented as circular). Of course, the individual experiencing something need not possess the relevant concepts or be able to specify the way in which the space around him or her is filled out. Acts of phenomenal representation can take place without conceptual representation.

Phenomenal representings (experiences) service the construction of conceptual representings (e.g. perceptual beliefs), and are able to be calibrated by learning in order to be more effective in servicing an organism's needs and desires (Dretske: 1995, p19). This latter point can be made clearer by an earlier distinction that Dretske made between analog and digital information (1981: pp136-153). On this distinction, acts of phenomenal representation carry much information; not all of which is likely to be used by the organism. They are informationally rich, or analog encoded, events. When a specific piece of information is extracted from such a token event via an act of conceptual representation (e.g. the recognition of a grape taste in a sip of wine), information contained within the analog encoded event is digitally encoded. Via a process of learning, the analog information in an organism's phenomenal representings can be calibrated to yield more and more specific, digitally encoded information. From only recognising the taste of grape, an individual can become a trained wine-taster who can recognise a great deal more from a mere sip of this beverage. All the relevant information that makes these discriminations possible is to be found in analog form in the phenomenal representings that are produced by the act of tasting wine even before the individual is able to make such discriminations. According to Dretske, the function of experiences qua experiences (rather than experiences qua specific phenomenal events with specific indicator functions) is to be found in their supplying a cognitive system with information for calibration and use in controlling and regulating behaviour (1995: p19). This is the adaptive role of experience that explains why creatures with the capacity to have experiences have been selected.

7. DRETSKE'S SENSE AND REFERENCE DISTINCTION

The distinction between sense and reference is a semantic distinction which originated in Frege's now classic essay *On Sense and Reference* (Frege: 1993). Frege introduced this distinction by discussing what is today termed the perspectival character of belief; a feature that gives rise to what has become known as the problem of referential opacity. The perspectival character of belief means that something is believed about an object under a certain description, where this description gives a property that the object has. Sometimes names can stand in for descriptions; they can be thought of as shorthand descriptions. For instance, the meaning of "Morning Star" can be given by a description such as "that object which is the brightest object in the morning sky"; and the meaning of "Evening Star" can be given by a description such as "that object which is the brightest object in the evening sky. The perspectival character of belief gives rise to the problem of referential opacity in that a person can believe something about an object under one description while not believing the same thing about it under another equally applicable description, simply because the person does not know that they both designate the same object. "Morning Star" and "Evening Star" are shorthand descriptions of different (relational) properties of the same object. Because of this, they can be said to provide different *senses* of the same *reference*.

Frege discussed this distinction in terms of statements of identity. While "a=a" and "the Morning Star is identical to the Morning Star" are trivial statements of identity, "a=b" and "the Morning Star is identical to the Evening

Star" are non-trivial statements of identity. The discovery that the Morning Star and the Evening Star are one and the same entity (Venus), for example, decreased the known number of heavenly bodies by one.

As a semantic distinction, the sense and reference distinction is one which concerns conceptual rather than phenomenal representings. Dretske, however, has applied this distinction to phenomenal representings in a manner which is analogous to Frege's original formulation of the distinction in terms of language. Dretske's distinction between sense (or aspectual shape) and reference is of fundamental importance to RT, as it is his notion of sense that essentially provides us with an account of the type of mental facts that he believes are available to a representational system. According to Dretske, the difference between the sense and reference of an act of sensory representation is the same as that between a represented property (a universal) and a represented object (a particular) (1995: p23). The sense of a representational act is given by the objective properties that it has the function of indicating, while the reference (when there is one) is the external object whose properties are being represented.

In cases of hallucination, the representational act has a sense, but it does not have a reference. For example, if I have a visual hallucination of Hannibal atop an elephant, then my sensory state can be characterised by its representational sense as representing (roughly) those properties which a veridical experience of Hannibal atop an elephant would have represented. I do not know if Hannibal actually ever rode an elephant himself, but this does not affect my example. A veridical experience of a model, or perhaps even a picture, of Hannibal atop an elephant would also have the same sense. My hallucinatory experience would, however, be without a reference, as the referents of sensory

experience are only those objects which stand in an immediate causal/contextual relation C to a perceiver at the time that a given representational act takes place. Of course, we could call whatever object I was looking at at the time of hallucination (say, a parked car), the reference of my (nonveridical) experience, but it is not necessary that I was looking at any object at all. I may have been staring into empty space, or I may even have been dreaming. In both of these latter cases, I will still be in a mental state that can be characterised in terms of its representational sense. It is a state that is representing the instantiation of various specific properties in a situation where I do not actually stand in relation C to any object at all.

This phenomenal representing is analogous to the thought (a conceptual representing) that griffins are raiding sheep from Canterbury farms. As the griffin is a chimera, that part of the thought which relates to griffins has a sense, but no reference. Note, however, that while the griffin constitutes an intentional object of my thought, Hannibal atop an elephant does not constitute the intentional object of my phenomenal representing. Rather, on Dretske's account, it is those properties that my experience represented as being instantiated that my experience was (strictly speaking) about. However, Hannibal atop an elephant may have been the intentional object of my associated (if any) conceptual representing. For example, if I believed that I was seeing (or even if I believed that I was hallucinating) Hannibal atop an elephant or something that looked like Hannibal atop an elephant, then Hannibal atop an elephant would have been the intentional object of my thought. This would also be the case if my belief was dispositional rather than occurrent, i.e. if I would have been disposed to judge that the experience was of Hannibal atop an elephant if asked by

someone (including myself). Even if I did not have any such thought (occurrent or dispositional), for instance if I had never heard of Hannibal or even of the things that Hannibal had done, I could still hallucinate the instantiation of that same selection of properties that I described above as (roughly) those properties which a veridical experience of Hannibal atop an elephant would have represented.

The reason why a token experience can represent an object (a particular) by representing the token properties of the object is that experiences represent by representing the instantiation of types of properties. A experience represents an object by representing its token properties as instantiations of certain property types. Experiences represent general features or properties. It is this fact about the generality of sensory representation that leads Tye to speak of the content of experience as being abstract (Tye: 1995, p138). It is also what Peacocke has in mind in his discussion that I alluded to in section 6 (Peacocke: 1992, p61).

Because experiences represent the token properties of particulars, *which particular* is being represented does not affect the experience, as long as the particulars share *the same properties and values of properties* that a representational system has the function of representing. For example, if the same shade of the objective colour property blue (say blue22) is instantiated in two different objects, a normal human's (veridical) colour experiences of either of these two objects will be the same. Likewise, even if there is no object that is being represented, an experience can be the same as one that arises in relation with an actual particular. In the case of sensory representation sameness of sense does not entail sameness of reference. In fact, this is also the case (to a large

extent a least) in the case of linguistic representation (cf Kripke: 1980 and Putnam: 1993). As with thoughts and utterances involving obviously indexical words like "me" and "here" and "now", the referents of acts of sensory representation are indexed to a person (in this case, the perceiver) and the place and time in which the event (the act of phenomenal representation) has occurred. It is because of this that Dretske, as I mentioned earlier, speaks of reference as being determined by a certain *causal or contextual relation C* (see section 4). For the same reason, Peacocke indexes the referents of experience by fixing specific origins and axes in relation to the body of a perceiver (Peacocke: 1992, pp61-62).

8. THE PHENOMENOLOGY OF EXPERIENCE

Dretske's notion of the sense of an act of phenomenal representation can be, thus, used to yield a naturalistic account of similarities between experiences of different particulars. Separate experiences of identical twins, for example, are experientially identical to human perceivers because such perceivers represent the token properties and values of properties of each particular twin as being of the same property type and value type.

Furthermore, Dretske's notion can also be used to yield a naturalistic account of experiential differences between humans and alien life-forms which are either occupying the same environment or which could, at least in principle, occupy the same environment as us. It certainly seems plausible, for instance, that humans are not the only creatures to experience the world (in the sense of being aware of objects via acts of nonconceptual sensory representation). If this is the case, then it is likely (given what is known about the senses of various creatures) that there are alien life-forms (like Nagel's bats) which do not have the same sorts of experiences as us, and that these differences cannot be accounted for solely in terms of the nature of the environment that they inhabit. If we occupied their environment and imitated their lifestyle, this would not give us much insight into their sensory experiences. Evolutionary processes have structured our sensory organs and our central nervous systems far too differently.

One (non-Dretskean) explanation of these experiential differences could be that the alien creatures experience different types of phenomenal objects than humans do. This explanation could not be easily accommodated by a naturalistic

approach to the mind. On Dretske's theory, however, it can be said that these creatures are simply experiencing different properties or a different range of values of the same properties as us. In this sense, they are experiencing the same external objects of the environment in different ways. For example, we can say that the experiences of bats and dogfish (Dretske's preferred example) differ from ours simply because their sensory states represent properties and values of properties that our's do not (Dretske: 1995, pp82, 84-88). Our experiences have a different representational sense even though the referents may be the same. The referents of our experiences are the same external, physical objects, not different internal mental objects. The difference between our mental states, then, is somewhat analogous to the difference between a person who has a thought about the Morning Star and a person who has a thought about the Evening Star. Both thoughts have a different sense, but the same reference. Likewise, the difference between them is given in terms of different representations of the same external object, not in terms of their thoughts possessing different mental objects.

While what I have said about Dretske's account of the sense and reference of acts of phenomenal representation may give the impression that I believe that he has made much progress in the project of naturalising qualitative mental states (experiences), I do not believe this to be the case. In fact, I believe that his approach is inadequate to achieve this task. A representational act may represent an object only by representing (some of) its properties, but the act itself can be characterised not only by the properties that it represents but also by the way in which these properties are represented. ***A full account of qualitative mental states would, I believe, have to tell us something about the way in which***

those properties which are phenomenally represented are, themselves, phenomenologically represented.

I am using the expression "the way in which those properties which are phenomenally represented are phenomenologically represented" to capture an important feature of the phenomenology of experience which I do not believe to be captured by Dretske's notion of phenomenal representings (or "phenomenal representations", as he calls them). Later (in chapter IV), I will identify this feature of experience with the qualia of an experience. More precisely, I will identify it as *a defining feature of the qualia* of an experience. In the meantime, it is most easily thought of as that aspect of our experiences which is inverted in traditional inverted spectra thought experiments (e.g. the *phenomenal*-red and *phenomenal*-green of colour experiences). It differs from Dretske's notion of phenomenal representation in that I believe that a token property of an object can be (veridically) phenomenally represented by two creatures with inverted spectra, such that both creatures phenomenologically represent the *same* token property but in a *different* way (e.g. phenomenal-red and phenomenal-green, respectively). This will be discussed more fully in chapter IV.

If it were the case that every trope of a particular value of a particular property could only be phenomenally represented in one specific way, then Dretske's theory might seem to capture the phenomenology of our experiences admirably well. The phenomenology of a sensory event would, as Dretske says, be determinable solely by specifying those properties that the sensory event was representing an object as having. The specific combination of particular properties and their values would, by specifying the representational content, specify the phenomenology. For example, if a particular value (shade) of the

objective colour property blue (say blue22) was being represented, then specifying that blue22 was being represented by a sensory event would also allow the specification of how blue22 was being phenomenologically represented as "that particular way in which blue22 is *always* phenomenologically represented". And if, as Dretske argues, representation is a matter of indicator functions, then determining the indicator functions of a sensory event would be the same as determining which of an external object's properties were being represented by that event, and (in the manner just mentioned above) how they were being phenomenologically represented. This would, indeed, seem to make phenomenology as objectively determinable as is the function of a biological organ such as the heart.

Let us assume that *all values of a given property* could be objectively determined independently of the particular representational system and sensory modality that is representing them. Note, however, that even if it was the case that every particular value of a particular property which was phenomenally represented could only be phenomenologically represented in one specific way (the same way that every sensory modality of every representational system represented it), we would still have to determine how this property was phenomenologically represented in order to *fully specify the phenomenology* of such sensory events. The phenomenology of a representational act is not given *solely* by the properties of the objects which are represented, but also by specific features of the act which constitute the representing of these properties. The fact that certain properties are being phenomenally represented is not the same as the fact that these properties are being phenomenologically represented in a certain way. *How these properties are phenomenologically represented is critically*

important, even if they could only be represented in one way. **Dretske's theory may eliminate the need for phenomenal objects by replacing them with physical objects but it does not eliminate phenomenal representings of physical objects. In fact, it posits these. And how something is phenomenologically represented is a feature of such phenomenal acts, not of the physical objects that are represented. The *properties of those objects* which are represented by representational systems are physical properties, they are not representational properties (in a non-Dretskean sense of being *properties of representational acts*). Likewise, *the phenomenal representing of properties* (a more easily determinable property of representational acts) is conceptually distinct from *the way that these properties are phenomenologically represented.* In other words, the fact that certain properties are represented by an experiential event is not the same as the fact that they are represented by that same token event in one certain way rather than another.**

9. UNIVERSALS AND TROPES

Because Dretske frames his notion of the sense and reference of an experience in terms of the distinction between universals and particulars, I do not want my claims to be misinterpreted as claims about universals (properties). I do not intend to claim that the universal *red*, for example, is sometimes phenomenologically represented as red and sometimes phenomenologically represented as green depending upon the subject in whose experience it is instantiated. The universal *red* is an intentional property with no colour of its own, whether in this world or in some Platonic realm of Forms. However, if we

accept Dretske's realism about properties and, therefore, understand colour (for instance) as being an objective property of external objects or their surfaces, then we can talk about a particular which instantiates a token of a particular colour property like red. The object (qua particular) possesses as a part of itself a determinable token of an objective property type which other particulars can also share (qua an universal), but not qua the particular instance (token) that its instantiation in this object constitutes. Such a token of a universal is known as a *trope*. When we look at an object it is the object's tropes that are represented by our experiences, even if these are represented in virtue of the fact that they are tokens of a certain type and are phenomenologically represented in the same way that other such tokens are phenomenologically represented.

Where I differ with Dretske is in the claim (that I make) that when an individual is looking at an object whose tropes (e.g. colour tropes) are being represented, the individual's token representings of the object's tropes are phenomenological representings which *could*, conceivably, have been of a *different* nature than they actually are. If two individual's with visual systems that have the same indicator functions are both looking at this same object, then they could both still be veridically representing the same tropes, but their respective token representings of the object's tropes are phenomenological representings which could, conceivably, be of a different nature than each other's. They could have inverted spectra. This means that their respective experiences of the world are *not* phenomenologically identical, even though their indicator functions are the *same* and they are experiencing the very *same* external object. This will be explained more fully when I come to discuss qualia in chapter IV.

10. THE DIAPHANOUS, PRESENTATIONAL CHARACTER OF EXPERIENCE

Acts of phenomenal representation characterise personal level mental events. While being representational in nature they are also, phenomenologically, presentational. This means that they are not experienced as being representational in the way that a plan of a building or a map of an area are experienced as essentially representational. In fact, experiences are not experienced as representational at all. When you look at the plan or the map you *see certain symbols* (words and pictures) and you *infer* that there is a door on the right-hand side of the front of the building, or that there is a street with a certain name running parallel to a part of a river for about 100 meters, etc. When you look directly at the building or the street, on the other hand, you a *simply see door or a nearby river.* The building may actually be a fake or you may have actually hallucinated the river, but your experience just presents these to you all the same; no inference is required. *Seeing* the building *as* a building, unlike just *seeing the building* (simple seeing), requires conceptualisation and, so, may be said to involve inference to some degree. At least, this is the case in those situations where one is consciously attempting to work out whether the building is real or not and, thus, is in the process of coming to see it as a building or as a fake. However, in most cases of seeing-as, as in all cases of simple seeing, inference (as a personal level phenomenon) is not involved as conceptualisation is immediate.

Furthermore, in the case of the use of representations like maps and plans you can *look at the symbols and then look at the objects that they represent*

and easily discern between these. However, if in this same manner, you try to *perceptually discern* your phenomenal representings from the external objects that are being represented, you will find that this cannot be done. All you can *see* are the represented objects, not the representings themselves. This is even the case when you use a finger to interfere with an eye in order to create a double image of that which is being seen. In such cases, it is obvious that you are having two phenomenal *representings* of the one object, rather than looking at two distinct objects, but both images locate the object as being external to you and as having features (e.g. colours and shapes) that belong to it rather than to your representings of it. Your representings do not seem to be red and yellow and rectangularly shaped -the object does. Likewise, you cannot discern the object behind the representings -you discern the object *with* the representings. The double image does misrepresent the number of features that the (single) external object actually has, and it also misrepresents a partial shift in the location of certain of the object's properties, but it still represents these misrepresented features as belonging to something external rather than as belonging to the representings. Experience is, in this way, diaphanous (cf. Moore: 1948). The object is seen with the experience in which a selection of the object's properties are represented.

11. INTROSPECTION

Dretske, like Evans and Shoemaker, takes introspection to be a conceptual rather than a perceptual faculty (Evans: 1982 and Shoemaker: 1994b). As such, introspection yields beliefs about our phenomenal representings

and not phenomenal representings of our phenomenal representings. I am happy to accept this characterisation of introspection, though I must note that Dretske's main argument for this position is that there is no phenomenology of introspection over and above the phenomenology of the phenomenal representings that introspection takes as its objects. However, a perceptual account of introspection that took proprioception rather than, e.g. vision, as its model would also have this consequence. Consequently, Dretske certainly does not refute the perceptual model of introspection.

While introspecting we can *think of our representings* in the same way that we think of representational objects such as paintings and, thus, describe features of these representings themselves, rather than the features of the objects that these represent. In such cases there is no phenomenology over and above the phenomenology of the experience itself (and, perhaps, the quasi-aural phenomenology of an internal monologue), and this phenomenology is presentational in nature. As such, such descriptions (conceptual representings of our phenomenal representings) are made in terms of the objects and properties that are phenomenally represented or misrepresented. Our phenomenal representings are taxonimised relationally (as, for example, a visual experience of a blue and yellow ball) even when it is the representings themselves and not the contents that they represent that we are interested in.

Dretske agrees with this much of what I have said about introspection; and it (what I have said) is consistent with RT. However, contra Dretske, I believe that introspection also gives us access to mental facts that are not the sort of facts that he identifies with representational facts. For example, when we are phenomenally representing something, we can concentrate our attention on

features (e.g. colours) of that which is being represented (e.g. a coloured ball) and *think about features of our representings* of these external features of the object. In this way we can become *introspectively aware* of the way in which our experiences are phenomenologically representing those external properties which they phenomenally represent. For example, we can become aware of the fact that the objective red of an object which is phenomenally represented in our experience of that object is phenomenologically represented in a way that we could designate as phenomenal-red rather than as phenomenal-green or phenomenal-blue.

CHAPTER III

MENTAL FACTS THAT ARE NOT REPRESENTATIONAL FACTS

In this chapter I will be concerned with developing the argument that I began in chapter II.8 against the adequacy of RT. In particular, I will argue that there are mental facts that are not facts about indicator functions. At the heart of such facts are qualia. While we cannot perceive our qualia (rather, we perceive *with* the qualia of our phenomenal representings) we are, *in a sense*, continually confronted with qualia in a way that *we can become conceptually aware of* when we introspect our experiences in the manner that I described in the last section of chapter II. The issue of qualia will be dealt with explicitly in chapter IV. Here (in chapter III) I will be primarily concerned with mental facts that are related to the fact that our experiences have the qualia that they do rather than being concerned with qualia per se.

1. TYPE-IDENTITY THEORY

Dretske's account of experience (RT) as being explicable in terms of representational facts (facts about indicator functions) suggests that, if it were true and if the type identity theory of mental events and brain events was true, it would be (in principle) possible to determine the experience an individual was having solely by determining his or her occurrent brain state and correlating it with the property or properties that this type of brain event has the function of indicating. That would give us the content of the individual's occurrent mental

state. In terms of what Dretske says about the content/vehicle distinction and its application to the difference between the mind and the brain this would give us knowledge of both the content and the vehicle of this individual's occurrent state. Furthermore, because all mental facts are supposed to be representational facts, and representational facts are merely facts about indicator functions, this would give us a full knowledge of the mental facts involved. Dretske, himself, says that we cannot gain this knowledge of mental facts solely from facts about the neurophysiological vehicles of representation, but this is because he does not assume the truth of type-identity theory. There are, indeed, good reasons to believe that type-identity theory is not true (e.g. Kripke:1980, pp144-155). It simply does not seem likely that any given type of mental event (say, an experience of the colour shade blue22) can be type-identified with a brain event.

Suppose, however, that type-identity theory was true and that all mental events (or, at least, all types of mental events which constitute phenomenal representings of colour) had been identified with brain event types, and an exhaustive list of these identities had been compiled. Obviously, compiling such a list would require more than just studying the neurophysiological vehicles of representation in isolation, so Dretske would be right in making the above claim even if type-identity theory were true. It would involve (*supposing that RT was also true*) studying which physical features of objects a type of mental event was correlated with and determining from this, and other considerations connected with the theory of natural functions being used, which property or properties the event is supposed to represent. However, once this task was completed, there would be *a (rather loose) sense* in which we could be said to gain a knowledge of mental facts solely from facts about the neurophysiological vehicles of

representation. For example, a particular woman could be determined to be in a mental state that could be characterised as an experience of the physical colour shade blue22 just by looking inside her head at the neurophysiological vehicles of her phenomenal representing; and this could be done even if there was no object with a blue22 trope present, i.e. even if she were hallucinating and this fact was not known to the neurophysiologist in advance.

I would argue that even if type-identity theory were true and an exhaustive list of type identities between mental events and brain events of colour representation were compiled, this would still not give us all of the mental facts. It would not enable us to determine *the way that* the token properties of a given object (in this case, the object's colour tropes) were phenomenologically represented by a token phenomenal representing of the object. We would need to know more than what property or properties it is the function of a type of representational act to represent, in order to determine the way that these were being phenomenologically represented by a token mental act of this type. In particular, we would need to know something about the mental properties of the personal level phenomenal representing that represented these objective, external properties. Knowing which physical property or properties tokens of a specific type of neurophysiological vehicle are supposed to represent would not tell us anything about the way in which a given neural event presented a subject with a particular external property or properties.

2. JACKSON'S MARY AND DRETSKEAN MARY

We can make this same point in terms of Jackson's well-known and widely discussed story of Mary the colour scientist who has all the physical information about colours and colour vision, but who has been kept in a black and white environment (Jackson 1982: p130; and 1997). Mary's knowledge is so impressive that she even knows the neurophysiological (and other physical) chains of events that result in an utterance of "The sky is blue" from someone looking up at the sky. Jackson claims that Mary learns something new about the world and our visual experiences of it when she steps out of her room and experiences colours for the first time. Mary learns *what it is like* to see colours. Her knowledge was, therefore, incomplete even though she had all the relevant physical information that there is to know. Consequently, physicalism is false. In terms of what I have said, we could say that what Mary learns are facts about *how* our phenomenal representings of colour phenomenologically represent colour. She learns how those properties that our colour experiences represent are represented (experienced) by a human subject.

Now, if Mary's knowledge of physical things is kept as complete as Jackson has specified, there remains the possible reply that Dennett has put forward: Mary knows so much more than we could possibly understand, that claiming that she learns something new is not warranted (Dennett: 1991, pp398-401). *We could not possibly imagine the situation as Jackson asks us to, because we could not know what sort of information Mary has, and what this sort of information would allow her to work out in advance about what would happen when she eventually came to see colours.* If we just imagine that she has all of the

relevant information that is available today, then it is not surprising that she would learn something new, but Mary is supposed to have "all the physical information."

I believe that what Dennett says here is correct, but I do not want to talk about exactly the same Mary as this one (Jackson's Mary), nor do I want to argue that physicalism is false or that it is true. Instead, I want to talk about *Dretskean Mary* and I want to argue that RT is false. I want to argue that there are mental facts which are not facts about informational functions. Dretskean Mary is in a very similar situation, but instead of possessing all the physical information, *she possesses all the representational facts about colour and colour vision*. Before looking at this Mary, though, I will look at what Dretske has to say about Jackson's Mary.

Dretske says that what Jackson's Mary lacks before leaving her black and white environment is a property awareness of colours (Dretske: forthcoming, pp12-13). A property awareness of colours is *an experience* (a phenomenal representing) of colours. Mary already knows which objects have which objective colour properties, and she knows that there is something in the brains of people that see coloured objects (e.g. red objects) that has the property of being a property awareness of red. Mary, herself, merely lacks internal states of this sort until she leaves her room. At this point in time, neurophysiological vehicles in her head acquire the property of being property awarenesses of red. Thus, upon leaving her room, Mary becomes property aware of colours but she does not become fact aware of anything that she was not previously fact aware of. Putting this another way, Mary gains an awareness of properties that figure in facts that she already knew. She does not gain a knowledge of any new facts.

Dretske writes that "The fact she was formerly aware of that she expressed as "Experience *e* is a p-awareness of red" can now ... be expressed as "Experience *e* is a p-awareness of *this*"" (forthcoming: p23).

Now, if Dretske is talking about Jackson's Mary (as he *is* doing), then this reply seems reasonable *-given Dennett's point about our inability to fully imagine Mary's scientific omniscience.* Note, however, that it is <u>*only*</u> because of Dennett's point that this seems reasonable. In other words, *if* we assume that Mary knew so much that she may *already* have *somehow* worked out *how* our phenomenal representings of colour phenomenologically represent the colours that they do, *then* it is reasonable to assume (as Dretske does) that she only gains colour experiences without learning anything about these experiences that she did not *already* know. If we do not *assume* that this could be the case, then Dretske's reply is inadequate. Mary would indeed learn *how* our phenomenal representings of colour phenomenologically represent the colours that they do. However, because RT is a far more limited thesis than is physicalism, I need not stipulate that Dretskean Mary has such an unimaginably vast store of knowledge. I will, therefore, *not* have to contend with Dennett's point. This is, in fact, the *only* reason that I am bothering to use Dretskean Mary rather than Jackson's Mary against RT -I want to limit the sources of knowledge available to Mary in order to show that an understanding of representational facts is not enough to deal with a situation like that which is provided by Mary.

Dretskean Mary is a Dretskean scientist who possesses all the representational facts about colour vision. She has the definitive theory of natural functions for the colour-indicating mechanisms of the human visual system. She, therefore, knows precisely which properties it is the function of

human colour vision to represent, and which particular colour properties are represented by any given representational state of such systems. As I did earlier (chapter III.1), I will assume the truth of the type identity theory. I will further assume that Mary can type-identify every token human neural event that represents colour with the appropriate type of experiential event. The result is that *Mary knows the indicator functions of all possible token human neural events that represent colour*. Mary also possesses a little black box which she gets assistants living outside her black and white environment to attach to a human subject's brain. The box relays information back to her about the token neural event that occurs in her experimental patient whenever this person's visual system represents something coloured. With the knowledge that Mary already possesses, she can use this information in order to immediately determine what colour property or properties are being represented. Knowing what physical colour tropes the object that this patient is looking at possesses, Mary also knows if the representational act is veridical or if experimental patient P is misrepresenting the colour properties of the object. This is something that P does not know.

According to RT, Dretskean Mary should know all the mental facts about her patient's colour experiences. However, this is not the case. P knows *how P's visual system is phenomenologically representing the property that it is phenomenally representing*, while Dretskean Mary does not know this until she steps outside of her room and experiences colours herself (assuming, of course, that she and P do not have inverted colour experiences). This fact that P knows may be a fact about a representational act, but it is not the sort of fact that

Dretske would class as a representational fact. It is not a fact about indicator functions.

**Remember, that Dretske claims that Jackson's Mary gains a property awareness (or experience) of colours but that she does not become fact aware of anything that she was not previously fact aware of. She merely gains an awareness of properties that figure in facts that she already knew. She does not gain a knowledge of any new facts. Applied to the case of Dretskean Mary, this response is inadequate. Dretskean Mary will learn that red looks like *this* to a person having experience *e* rather than, for example, like the way objects with the objective property of green look to people in green-representing neural states. This is not to be taken as the claim that she learns that red objects look red and not green, where this is just taken to mean that token neural events that represent objective red, represent red *as being red and not green.* She already knew *this.* The fact that a token neural event phenomenally represents red just means that it represents (veridically or not) something *as being red* (in a nonconceptual sense of "as") and not as being green. Rather, it is the claim that red objects are phenomenologically represented in a way that *for all Dretskean Mary knew prior to having colour experiences herself* green objects may have been phenomenologically represented by her phenomenal representings. She did not know this beforehand as it is something that RT did not equip her to know. If this is right then there are mental facts that are facts about representational systems such as people, but which are not facts about informational functions. This result would falsify RT. Remember, of course, that my claim is about Dretskean Mary and not Jackson's Mary. *With the kind of knowledge Jackson's*

Mary has, she may well have already known that which Dretskean Mary did not know.

**What is more, for all her knowledge, *Mary may not* even be in as good a position to know mental facts about similarity and difference relations between phenomenal representations of external colour properties as someone with a better knowledge than her of what Dretske calls facts about representational systems (where these are taken as including facts about functional organisation), but a very limited knowledge of the relevant indicator functions. I will call this person Mike. *Facts about similarity and difference relations between phenomenal representings of external colour properties do seem to be good candidates for mental facts. They also seem to be good candidates for facts that are functionally definable, i.e. definable in terms of causal relations between inputs, outputs, and other internal states* (Shoemaker: 1975). Consequently, Mike may be able to gain a knowledge of these sorts of facts while being confined to a black and white environment. Mary, on the other hand, might not. Mary may not be able to learn these phenomenal facts via her knowledge of the properties that visual states indicate because these facts may depend upon innate features of our perceptual apparatuses. The same objective properties may be able to be perceived by different creatures as having very different similarity and difference relations (Shoemaker: 1990, p119). If this is the case then there are even more mental facts that are facts about representational systems, but which are not facts about informational functions. However, as I will show later (chapter III.5.b), it is possible that similarity and difference relations can be captured by RT. *In any case, my earlier claim that Dretskean Mary would learn that red objects are phenomenologically represented in a way that for all she knew prior to having*

colour experiences herself green objects may have been phenomenologically represented by her phenomenal representings, is independent of this claim about similarity and difference relations. Consequently, it will be unaffected by the result that this latter claim may be wrong. Furthermore, if the former claim is found to be true, then this is all that I need to refute RT.

3. RT AND FUNCTIONALISM

If I am right and RT as it is currently formulated by Dretske is refuted, it could perhaps still be saved by weakening the scope of its claim. It could, for instance, be put forward as a thesis which covers those areas of our mental life which elude functionalism. It could, perhaps, be held to supplement a weakened version of functionalism, rather than to supersede it as a theory of mind. Lycan, for instance, defends a hybrid functionalist-representationalist theory of mind (Lycan: 1996). Functionalism is currently the most popular theory of mind among philosophers. The central claim of functionalism is that the essential or defining features of any type of mental event are given by the event's causal relations to sensory and conceptual input, to behavioural output, and to the other types of mental states that mediate between this input and output. In its strongest form, all features of mental events are held to be exhausted by their causal roles. Qualia-related problems have undermined support for functionalism in its strongest form (e.g. Block & Fodor: 1972; and Shoemaker: 1975).

However, weakening RT and allying it with functionalism would be against the spirit of Dretske's RT and, therefore, it is something that Dretske himself would eschew. In any case, such a move would not help to save RT

because *that which eludes functionalism also eludes RT. Just as functionalism cannot functionally define an individual quale and, therefore, avoid inverted qualia scenarios* (Shoemaker: 1975), *RT also cannot give a satisfactory definition of a quale in terms of indicator functions*. While I will pursue this problem for RT in chapter IV, it is worth stressing that this not only has the result that RT cannot avoid *problematic* inverted qualia scenarios, but that it is also *at the heart of my claim* that upon her release Dretskean Mary would learn that red objects are phenomenologically represented in a way that for all she knew prior to having colour experiences herself green objects may have been phenomenologically represented by her phenomenal representings. Dretskean Mary learns this new fact because her knowledge of indicator functions did not provide her with any knowledge about the way that indicated colour properties are phenomenologically represented by the experiences that represent them. Because I will (in chapter IV.3) be identifying this type of mental property as a *defining feature of qualia*, we can say that Dretskean Mary's prior knowledge did not provide her with an adequate definition of an individual quale such as a red quale. Add to this problem, the fact that RT *may not be able to* capture some things that functionalism can manage to capture (e.g. similarity and difference relations that hold between experiences in virtue of the way that properties are represented rather than the way that objects are represented), and RT would actually be less impressive than functionalism alone. Even if this latter claim is incorrect, RT supplemented by functionalism would add little to functionalism besides a whole host of problems that it currently does not have to deal with (e.g. problems to do with determining indicator functions).

It is worth noting that Shoemaker has argued that functionalism does not have to worry about the issue of absent qualia which Block and Fodor claim to be a problem for functionalism. The problem of absent qualia is also known as the problem of phenomenal zombies. In its application to functionalism, this is the problem that is posed by the fact that it seems to be a possibility that there could be functional duplicates of creatures which have qualitative states, but which lack qualitative states themselves. This would mean that there is more to qualitative mental states than is given by their causal roles, and this in turn would mean that functionalism is incomplete in an important area relating to our mental lives. Shoemaker argues that, because phenomenal similarity and difference relations can be functionally defined, there can be no cases of absent qualia in functional duplicates. Dretske, on the other hand, admits that RT has the consequence that absent qualia are possible if molecular duplicates of representational systems miraculously materialise out of nowhere and, therefore, lack any history which could have bestowed functions upon their states (Dretske: 1995, pp141-151). If Shoemaker is right, then the problem of absent qualia is another problem that RT has, but functionalism does not. However, it is debatable whether or not Shoemaker is right (e.g. Block: 1980; and Shoemaker: 1981). In any case, I will not be concerned with the problem of absent qualia. My main concern is, rather, the problem of defining an individual quale, as there is general agreement among those philosophers who accept that there are qualia, that this is a problem for functionalism. I believe that it is also a problem for RT. I am also interested in pursuing the issue of phenomenal similarity and

difference relations between experiences because this issue is not taken to be a problem for functionalism at all -*but it still may well be a problem for RT.*

4. DRETSKE'S RESPONSE TO THE PHENOMENAL SIMILARITIES AND DIFFERENCES ARGUMENT

I have claimed that RT *might not be able to* capture similarity and difference relations between experiences, in virtue of the way that they represent the properties that they represent. Having recently spoken with Dretske about this issue, I will discuss some of his thoughts about it. What Dretske had to say was consistent with RT and it does provide a workable and empirically falsifiable response to my argument. It allows us to further elaborate the details of RT.

The claim that I made to Dretske was that while RT provides an account of similarities and differences that hold between two representational systems' experiences of the same objects, *it does not provide an account of similarities and differences that hold between their experiences of the same range of values of the same properties.* Following Shoemaker, I suggested that there could be a creature that represented the same objective colour properties as us, but which represented them in such a way that different similarity and difference relations held between its experiences than those that hold between ours (Shoemaker: 1990, p119). I gave the example of a creature looking at three balls with objective colour properties that we (upon the basis of our own experiences) would describe as red, orange, and green, respectively. In other words, these balls have physical property tropes that we would phenomenally represent as red, orange, and green. I stipulated that it was whatever microstructural features of these balls

that our experiences represent in this way, that this creature also phenomenally represents. If the states that we enter when we look at these objects have the function of indicating what we could call objective-red, objective-orange, and objective-green, respectively, then the sensory states of this creature also represent these exact same objective colour properties. However, I claimed, it seems clearly conceivable that if the creature was cognitively sophisticated enough, it might look at these three balls and truthfully claim that the red ball resembles the green ball in virtue of colour more than it resembles the orange ball. Even if it were not sophisticated enough to make such a claim, its experiences would still present the red and green balls as being more similar than the orange ball. Put another way, it would be easier for such a creature to mistake a red ball for a green ball than for an orange ball. I claimed that because the creature's colour representing states have the same functions as ours, and yet these states represent that which they represent in a different way than our states do, there are, therefore, mental facts about phenomenal similarity and difference relations that are not facts about informational functions. Therefore, I claimed, RT is false.

Dretske replied a creature could not represent red as being more like green than orange because this would be contradictory. It seemed to me at the time (*and it may well have been the case -though not necessarily*) that he was making a conceptual point about what *we* (humans) call colour. For example, if a person said that red objects looked more like green objects of the same type than they looked like orange objects, we would not know how to take this as a serious claim about colours. We might suspect that either this person had not mastered our colour vocabulary, or that there was something seriously wrong with him or

her. Part of *our conceptual undersanding of red* is that red things are more similar to orange things than to green things, in virtue of colour. In this sense, the claim that red is more similar to green than to orange is contradictory. However, my claim is ontological rather than conceptual. *If we accept* a Dretskean position of colour realism, whereby the extension of our colour terms are objectively determinable features of objects, *then there seems to be no contradiction* in the claim that *these same features* can be phenomenally represented in such a way that the relations of phenomenal similarity and difference that hold between them can *differ* between two different representational systems that have indicator functions relating to these *same extensional properties.* This would, in fact, be the case *if such similarity orderings are imposed upon represented properties as a product of innate features of the perceptual apparatuses of the systems in question.* I will explain this claim in a little more detail.

Essentially, my argument about Dretskean Mary not knowing mental facts about relations of phenomenal similarity and differences, presupposes a Quinean view about quality spaces such as the human colour quality space. Quine notes that in order to explain a child's ability to learn to use colour words correctly we have to attribute a pre-linguistic quality space to the child with which certain physical stimuli can be discriminated from one another, and in such a way that some stimuli are more difficult to discriminate than others (Quine: 1960, p83). While this view identifies colours with external stimuli, Quine adds that *the spacing of such stimuli is "within" the child* (1960: p84). Thus, his view seems to be that a human-relative similarity ordering is placed on the physical (colour) stimuli that are perceived by people. Shoemaker notes that

the similarity orderings that exist among sensory events produced in specific kinds of creatures correspond to the similarity orderings of physical stimuli that are relative to such creatures (Shoemaker: 1991, p509). ***If this is the case*****, then Dretskean Mary lacks a knowledge of mental facts about relations of phenomenal similarity and differences among colours, simply because she lacks knowledge about the way the human visual system structures the relevant stimuli via sensory events with a human-relative similarity ordering.**

In conversation, Dretske did go on to say that the similarity and difference relations that hold between colours are presumably also due to objective features of objects, just as those things that make shapes similar and different from one another are objective features of objects. I believe that this, ***in itself*****, poses no insuperable problem for my position as I can readily acknowledge that this is the case, while insisting that our colour quality space may still be Quinean. For this reason, I will show that** ***it is crucially important for RT that it is not only the case that the similarity and difference relations that hold between colours are due to objective features of objects, but that there is also an objective (perceiver-independent) similarity ordering that holds among them and that this ordering must be phenomenally represented in such a way that it is preserved as part of those phenomenal acts whereby colour properties can be said to be represented.*** **In other words, RT requires that the similarities that we perceive as holding between colours be an objectively determinable perceiver-independent feature of coloured objects which we phenomenally represent as part of our phenomenal representings of the colours of objects. If this is not the case, then Dretskean Mary lacks a knowledge of mental facts about phenomenal similarity and difference relations among colours; in which case RT is false.**

Later (chapter III.5), I will look at a way in which objective similarity orderings could be supposed to be *preserved and represented* as an informational function of a given cluster of property-representing states. *If empirically accurate, this would refute my claim that RT cannot adequately capture such similarity orderings between the way colour properties are represented.* It would also suggest a way in which Dretske's claim about the contradictoriness of the idea that red could be more similar to green than to orange could also be interpreted as an ontological, rather than just a conceptual, claim. As I have said, if RT can deal with such facts in such a way, then my claim that RT cannot adequately deal with some mental facts that functionalism can will be refuted. It must be noted, however, that it could be the case that RT *still* might not be able to adequately capture phenomenal similarity orderings between the way other properties (*especially other secondary qualities*) are represented. If RT cannot capture *these*, then RT would still be false. However, as I will show, it remains an open question as to whether similarity orderings among *colours* are perceiver-dependent or perceiver-independent features, and the balance of available evidence is certainly not in favour of the latter position; although the possibility that this position is correct and that, therefore, objective similarity orderings may be preserved and represented as an informational function of physical colour-representing states, still exists.

Furthermore, even if *(as Dretske requires)* such a position does turn out to be the correct one, my claim that RT has the same difficulties as functionalism will stand -in particular, *the problem of satisfactorily defining an individual quale.* In Dretske's case, such a definition needs to be given in terms of indicator functions and this, I will argue (in chapter IV), is not possible.

5. PHENOMENAL SIMILARITIES AND REPRESENTATIONAL DUPLICATES

(a) Thought-experiment

Let us imagine that there are two simple creatures (A and B) that are Dretskean representational duplicates. They both have only one sensory modality (a rudimentary analogue of colour vision) and they can enter only one of three possible types of phenomenal state (1, 2, or 3) at any one time -each of which represents a specific objective type of colour property which it is supposed to represent as being instantiated. To make the case even more specific, let us assume that token states of these types are tokened when the following shades of colour are instantiated in objects:

State 1 = Red21;
State 2 = Orange33; and
State 3 = Green18.

The numbers correspond to specific, arbitrarily chosen, shades of the colour that they are attached to. Thus, whenever A or B enter type 1 states, their visual systems represent that those objective properties of an object that the human visual system represents as red21 are instantiated. Every instantiation of such features that the visual systems of normal colour-sensitive humans would (under standard conditions) veridically represent as red21 are also, if veridically represented by A and B, phenomenally represented as red21. Any such properties that the visual systems of normal colour-sensitive humans would not (under standard conditions) veridically represent as red21 are also not veridically represented by A and B as red21. The same applies, mutatis mutandis, to type 2 and type 3 states. Standard conditions are the same for both

A and B. Each of the three types of states represent those objects that they represent as instantiating a particular shade in such a way that they are easily discriminable from objects that are represented as instantiating either of the other two shades. Likewise, two objects that are represented as instantiating the same particular shade are represented in such a way that they are qualitatively identical to each other in virtue of colour. By "qualitatively identical" I mean that if A or B were cognitively sophisticated enough to be presented with two such objects one at a time, were told that the first object was called j and the second was called k, and then shown the two objects one after the other in random order (as we will see below A and B can only represent one colour trope at a time) they could not distinguish j from k.

It may be argued that in order to phenomenally represent colours A and B would, like humans, also have to represent a variety of spatial properties like the size and shape and perhaps even the distance away of the coloured surfaces that they represent. They might, therefore, be able to distinguish j from k in virtue of any such differences as these unless j and k are qualitatively identical in respect to these properties also. However, in order to keep A and B as simple as possible, I will stipulate that A and B can only represent the three shades that they do when they are standing within a very narrow range of close distances from the external object that they are representing, that they can only represent a relatively small part of a coloured surface (say an area of about 5 centimeters x 5 centimeters) at any one time and that unless *all* of that represented area is homogeneously red21, orange33, or green18, it will not be phenomenally represented at all. Thus, if shape and size can be said to be represented at all, it is only in terms of the (5cm x 5cm) size of the creature's visual field; the visual

field, itself, not being something that is represented by either creature's experience. It can also be stipulated that these restrictive features do not adversely affect the survival chances of A and B in the environment that live in. Furthermore, even though they each only have three types of representational states worthy of being classified as experiences, they also have a whole range of hard-wired behavioural routines that enable them to survive in their adapted environments.

Now we get to the important part. Even though A and B are Dretskean representational duplicates (they both have three types of visual or, perhaps, quasi-visual states which have the same indicator functions as each other's), the similarity and difference relations that hold between these states are different in the following way:*Creature A's type 1 states are more similar to its type 2 states than they are to its type 3 states; while creature B's type 1 states are more similar to its type 3 states than they are to its type 2 states.* This means that creature A's experiences represent objects that have red21 tropes as being more similar to objects with orange33 tropes than to objects with green18 tropes. This is the same way that our experiences represent such objects. Objectively red objects just look more like objectively orange objects than they look like objectively green objects. On the other hand, creature B's experiences represent objects that have red21 tropes as being more similar to objects with green18 tropes than to objects with orange33 tropes. If creature B was cognitively sophisticated enough it would, if asked, say that objectively red objects just look more like objectively green objects than they look like objectively orange objects.

(b) <u>Discussion of thought-experiment</u>

Contrary to Dretske's RT, I have just described a situation in which the facts about the informational functions of A's and B's states are both the *same* (they both have phenomenal states that represent the same values of the same properties), but there are *different* mental facts in both cases. If such scenarios are possible they falsify RT. However, Dretske could argue that the particular scenario that I have looked at here is incorrectly described. I contend that this is a case in which two creatures are phenomenally representing the same values of the same properties, but are phenomenologically representing them in such different ways that there are different similarity and difference relations holding between their respective representings. Dretske, on the other hand, can argue that A and B must be, in effect, perceiving different properties because those features which make something that is red21 more similar to something that is orange33 than it is to green18 are important objective features that are (at least partly) constitutive of what red21 actually is. Therefore, *if these features are not represented*, then red21 is not represented.

In reply to this I must stress that *it does not follow from the fact that certain features are not represented in a certain way by a phenomenal representing that they are not represented at all*. I have said that A's and B's states (when veridical) represent every instantiation of objective features that the visual systems of normal humans would (under standard conditions) veridically represent as red21, etc. Any such properties that the visual systems of normal colour-sensitive humans would not (under standard conditions) veridically represent as red21, etc. are also not veridically represented by A and B as red21, etc. I assume that this means that I am including as part of red21, for example,

those objective features which make something that is red21 more similar to something that is orange33 than it is to green18, *relative to us.*

Of course, it is still possible that while A's and B's states are tokened when the same sets of microstructural features are present, both creatures' states represent different subsets of those features. This would be an analogous situation to that in which two representational systems represent different selections of properties of the same objects. However, I will stipulate that this is not the case. Both A's and B's states represent the *same* selections of microstructural features as our's do and, yet, *different* similarity relations hold between B's states than our's. If there are good reasons to believe that this is not possible then this stipulation amounts to nothing, but I do not believe that there are such reasons. And, if there are no such reasons, then I can accept that the similarities that *we* perceive among colours *are given by* objective features of objects without conceding that Dretskean Mary knows the relevant facts about the similarities and differences that hold among our colour experiences.

It certainly seems possible that those objective features that make two objects look more similar to each other in virtue of colour than they do to a third object, relative to a certain kind of creature, *can be* objective features that are represented by another kind of creature, *without* that creature representing these as more similar to each other than to the third. As long as it is recognised that there are, in this case, three distinct colour properties, then it must be recognised that these are still *different* properties. And if they are different properties, then it seems possible that they can be represented *without being represented as having the same similarity or difference relations among them that our experiences represent them as having*. This could be because there are *good evolutionary*

reasons for a creature to have sensory states that enable it to make the same discriminations between physical stimuli that a second kind of creature can make; while there are also good evolutionary reasons for it to have sensory states which enable it to easily discriminate between stimuli that the other creature finds it difficult to distinguish between, and no such reasons for it to be able to easily discriminate between the stimuli that the second creature finds it easy to discriminate. Note, that in both cases it will still be objective features of the stimuli which are responsible for them looking more similar to (or different from) other stimuli, relative to one or the other creatures, because it is objective features which make the stimuli different from each other (and, therefore, discrimin*able*) in the first place. Furthermore, *the same objective features* that make two different stimuli easily discriminable from each other, relative to the first creature, can be among the features of the two stimuli that the second creature represents, and vice versa. The only difference is that innate features of their perceptual systems structure the represented stimuli via sensory events of a different character.

This, in turn, means that an understanding of indicator functions would not be enough for an understanding of mental facts about phenomenal similarity and difference relations holding between phenomenal representings of objective properties of objects. *However, I believe that Dretske can argue that there are objective similarity orderings among certain types of properties that are preserved and represented as an informational function of a given cluster of property-related mental states, such as those that represent colour.* Dretske could, for example, argue that it is not just the case that there is an objective cluster of physical stimuli with certain specific features that provides the extension of our

phenomenal and conceptual representings of red and another such cluster that provides the extension of our phenomenal and conceptual representings of green and so on for all colours and (sliced more finely) for all shades of colours. He could also add that each of these physical features has an objective, perceiver-independent value via which they are objectively related to each other, *such that they can only be properly be said to be veridically represented if that value is preserved as part of the representing of the properties.* The value of red, for instance, would be such that it is closer to the values of either blue or yellow than it is to green; while it is even closer to orange than it is to yellow. Consequently, we would expect our experiences of red objects to be more similar to our experiences of orange objects than they are to our experiences of green objects. Experiences that did not preserve this similarity ordering *would not be veridical experiences of these same properties.* According to Tye, the properties that our colour experiences represent are, in fact, ontologically of this nature (Tye: 1995, p148). He contends that these external properties can be represented in a three-dimensional space as a closed circular loop with each physical value as coordinates in this space. This would relate each value to each other value in an objective similarity/difference ordering whereby coordinates that were closer to each other were more similar than coordinates which are further away. As I will show, however, this claim of Tye's is mistaken, but it does illustrate a way in which physical colour properties (whatever properties these happen to be) could be found to have perceiver-independent features which are responsible for the similarity orderings we perceive to hold among colours. If physical colour properties do happen to turn out to have such features, then my claim that

Dretskean Mary does not know facts about the resemblance and difference relations that hold among P's colour experiences would be wrong. However, I may still right about the other type of facts that Dretskean Mary does not know. Likewise, if physical colour properties do have such features, then I am wrong in claiming that creature B could represent three of *the very same physical colour properties as us*, but such that the resemblance and difference relations between them were different. Again, however, if my other claim about the facts that Dretskean Mary does not know is right, then I could transform creature B into a Dretskean representational duplicate of A in terms of *represented properties and represented similarities*, but with inverted qualia and argue that someone like Dretskean Mary would need *more than just all the representational facts* to determine that this was the case. The problems that such inverted qualia scenarios pose for RT will be looked at in chaper IV.

Furthermore, *it is certainly not necessarily the case that* what we call red or orange or green are objective properties with objective perceiver-independent values that relate them to each other in an ordering that must be preserved if they are correctly to be said to be veridically represented. *This is just one possibility*. And, if it is not the case that colours are like this, then the similarity ordering that we perceive colours as having may be a similarity ordering that is imposed upon physical stimuli by innate features of the human perceptual system. Now, if this actually happens to be the case, *then there certainly could be a creature such as B* that could veridically represent only a small selection of such properties but with a different similarity ordering obtaining among these. B could represent these very same features of objects that we represent but in a different way. In the next section, I will discuss what is known about the physical

features of objects that correspond to colours, and show that *Tye's particular identification of colour properties is wrong*. This discussion will show that *it is still an open question* as to whether or not the similarity ordering that we perceive to hold among colours is a perceiver-dependent or a perceiver-independent feature of objects. Put another way, it is an open question whether our colour quality spaces are Quinean or Tyean spaces. If they are Tyean spaces, then the human visual system *could* have the function of representing colours *as* properties that have a given, objective ordering. This could be done *via* sensory events which mirror this ordering *by* representing stimuli which are objectively close together *as* being more similar than stimuli which are further apart in the ordering. This would, *provided that our visual systems was found to have such a function*, make facts about colour similarities facts about indicator functions. If our colour quality spaces are Quinean, however, then facts about colour similarities could not be facts about indicator functions and RT would be false. The same also holds of all our other quality spaces too (if just one such space was Quinean, we could construct a Dretskean Mary type of thought experiment involving that quality space and show that RT is false), but I am only concerned here with colours. Note, also that I will use the term "Tyean" as a general term to designate a quality space which involves stimuli with a perceiver-independent similarity ordering. In the case of colour spaces, therefore, a Tyean space is *not* to be strictly identified with one which identifies colours with triplets of average relative reflectance as I will show that Tye does.

6. DRETSKE'S OBJECTIVISM ABOUT COLOURS

As has been suggested by the discussion so far, Dretske's theoretical approach to the study of experience commits him, like Tye (1995:pp146-150), to a position of objectivism about secondary qualities such as colours. Colour objectivism is a position that takes those properties that our colour experiences are of to be independent of our experiences in the sense that if all the perceivers of colours in the world were annihilated or were made colour blind colours would still exist as part of the world. Objectivism requires that those physical properties which are candidates for being colour properties, be distal properties that our visual systems can approximately track or detect. Proximal properties (such as patterns of irradiation on the retina) are not candidate properties because they do not qualify (as colour properties must) as properties that we see (Thompson: 1995, p111). Exactly what properties are to be identified with colours, however, is a source of difficulties for objectivists. There does, however, seem to be general agreement that the only relevant candidate properties are spectral reflectances and emittances, and that these are anthropocentrically grouped in the sense that it is only relative to creatures such as ourselves that those spectral reflectances and emittances that we experience form interesting groupings (Hardin: 1996, p339).

One way of identifying colour properties with physical properties is to argue that every different spectral reflectance profile of an object is a particular determinate colour in the sense of being a perceiver-independent property of objects. The major problem with this, however, is provided by the well-known phenomenon of metameric matching: depending upon illumination conditions, a host of different spectral reflectance profiles can give rise to the same colour experience (e.g. red). It is because of this problem that Hilbert says that "colour

perception and colour language give us anthropocentrically defined colours and not colours themselves"(quoted in Thompson: 1995, p115).

Hilbert still thinks that anthropocentrically defined colours are, like "colours themselves", objective. They are specifiable in purely physical terms even though they are defined in terms of the human visual system. Anthropocentric colours correspond to triplets of average relative reflectance (Thompson: 1995, p116). It is this identification of triplets of average relative reflectance with colours that Tye has recently come out in support of (Tye: 1995, pp146-148). Furthermore Tye even uses it as a model to posit identities between other secondary qualities (such as smells, tastes, and sounds) and physical properties (1995: pp149-150). As such, it is possible that Dretske could also use this identification of colour properties to support his own representationalism. He does *not* do this however, but considering that Tye believes that this posited identity yields an identification of colour similarities with perceiver-independent features of objects, it is worth briefly considering the viability of such an approach.

The problems associated with the identification of colours with triplets of average relative reflectance have been recently discussed by Thompson (1995). There are two particular problems with this identification, both of which relate to Hilbert's objectivist interpretation of retinex theory. It is this interpretation which grounds the posited identity between colour and triplets of average relative reflectance. The first problem is that Hilbert substitutes axes of integrated reflectance for the axes of lightness which retinex colour space is composed of. This is a problem because integrated reflectance can be measured by a meter, while lightness is a perceptual property that is only measured by the

visual system (Thompson: 1995, p125). Unlike integrated reflectance, lightness is not an external property with which we can identify colour (p127); and, while lightness triplets serve as designators of triplets of average relative reflectance, lightness is not always correlated with reflectance (p128). Thus, if colours are type-identified in terms of reflectance, there will be important generalisations about the colours we see that will remain uncaptured. One such generalisation is that colours as we see them (i.e. anthropocentric colours as opposed to Hilbert's objective "colours themselves") depend upon the amount and the quality of illumination (p128). ***The end result is that colours would have to be identified with lightness triplets rather than with triplets of average relative reflectance, and that lightness triplets do not provide us with a physical, perceiver-independent specification of anthropocentric colour.***

The second problem is that the lightness colour space of retinex theory fails to model important features of the phenomenal structure of colour. Specifically, it does not successfully model opponent relations among colours (i.e. red-green and blue-yellow opponencies) and the unique/binary structure of hue (e.g. red, green, blue, and yellow versus orange, purple, etc.) which results from these opponent relations (Thompson: 1995, pp125-126). ***There is no objective perceiver-independent way of modelling these out of the lightness colour space.*** **Furthermore, the identification of lightness triplets with colours, leaves us with little understanding of how single chromatic perceptions can be formed from different values of the achromatic quality of lightness. The result is that lightness comparisons do not allow us to understand the phenomenal structure of colour (Thompson: 1995, p129).**

Because of these problems, the identification of colours with triplets of average relative reflectance does not work. Consequently, *this particular identification* does not succeed in demonstrating that our colour quality space is Tyean. Unlike Tye, Dretske *does not* treat the issue of which physical properties are to be identified with colour properties (and, therefore, which physical properties human colour vision has the function of indicating) as relatively unproblematic. Instead, Dretske makes the point that even though the conditions that cause us to see red are many and varied, all this shows is that the property that red is may not be obvious from the variety of conditions that now cause us to experience red (Dretske: 1995, p93). It does not show that that which we experience when we experience red is not an objective property. For this reason, Dretske (following Hilbert) says that "colour is whatever property it is the function of colour vision to detect". As such, it is still always possible that objective anthropocentric colour properties *could* turn out to be properties that have objective, perceiver-independent relations among themselves that are mirrored in the similarities of our phenomenal representings of colour, such that any creature that can be said to veridically represent the same colours that we do must also represent these same similarities as holding between them. However, whether colours are actually like this or not is an open question. If they are not, then colour similarities are mental facts that are not facts about informational functions.

Furthermore, I must stress again that even if it does happen to be the case that such similarities are perceiver-independent features of objects, Dretskean Mary would still not know how those physical features of objects that we represent when we represent saffron (a shade of orange) were

phenomenologically represented by our phenomenal representings until she left her black and white environment. Consequently, she would *also not know* if P's phenomenal representings phenomenologically represented the colour property saffron in the same way as her own did solely on the basis of Dretske's representational facts. This is, of course, the problem of inverted qualia. It will be looked at in more detail in the next chapter.

CHAPTER IV

QUALIA

In chapter III.2 I claimed that, upon Dretskean Mary's release from her black and white environment, she learns that red objects are phenomenologically represented in such a way by her phenomenal representings that, ***for all she knew prior to having colour experiences herself,*** **green objects may have been phenomenologically represented. She did not know this prior to her release because RT did not equip her to know it. As I have stressed, if this is right then there are mental facts that are facts about representational systems such as people, but which are not facts about informational functions; and this falsifies RT. This claim that I have made about Dretskean Mary is a claim about her knowledge of qualia.**

I will now discuss what qualia are and what Dretske says qualia are. In particular, I am interested in assessing the adequacy of RT in dealing with qualia. I will argue that RT can adequately deal with qualia as defined by Dretske, but that it cannot adequately deal with the conception of qualia that is associated with the traditional problems of qualia for functionalism and materialism in general. To be more specific, just as is the case with functionalism, ***RT cannot give a satisfactory definition of an individual quale and, therefore, it cannot avoid problematic inverted qualia scenarios.*** **Furthermore, it is because RT cannot give a satisfactory definition of an individual quale that Dretskean Mary learnt something new when she left her room. As such, contra Dretske, RT**

cannot satisfactorily deal with problems that have dogged other naturalistic theories of mind.

1. QUALIA INTRODUCED

"Qualia" is a philosophical term which is used to refer to properties of qualitative mental events (experiences) which are responsible for the qualitative character of such events. Qualia are, paradigmatically, phenomenal properties of occurrent, personal level mental states; or *phenomenally characterised features of experiences.*

The issue of whether or not qualia are (wholly or partly) intentional properties is an important issue. If qualia are identified with, or explained in terms of, intentional properties this could allow an identity to be postulated with, or an explanation to be framed in terms of, another personal level phenomenon. A naturalistic theory of intentionality could then be employed to naturalise both intentionality and qualia at one stroke. This is a very attractive option to the naturalist as greater results have been achieved in naturalising intentionality than is the case with qualia.

This is, in fact, Dretske's own strategy. He identifies qualia with intentional properties which he claims are given by indicator functions. This allows him to account for both intentionality and qualia naturalistically. Regardless of the virtues of this strategy, the identification of qualia with these intentional properties must be shown to leave out nothing which is important to what qualia are. *I believe that Dretske fails to do this.*

2. DRETSKE'S ACCOUNT OF QUALIA AS INTENTIONAL

Dretske provides an *intentional* characterisation of the personal level phenomenon of sensory experience which I believe to be correct, but which I do not believe captures the full *representational* richness of experience. Consequently, even if he can fully account for the intentional properties of experience, this is not sufficient to account for all the representational properties of experience. In particular, I will argue that it is not sufficient to account for an experience's qualia. *Qualia are mental properties that have representational features other than intentional features.*

As I have shown (chapter II.6), Dretske makes a distinction between two types of personal level representings: conceptual and nonconceptual representings. Nonconceptual representings are phenomenal representings, or experiences. Dretske shows that experiences display all the familiar properties of intentionality. They have aboutness, in the sense that they are about things in the world (1995:pp28-30); they can misrepresent an object as being F when there is no object there (hallucinations) or it is not F (illusions) (1995:pp27-28); and they have aspectual shape (1995:pp30-32) or sense (1995:pp23-24). This provides a good case for the representational nature of experience. Furthermore, I agree with Dretske that the only properties that experiences have are representational properties. The main problem that I have with RT is that I believe that there are *other representational properties* that experiences have besides intentional properties, and that RT cannot account for these. I will look at these in the next section.

Dretske identifies the qualia of an experience with the aspectual shape or sense of the experience, though he does not explicitly state this as such. Rather, he simply offers the same definition of qualia as I have shown him (chapter II.7) to make of aspectual shape/sense. For example, he writes that "the quality of experience, how things seem to us at the sensory level, is constituted by the properties things are represented as having" (Dretske 1995:p1). Elsewhere, he writes that "the subjective quality of an experience, the phenomenal appearances, are the way experience systemically represents things to be" (1995:p22). While he uses the expression "quality of an experience" rather than "qualia" in these examples, he does explicitly use the latter to designate the same thing in his chapter on qualia (1995:chapter 3). For instance, he states that his representational approach to the mind identifies "qualia with the properties that the experience systemically represents things as having" (1995: p72).

I agree that we can *partly* identify those things that are given by such terms as "phenomenal appearances" and "the quality of experience" and (more commonly) "qualia" or "the qualitative character of experience" with the aspectual shape or sense of phenomenal representings. In other words, I agree that the properties that an experience represents contribute to the qualitative character of experience. However, I do not agree that aspectual shape (an intentional property of experiences) is the sole determinant of qualitative character. Therefore, we cannot simply identify qualitative character and aspectual shape; there is also something else that contributes to qualitative character. To be more precise, it is an experience's *representings of external properties* that constitutes the qualitative character of experience and *it is only because of this*, that the properties that an experience represents contribute to the

qualitative character of the experience. ***These representings of external properties are themselves to be identified with qualia*****, and they have properties other than intentional properties.**

Qualia are those properties that are constitutive of the qualitative character of experience. They are phenomenal, or mental, properties. Dretske also says that qualia are phenomenal properties, but his explication of what these particular phenomenal properties are identifies them with intentional properties (1995:p73). ***Dretske's notion of qualia identifies qualia with those objective properties of external objects that it is the function of sensory events to represent.*** **Intentional properties are, indeed, genuine (relational) mental properties. Even when intentional properties are given a Dretskean twist and are equated with indicator functions, they remain mental properties; they simply become naturalised mental properties. In other words, Dretske does not merely identify the qualia of an experience with the colour, shape, and other tropes of an object which is being looked at. Rather, he identifies the qualia with the tropes that an experience** ***represents*** **(*****qua instantiations of properties which the experience has the function of representing*****) as being instantiated in an object. These are** ***intentional properties*** **which can, when the experience is veridical, be** ***specified*** **in terms of an object's tropes; but they are not identical with them. In nonveridical experiences,** ***the properties that are represented as being instantiated*** **do not correspond to the tropes that the object which is represented actually has. The tropes that the object is represented as instantiating are different from the tropes that are instantiated in the object and, so, such cases make it clear that it is** ***represented properties rather than physically instantiated property tokens (tropes)***

which provide the qualia of the experience. Represented properties are intentional properties of experiences.

As such, Dretske does identify qualia with mental properties. It is my contention, however, that this identification is incomplete. I believe that the qualitative character of experience can be partially described in terms of the external properties that an experience represents as being instantiated, but that there is more to it than this. In particular, the qualitative character of experience is *constituted by mental properties which not only represent external properties but which represent them in one way rather than another. I identify these mental properties with qualia.*

3. QUALIA DEFINED AS *BOTH* REPRESENTATIONAL *AND* INTRINSIC *PROPERTIES OF EXPERIENCES*

I believe that Dretske's characterisation of qualia solely as intentional properties of experiences is inadequate, but I still believe that qualia can be fully characterised as *representational properties* of experiences. However, some of the properties which I characterise as representational properties are more usually characterised as intrinsic properties. I will explain this and, then, I will move on to offer a definition of an individual quale which I believe to capture that which is so problematic for naturalistic theories like Dretske's to account for.

I have claimed that qualia can be fully characterised as *representational properties* and I mean this in two different senses:

(1) *Firstly,* a token quale is a *representing* of a trope. The trope, itself, is the *represented of the quale.* When there is no trope that is being represented, the represented is an uninstantiated trope of a certain type. When a quale represents

a trope as being a trope of another type than that which it actually is, the trope is simply misrepresented.

(2) *Secondly, qualia are properties of phenomenal representings* of objects. When I have a veridical experience of a tomato, for instance, those particular microstructural features of the tomato which are what my experience represents as red constitute *the red trope of the tomato*; while the representing of the red trope constitutes *a quale of my experience.* Being a part of my token *representing* of the tomato, rather than a part of the token tomato (as the red trope is), makes the red quale *a property of my phenomenal representing* of the tomato.

Being a property of a representing (2 above) is, I contend, a representational property, but it is not an intentional property. In this second sense of how qualia are *representational properties*, qualia can also be characterised as *intrinsic properties* . This is because:

(a) qualia are *properties of mental events*, rather than properties of that which is represented by such events; and

(b) qualia are *nonrelational* (and, therefore, intrinsic) mental properties *to the extent that they are not exhaustively characterised by* their intentional properties of being *about* tropes of certain types.

I have just explained how qualia are fully representational, but not solely intentional, properties of experiences. I have also shown how certain of their representational properties can be characterised as intrinsic properties. I will now offer the following <u>definition of an individual quale</u> which builds on this discussion: *A token quale can be <u>defined</u> as an intrinsically mental property that represents a token trope of a certain type <u>in a certain way</u> rather than another possible way.* The type of tropes that can be veridically represented by a token

quale are those which constitute instantiations of the particular type of objective property which tokens of that type of quale represent. For example, emerald-green qualia veridically represent those tropes of emerald-green objects which constitute the objects' emerald-green colour.

Representing tropes in certain ways rather than other possible ways may seem like an intentional property of a quale that could be captured by a notion of the sense or aspectual shape of a trope, but it is not. For, while the sense of an object is constituted by a selection of the object's properties, the way that a quale represents a trope is not by representing a selection of the trope's features. For, I claim, the exact same set of microstructural features which constitute the emerald-green trope of an object (for instance) could be veridically represented in different ways by different creatures. This is because the way that the trope is represented by a quale is not determined solely by features of the trope itself but also by innate features of the perceptual systems which phenomenally represent tropes as instantiations of certain properties. Note, that this claim is independent of whether or not the human colour quality space is Quinean. Even if our colour space is Tyean and a perceiver-independent similarity value which must be preserved by our phenomenal representings is part of the represented features of an emerald-green trope, an emerald-green trope could still be veridically represented in *different* ways by two creatures representing the *same* trope.

What this means is that a different type of quale would be tokened in each case. Both would be emerald-green qualia in the sense that they both represent an emerald-green trope *and* they both represent this trope as being more similar to a blue trope than to a red trope, but they would be different qualia in the sense that they represent the emerald-green trope *differently*. For

instance, two creatures with symmetrical colour spaces could have inverted qualia such that the quale that represents the emerald-green trope in one creature's experiences is similar to the quale that represents a trope of a certain shade of red in the other creature's experiences, while both creatures' qualia represent emerald-green tropes as being more similar to blue tropes than to red tropes.

For this reason I will italicise, and add an asterisk to, "emerald-green qualia" (thus: *emerald-green* qualia*). This signifies that while qualia can be taxonomised via the external properties that they represent (e.g. the property of being emerald-green), this taxonomy does not capture the differences that could exist between two qualia which are representing the same trope of an object. As such, I will use the same procedure of italicising and adding an asterisk in all cases in which I am characterising a quale in terms of the property that qualia of this type represent. While sometimes I will refer to the qualia which represent properties as specific as emerald-green, other times I will refer more generally to the qualia associated with colour properties (thus: *colour* qualia*) or even to the qualia associated with external properties in general (thus: *property* qualia*). This will make such terms neutral as to *the way that* the qualia concerned are phenomenologically representing the token properties that they represent, while recognising that these token properties *are* being represented by such qualia in certain ways which could be different in another creature's experiences.

4. PHENOMENAL COLOUR

Phenomenal colour is often held as being paradigmatic of qualia. I agree that it is, but in order to appreciate why this is so, certain issues relating to colour have to be made clear. The first of these is the issue of whether or not colours and phenomenal colours the same thing. And, the other issue is how to define phenomenal colour in order to successfully use it as a paradigm of qualia. In this discussion I will use the example of purple, but it should be understood that purple can be substituted for any other colour. I am talking about colours in general, but (as elsewhere) it is convenient to use a single colour by way of example.

Like Dretske, I am happy to assume that colours are objective microstructural features of objects. This means that our colour words refer to *these* objective features. If I say that an object is purple, then I am making the claim that the object has a certain trope of a property that can be (in principle) specified independently of it ever causing phenomenal representings in, or ever being picked out or identified by, anyone or anything. As such, *it is the object that is purple.* The phenomenal representing of the object's colour is *not* purple. It is a representing *of* purple. If Dretske is correct, that which makes a token phenomenal representing into a *representing of purple* is the fact that tokens of this type of representing are produced by a representational system (colour vision) which has, as one of its *systemic functions,* the function of representing purple microstructural features of objects. The token representings of purple tropes derive their systemic indicator functions from this function of the colour vision system. *They are representings of purple because they are supposed to be representings of purple.*

In order to differentiate the phenomenal representing of purple from the purple trope of an object, I have used the term "purple qualia". However, I have noted that the phenomenal representing of a purple trope has a structure which is not captured by this term. One creature's purple qualia may represent purple objects *in the same way that* another creature's purple qualia represents orange objects. Both creatures may have phenomenal representings of the purple trope of an object which phenomenologically represent this same trope differently. They both represent the object as being purple because (presuming Dretske is right) that is what their representational states have the function of representing the object as, but the phenomenal character of their representings differs in the sense that the asterisk in *purple* qualia* designates something different in each case. Because what I am terming "purple qualia" can differ in this important qualitative respect, it is not purple qualia that should be taken as paradigmatic of qualia. As such, it is not purple qualia which I will identify with the phenomenal colour purple. Rather, it is that which the asterisk designates in *purple* qualia* which I will identify with phenomenal purple. Consequently, in order to specify what the asterisk designates I will use the terms such as "ph-purple" and "ph-orange" (the ph- being short for phenomenal). Thus, one of the creatures which were mentioned previously may be said to have *ph-purple purple qualia*, while he other has *ph-orange purple qualia*. Saying that one creature has *ph-orange purple qualia* is simply the same as saying that the way that its purple qualia represent purple objects is the same way that the other creature's purple qualia represent orange objects. I will continue to italicise such terms in order to signify their closer affinity to terms like "*purple* qualia*" rather than "purple

qualia". As this may be confusing, and because *these distinctions are of paramount importance to the rest of this paper*, I will recapitulate the meanings of these terms:

1. Purple: those token microstructural features of an object which humans phenomenally represent when they veridically perceive a purple object. Purple objects are objects with purple tropes.

2. Purple quale: that part of a phenomenal representing which represents an object as instantiating a purple trope.

3. *Purple* quale*: a purple quale understood as phenomenologically representing the purple trope of an object in a certain way rather than another way.

4. *Ph-purple purple quale*: a *purple* quale* which phenomenologically represents a purple trope in a way that can be characterised as phenomenal purple as opposed to, e.g. phenomenal orange. The term with the ph- prefixed to it specifies that which the asterisk in *purple* quale* stands in for, while the term that follows the prefixed term (in this case "purple quale") retains the same function as it does in 2 above.

Number 4 should not be taken as implying that someone who perceived a purple object by tokening a *ph-orange purple quale* is misperceiving the purple of the object. Remember that the purple that is perceived is the purple trope of an external object *not* the purple of a quale. Purple is veridically perceived *by* the tokening of a purple quale. What makes a quale a purple quale is simply the fact that it is a perceptual representing which represents purple. Thus, if Dretske is correct about how representational states come to represent objective properties, then a purple quale is merely a token representing which has the *function* of representing the purple of an object. In the case of properties which humans

represent (e.g. purple), ph- ascriptions such as the ph-orange in *ph-orange purple quale* can be thought of as convenient first-person descriptions of the way that a particular (purple) quale may represent a (purple) trope of an object. It is a way of saying that, perhaps, the way that your purple qualia represent purple tropes is the same way that someone else's orange qualia represent orange tropes. From your own experience, you know what way your own purple qualia represent purple and what way your own orange qualia represent orange. In this way, saying that someone veridically tokens a *ph-orange purple quale* when they look at a purple object allows the difference in inter-personal experiences to be understood in first-person terms. It is not a claim that one person's experience is more privileged than the other's in the sense that one gets to see purple as it really is. Provided that both individual's token purple qualia when they look at purple objects, *they both see purple as it really is*. Saying that a token quale is a *ph-orange purple quale* is the same as saying that it represents purple. As such, a man who has *ph-orange purple qualia* can only veridically see purple objects by tokening *ph-orange purple qualia.* His purple qualia simply are *ph-orange purple qualia*. If the man tokens *ph-purple orange qualia* when he looks at purple objects, then he is misrepresenting purple as orange.

It should be clear from what has been said that not only are purple qualia not purple (rather, they *represent* purple) but *ph-purple purple qualia* are also not purple. The only things that are purple are those objects which have those microstructural features which our purple qualia represent objects as having. *Ph-purple is a feature of our purple qualia.*

It is possible that all normal human beings (e.g. those who do not suffer from blindness or some degree of colour-blindness) see colours *with* colour qualia

which phenomenologically represent coloured objects in different ways. Ned Block (for example) says that, "given that there are enormous differences among normal people in the physiological machinery of color vision", there is no good reason to believe that the same type of red quale is produced in all of us when we look at red objects (Block: 1990, pp56-57). By this, Block means that we do not have a good reason to assume that we all have what I term the same *red* qualia*, even though we may all have red qualia. As Block points out, this is a good reason not to identify colours with phenomenal colours (*colour* qualia).*

On the other hand, it is possible that all normal human beings see colours *with* colour qualia which phenomenologically represent coloured objects in exactly the same way. In other words, it is possible that all our red qualia are *ph-red red qualia.* If this were the case, then ph-red could be defined as *the way* that people's red qualia *normally or typically* phenomenologically represent red. This would allow for a criterion of normalcy that might seem to give legitimacy to the ascription of ph-red to one creature rather than another which also has red qualia. Now, if there is such a criterion of normalcy that could be discovered (in principle), then we could have principled grounds for saying of an alien creature which is discovered to have red qualia that it has *ph-red red qualia* or, e.g., *ph-green red qualia.* Saying that it had *ph-green red qualia* would be the same as saying that it had red qualia which represented red phenomenologically in the same way that our green qualia (which could be characterised as *ph-green green qualia*) represent green. Also, if half the human population was found to have *colour* qualia* that were inverted vis-a-vis the other half of the population (e.g. ph-red was inverted with ph-green, etc.) and one half was found to have some sort of genetic or other biological defect that correlated with this inversion,

then that part which was normal could be characterised as having *ph-red red qualia*, etc. while the other half could be characterised as having *ph-green red qualia*, etc.

Remember, of course, that by describing the *red*qualia* of the abnormal half of the population as *ph-green red qualia*, I am assuming that this part of the population still have phenomenal states which *represent red.* If, on the other hand, the defect that was discovered was one that made these people token green qualia when they looked at red objects, then their qualia would be *ph-green green qualia* which were merely misrepresenting red objects. Leaving aside difficulties associated with the asymmetrical nature of the human colour quality space, we can assume that these abnormal individuals learnt to use colour words in the same way as the other half of the population. However, assuming that Dretske is right about how phenomenal states represent those properties that they do, then these individuals would be misrepresenting the colour properties of objects in the sense that states that are *supposed to* represent red would be tokened by green objects instead of red objects, etc.

Because I am *not* claiming that ph-colours are real colours, failing to establish a criterion of normalcy for our ph- ascriptions would be of no consequence for our ordinary understanding of colours as belonging to objects. The relevant microstructural features of objects are the real colours of objects. Ph-colours are merely features of our colour qualia. They are those features which phenomenologically represent colours in a certain way. They are, therefore, important for an understanding of phenomenological facts, but not for our understanding of facts about objects.

5. INVERTED SPECTRA

Because phenomenal colour is paradigmatic of qualia, most discussions of qualia involve colours. Likewise, most discussions of inverted qualia are discussions of inverted spectra. Shoemaker puts forward John Locke as the originator of the idea of inverted spectra in the philosophical literature. In *The Essay,* Locke entertained the possibility that "the same Object should produce in several Men's Minds different Ideas at the same time; for example, the Idea, that a Violet produces in one Man's Mind by his Eyes, were the same that a Marigold produced in another Man's, and vice versa" (quoted in Shoemaker, 1997: p643). As used in this passage, at least, the word "Idea" was being used in the same way that I have been using the term "*colour* qualia*". It was part of Locke's supposition that two people's phenomenal representings of colour (their colour qualia) could differ in ways which were not manifested in differences in linguistic and other behaviours, i.e. they could differ in their *colour* qualia* (where the asterisk stands in place of ph-violet and ph-marigold, respectively - and these refer to the colours of the named flowers). *This interpretation is supported by the fact that it was not part of Locke's supposition that one of them would be seeing the colour incorrectly.* Consequently, that which was supposed to differ in their "Ideas" was not their veridicality (a relational property), but some intrinsic feature of their Ideas.

In recent years, this feature of the indiscriminability of spectra inversions has been used to formulate inverted spectra objections against functionalism. For, it is argued, if a spectrum inversion cannot be detected in functionally identical creatures, then there are mental features which

functionalism cannot capture. If such inversions were detectable in linguistic and other behaviour, then the creatures involved would not actually be functionally identical. Such discussions of inverted spectra generally involve inversions of 180 degrees in the hue circle of phenomenal colour space, such that the opponent hue pair of red and green are inverted, the opponent hue pair of blue and yellow are inverted, and all the binary hues are inverted with those which lie on the opposite side of the circle from them. It is then posited that there could be creatures that phenomenally represent green objects in the same way that other creatures phenomenally represent red objects and so on. However, the human colour quality space has been recognised as being asymmetrical. This means that such inversions would be, unlike Locke's discussion presumed, detectable inversions. As such, inverted spectra discussions today generally posit inverted spectra in imaginary creatures with symmetrical colour spaces and no other functionally detectable connections between these and behaviour. Again, like Locke's discussion, they do not normally presuppose that either creature is perceiving colours incorrectly.

6. PHENOMENAL APPEARANCES

I will now look at Dretske's discussion of phenomenal appearances. This is important because Dretske goes on to discuss the issue of inverted qualia in terms of phenomenal appearances.

Dretske discusses a situation in which a child (Susan) of normal eyesight and intelligence, who has never seen a dog before and does not know what a dog looks like, first sees a dog (1995:p66). This dog is a French poodle. Dretske claims that Susan will not conceptualise what she sees as a dog, let alone a French poodle. She will not have the belief that this is what she is seeing. However, her visual experience will represent the dog as a poodle in the sense that it (the experience) will represent the dog as having *the manifest properties of poodles.* These are those properties which make poodles look different from other dogs and other objects generally. Pictures of poodles and fake poodles will also have these properties and, thus, Susan's visual system will sensuously represent these in the same way that it represents real poodles. Such things will "produce poodle qualia in Susan". Dretske, thus, believes that there is a sense in which a poodle can look like a poodle to someone who has no concept of a poodle. At the same time he acknowledges that there may be a sense of such words as "look", "seem", and "appear" in which something could not look like a poodle to someone who did not have the concept of a poodle (1995:p67). He calls this latter sense of such words the doxastic sense: looks-d, seems-d, appears-d, etc. In this sense, the poodle does not look-d like a poodle to Susan; she does not believe that she is seeing a poodle. In the other sense, the poodle looks the same to Susan as it looks to someone who does possess the concept POODLE. This, he calls the phenomenal sense: looks-p, seems-p, appears-p, etc.

According to Dretske, saying that a poodle looks-p like a poodle involves two claims (Dretske: 1995, p68):

1. It looks-p to S the way poodles normally look-p to S; and
2. It looks-p different to S from other breeds of dogs.

Number 2 is a discriminatory clause. It prevents dogs looking-p like poodles to half-blind people and to animals without the necessary visual acuity. To such people and animals, all (or, at least, many) dogs will look-p the same (Dretske 1995:p68). Dogs, therefore, provide the relevant *contrast class* for something to be said to look-p like a poodle. Dretske suggests that toads are one kind of animal to which poodles do not look-p different from other breeds of dogs. Instead, he says, they look-p to toads the way they look-p to half-blind people. *They look-p like blurry spots*; the same way that everything else that is this size and moving (toads do not seem to visually detect stationary objects) looks-p to a toad. This is because toad's cannot see the *differences in form and detail* which make poodles look different from other breeds of dogs to us (Dretske: 1995, pp69-70).

This suggests that if this is put in terms of what I have said elsewhere (chapter II.6) about how acts of phenomenal representation, though nonconceptual, can be viewed as representing purported fact, Dretske is making the point that the phenomenal representings of both the half-blind person and the toad are not as determinate as those of a fully-sighted person. The space surrounding a half-blind person or a toad (i.e. the space comprising that area that is being represented by such an individual at a particular time) could be filled with a variety of different objects that have those properties that make them look-p very different from other objects to a fully sighted person, and still be consistent with the veridicality of their original, unchanged, representings. The area around a fully sighted person could only be filled with tokens of the same type of object being represented, or with good fakes or pictures, in order for his or her representings to remain veridical.

My major difference with Dretske is most clearly seen at the level of relatively *simple represented properties* like colours and shapes, rather than at the level of more *complex represented properties* such as the property of being a poodle. Such complex properties admit of separation into a range of simple properties such as colours and shapes in much the same way that David Hume distinguished between simple and complex impressions of sensation. For instance, Hume wrote that "Tho' a particular colour, taste, and smell are qualities united together in this apple, 'tis to perceive they are not the same, but are at least distinguishable from each other" (Hume: 1985, p50). Likewise, in the case of a visual representing of a moving poodle, the representing can be separated into a range of represented properties including colours and shapes and movement.

Dretske's definition of phenomenal appearances is meant to apply to the simple represented properties as much as to the complex ones. For instance, he writes that objects cannot look-p *red* to people who are colour blind and who, therefore, cannot discriminate one colour from another (Dretske: 1995, p68). Presumably, therefore, saying that an object looks-p red to S involves the following two claims:

1. It looks-p to S the way red objects normally look-p to S; and
2. It looks-p different to S from other coloured objects.

If the indeterminateness/generality of red is deemed a problem here, I could substitute red for a very determinate shade of red (e.g. red21), but the example should be clear enough without doing so. Note, also, that the notion of an object being represented as having the manifest properties of something (e.g. a poodle) does not apply so well when we are talking about relatively simple

represented properties like red. Red *is* one of the manifest properties of ripe tomatoes, for example, but the manifest properties of red things (in virtue of which they are *red* things) consist *solely* in the property of being red. The phenomenal red (*red* quale*) of an experience of a red object may be specified in terms of the three variables of hue, saturation, and lightness, but these variables combine in the representing of the red of a red object to form a *monadic* property of the experience in a way in which the manifest properties of poodles do not combine. Consequently, they are not separable in the same way that the manifest properties of poodles are.

7. DRETSKE ON INVERTED QUALIA

Dretske notes that an object can look-p F to two different people or other animals without it looking-p the same to them (1995:p69). They could both be equally good at discriminating F's from other objects and a given object could look-p to each of them the same way such objects usually look-p to each of them. They could, however, have inverted qualia despite meeting the conditions for having something look-p a certain way to them. This is because, even if we know that the conditions for something looking-p a certain way to a particular representational system are met, this does not tell us which properties the representational system is *representing* as being instantiated and it is these represented properties that Dretske identifies with qualia. An example of a represented property is provided by that which I have termed purple qualia. In the case of purple, therefore, Dretske's qualia are to be *solely* identified with that

part of a phenomenal representing which represents an object as instantiating a purple trope.

Dretske notes that we must be careful when making inferences about appearances-p from discriminatory data (1995:p71). This is a consequence of his acknowledgment of the possibility of inverted qualia. Thus, Dretske gives an account of experience that involves phenomenal representings with a certain qualitative character (rather than just cognitive phenomena such as judgments) and acknowledges the possibility of qualia inversion. However, while conceding this much would usually be accompanied by the admission of a problem for materialistic theories of mind, Dretske does not go this far. Instead, he says that the possibility of inversion is only a problem for behaviourists and functionalists who believe that experience must be defined behaviourally or functionally (1995:p72). Dretske regards RT as succeeding here in two important respects. Firstly, it respects the intuition that the qualitative nature of experience does not necessarily express itself in behavioural or functional performance; and, secondly, it makes this qualitative nature objectively determinable. On Dretske's account, qualia are as objectively determinable as the functions of bodily organs. This leaves us, he claims, with practical but not theoretical obstacles to specifying an organism's qualia. This result is yielded from Dretske's identification of qualia with experienced properties which, in turn, are identified with the properties that are systemically represented by a representational system. These latter properties are identified with those properties that the senses have the natural function of providing information about (Dretske 1995:p72).

I believe that Dretske is wrong. I believe that the problem of inverted qualia *is* a problem for RT, and that RT *does not* make the qualitative aspects of

experience objectively determinable. ***We could know all about the functions of a particular sensory system without this telling us anything more about the qualia of the various states of this system other than that the qualia of such states are constitutive of the way in which those properties of external objects that are being represented are being phenomenologically represented.*** **Knowing what properties of an external object a mental act is representing does not tell us about the way that the act is representing these properties. If we do not know how the act is phenomenologically representing these properties then cases of qualia inversion are not averted, because it is** ***this feature of*** **experience that is inverted in problematic inverted qualia thought experiments. Put in terms of the terminology I employed earlier, knowing that an experience of purple is constituted by a purple quale does not enable us to know what the asterisk in** ***purple* quale*** **designates. For example, it might designate ph-purple or it might designate ph-orange. Consequently, we do not know how the act is representing purple and, so, cases of qualia inversion are not averted.**

8. A DRETSKEAN QUALIA INVERSION SCENARIO

I will now consider why Dretske thinks that RT can deal with inverted qualia. The essence of his case in support of this claim will be seen to reside in inverted qualia scenarios which are relatively unproblematic when compared with more traditional examples of qualia inversion. In other words, Dretske's claim that RT can adequately deal with cases of inverted qualia is mistaken.

Dretske offers three of his own examples of inverted qualia cases and argues that these tell against defining sensory qualities in terms of discriminatory

performances (1995: pp69-78). All three versions involve highly degraded quality spaces. This, Dretske contends, is to get around the problems that have been highlighted regarding disruptions to the organisation of highly structured quality spaces such as those of humans' which would contribute to differences in discriminatory powers and would, therefore, be detectable differences (1995: n7, p177).

I will consider only one of Dretske's inversion cases as they are all essentially the same in that they concern the representation of different values of the same properties. In the case which I will consider, Dretske asks us to imagine that Susan has had her vision severely impaired, such that moving objects look-p the same to her as they do to Arthur the toad. This example involves us supposing that toads' visual systems, like human visual systems, can (to a very limited degree) represent shapes of middle-sized moving objects. Dretske doesn't seem to be concerned in this example with properties other than shapes. The main difference that Dretske's example seems to be supposing to exist between normal humans and normal toads with respect to the visual representation of the shapes of middle-sized moving objects is a difference in *the values* of the shape properties which are represented (1995: p76). For example, he writes that

> toads cannot see differences that normal human beings see, *differences in form and detail* that make poodles look different from bulldogs to us.... So poodles, I infer, must look-p to toads the way poodles look to us when we see them in poor light and without our glasses -pretty much the way bulldogs and terriers look (Dretske: 1995, pp69-70).

According to Dretske, if a representational system is damaged, the information its various states carry (what they correlate with) can change, but what this information means (what it has the function of correlating with) will remain as it was before the damage occurred (1995:p77). Consequently, he

claims, that *not only can we imagine* Susan's vision becoming impaired such that moving objects look-p the same to her as they do to Arthur the toad, *but we can also imagine* visually-impaired Susan seeing-p every dog as a poodle. We can imagine her hallucinating poodles whenever she looks at dogs (1995: p74). Dretske notes that this example presupposes Susan's visual impairment is due to damage to some part of the visual system which is responsible for *delivering* information about certain external properties to those parts of the brain which process such information. The damage is not to this part of the brain itself -so it does not affect Susan's ability to dream, imagine, or hallucinate such properties (1995: n12, p178).

On the other hand, Dretske contends, we *cannot* imagine a toad (Arthur) with similar visual acuity seeing-p every dog as a poodle. Susan has (hallucinatory) sensory states that mean that the manifest (shape) properties of poodles are being instantiated in an object being looked at even when these states are no longer carrying this information. Arthur's sensory states *could be carrying the same information* as Susan's are carrying now that her vision is impaired, but his states do not mean that the manifest (shape) properties of poodles are being instantiated in an object being looked at. This is because Arthur has no sensory states that are *supposed to* carry this information.

What it means for Arthur's states to be carrying the same information as Susan's is that they both have states which *correlate with* moving objects which instantiate the manifest (shape) properties of dogs. For instance, when they look at moving dogs they might both enter tokens of the types of states which they enter whenever they look at moving dogs, such that these states correlate with the presence of moving dog shapes. Now, if they are both looking at *the same dog*

(say, a moving, American Pitbull Terrier), then they will both enter states of the type that correlate with the presence of moving *dogs*. Now, *neither* Susan's nor Arthur's sensory states *are* carrying *the information that* (=*indicating that*) the manifest (shape) properties of *poodles* are being instantiated in an object. This information could not be carried by *any states* because these properties are *not* being instantiated.

However, because the type of state that Susan enters when she sees a dog is the type of state that has the *function* of representing the manifest (shape) properties of poodles, Susan has (hallucinatory) sensory states that mean that the manifest (shape) properties of poodles are being instantiated in an object being looked at even when these states are no longer carrying this information. Consequently, Susan's state has poodle shape qualia while Arthur's does not. This is because, according to Dretske, qualia are represented properties and those properties which Susan is (nonveridically) representing are the manifest (shape) properties of poodles.

This inverted qualia scenario is used by Dretske to show that two representational systems can be functionally equivalent (equivalent in discriminatory powers and capacities) even though they are occupying different representational states. Neither Arthur nor visually-impaired Susan can distinguish poodles from other dogs; and all dogs look-p the same to Arthur, and they all look-p the same to Susan. However, dogs look-p different to Arthur than they do to Susan. The states that Susan and Arthur are occupying when looking at a moving dog have different indicator functions; when functioning properly they are supposed to indicate *different values of shape properties.* Because qualia are given by indicator functions, their states have different qualia. Dretske says

that this puts him in agreement with Block and Fodor (1972) and Shoemaker (1975) that qualia are not functionally definable (Dretske: 1995, pp77-78). Experiences can be different without this difference being expressed in discriminatory performances. He does insist, though, that qualia are still physically definable as long as there is a physical description of the conditions in which systems have indicator functions (1995: p78). For instance, if this is the case, then the qualia of Susan's hallucinatory states could be given by a specification of the values of shape properties that these states were representing as being instantiated, which would be different from the values of shape properties that Arthur's states were representing as being instantiated.

I would argue, however, that the fact that two systems occupy different representational states (states with different indicator functions), despite being the same in discriminatory performance, tells us relatively little about the qualia of these states. What Dretske gives us here is an argument to the effect that the properties that a state is supposed to indicate (the property qualia of the state) are not definable in terms of discriminatory powers and capacities. Now, *qualia can be partially specified* in terms of the properties that they represent. However, as I have shown, there is another feature of qualia which is not captured by such specifications and it is this feature which generally provides the problem in inverted qualia scenarios. This feature (= the way that a property is phenomenologically represented) is the major defining feature of a quale; and, about this, Dretske's discussion is silent. Thus, even if Dretske's example can be used to provide an argument to the effect that represented properties (property qualia) are not definable in terms of discriminatory powers and capacities, it still tells us little about those features of the qualia which do the representing. Only if

we accept Dretske's assertion that qualia are to be *solely* identified with those properties that it is the function of a representational state to represent, does it follow that qualia are physically definable in terms of indicator functions.

9. PROBLEMATIC INVERTED QUALIA SCENARIOS FOR RT

A different type of inversion scenario that Dretske does not discuss in *Naturalizing the Mind*, but which he would allow as a possibility is provided by the following. Suppose that there are two creatures (Y and Z) which have symmetrical colour quality spaces which are inverted vis-a-vis one another. They represent the same properties and the same values of these properties *differently*. When Y and Z look at a red object, they both meet the two conditions for having it look-p red to them. The object looks-p to Y the way red objects normally look-p to Y and it looks-p different to Y from other coloured objects. The same is true with regard to Z. However, the object still looks-p different to Y than it does to Z. They have inverted qualia despite meeting the conditions for the object looking-p red to both of them. Dretske would explain this by saying that at least one of these two creatures is *misrepresenting* the object. For instance, Y may be representing the object as red while Z is representing it as green. Both creatures do, indeed, have inverted qualia, but this can be accommodated by RT. When Y looks at the object a red quale is tokened while a green quale is tokened by Z. Z is misrepresenting the object in the sense that a quale which is *supposed to* represent green is tokened when looking at red.

Now, it is easy to imagine that Z's visual system was studied and that it was found that states in Z that are supposed to represent green were being

tokened when Z looked at red objects. In such a situation, Dretske's explanation would be correct. But it is just as easily imaginable that, for example, Y and Z are Dretskean representational duplicates with *different evolutionary histories.* Presuming that systemic functions are acquired via evolutionary history, both Y and Z could have colour-representing states (colour qualia) which differ in that feature which is designated by the asterisk in *colour* qualia.* Z's ph-green states may have acquired the function of representing red, just as Y's ph-red states may have acquired the function of representing red. As such, when Z looks at a red object a *ph-green red quale* is tokened while when Y looks at a red object a *ph-red red quale* is tokened. This would be a scenario that RT could not explain, as the salient mental difference between Y and Z is not a difference in indicator functions. Y and Z both represent the *same* colour properties *and* the *same* values of colour properties, but in *different* ways.

We can imagine such a situation arising by imagining a stage of evolutionary history when some of Z's previously colour-blind ancestors gained mutations which allowed them to go into ph-green states when they saw red objects. Such states would be indicators of red. Consequently, these ph-green states could be selected to *represent red.* They could *not* be selected to represent green because, as Dretske notes

> **an item cannot be naturally selected to do X unless it actually does X. It has to do X because the way it gets selected is by having its performance of X contribute in some way to the survival and reproductive success of the animals in which it occurs. It is this contribution to reproductive success that, when it is selected for, confers a function on a system, and the function it confers is doing what the system *did*... that increased fitness (Dretske: 1995, p165).**

Likewise, saying that Z's ph-green states misrepresented red would not make sense. They could only misrepresent red if they were selected to represent something else (e.g. green), for it is selection which confers the functions that Dretske requires for representation. Because they were selected to represent red, red *is* what they represent: they are red qualia. They merely represent red in a way that is different from the way that Y's red qualia represent red. Z's red qualia are *ph-green red qualia*, while Y's red qualia are *ph-red red qualia.*

This scenario involves two creatures with phenomenal representings that have the *same* indicator functions (and, therefore, the *same* colour qualia) but *different colour* qualia*. Of course, they both have *red* qualia*, but in both cases the asterisk designates something different. It designates that which I termed ph-red and ph-green, respectively. The first creature has *ph-red red qualia* while the second creature has *ph-green red qualia*. Such cases severely undermine the case which Dretske makes for the identity of qualia (solely) with represented properties, as the experiences of the creatures have qualitative features which can *vary independently of the properties that they represent.*

Another way in which this same point can be made is by positing two creatures which are looking at the same object and which are having phenomenal representings with *different* indicator functions but with the *same* qualitative features. Such creatures could both be phenomenally representing different properties of the same object, but phenomenologically representing them in the same way. An example of this sort of case is provided by Paul Churchland. Churchland asks us to imagine that there are alien beings that can perceive the temperature of an object by looking at it rather than by feeling it as we do (Churchland: 1987, pp8-10). These beings have eyes with retinas

consisting of rods which are sensitive to electromagnetic radiation at a wavelength in the far infrared. Because such radiation provides a reliable indicator of an object's temperature, images of it form on the creatures' eyes with differing degrees of 'brightness' in accordance with how hot an object is, and these creatures enter sensory states which represent objects as hot, warm, or cold. Churchland also asks us to imagine that the sensory events of these beings phenomenally represent the world such that hot objects look how we would describe as white, warm objects look how we would describe as grey and cold objects look how we would describe as black. Now, let us assume that a person and one of these beings are looking at an object (the same object) that is situated in front of them both. In order to keep the example simple, let us suppose that each being is focussing on only one of the properties (a different property in each case) of the object. The human is looking at the objective colour of the object, while the alien is looking at its temperature. The human is phenomenally (i.e. nonconceptually) representing the colour of the object phenomenologically in a way that we would identify as white, and the alien is phenomenally representing the temperature of the object phenomenologically in the same way. Thus, we have a situation in which qualia and indicator functions cannot be identified. They cannot be identified because both beings have sensory states which have different indicator functions and which are, therefore, phenomenally representing different properties of the same object. According to RT, the human's sensory state is characterised by a colour quale because it is representing the object's colour and the alien's state is characterised by a temperature quale because it is representing the object's temperature, and that's

all there is to the matter. However, there is an additional mental fact here which RT cannot capture. Namely, the fact that both beings are phenomenologically *representing different properties in the same way.* One's experience is characterised by a *colour* quale* rather than just by a colour quale, while the other's experience is characterised by a *temperature* quale* rather than just by a temperature quale. As such, there is an important sense in which the qualia of both of their states is the same. In this case, the asterisks attached to "colour" and to "temperature" designate the same feature of these differently taxonomised qualia. The human's experience can be specified as being characterised by a *ph-white colour quale* while the alien's experience can be specified as being characterised by a *ph-white temperature quale.*

With the above examples, I have attacked the identification of qualia solely with indicated properties in two different ways. First, I posited creatures with states that have the same indicator functions but different qualia. Next, I posited creatures with states that have different indicator functions but the same qualia. The main conclusion to be drawn from the discussion of these cases is that while the notion of represented properties is an important part of what qualia are, there is an important and defining *feature of qualia* which is not captured by this notion alone. Consequently, there is an important feature of our experiences which eludes RT.

10. DRETSKE'S DEFENCE OF HIS ACCOUNT OF QUALIA

It might be thought that Dretske may just be making a different use of the word "qualia" than the one that I am making. His identification of qualia

with systemically indicated properties may be purely stipulative. However, Dretske cannot make this identity stipulative, because he is interested in addressing existing philosophical issues and problems of qualia. He wants to show that RT can enable progress to be made here (1995:ppxiv-xv). Creating another use for the word which *does not* (as I believe that my usage *does*) capture the traditional problems of qualia would simply amount to changing the subject. It is, of course, possible for Dretske to argue that his usage is (contrary to what I have said) actually in line with the more usual usage, and that he *does* address traditional problems of qualia. In fact, this is what Dretske does. He puts forward an argument which, if successful, would refute what I have just said about the inadequacy of his account of what qualia.

Dretske produces the following argument (A) out of what he says are two facts that almost everyone accepts:

> The first fact is that qualia are supposed to be the way things seem or appear in the sense modality in question.... [I]f a tomato looks red and round to S, then redness and roundness are the qualia of S's visual experience of the tomato. If this is so, then (second fact) if things ever are the way they seem, it follows that qualia, the properties that define what it is like to have that experience, are exactly the properties the object being perceived has whenthe perception is veridical (Dretske: 1995, pp83-84).

Just in case the second fact does not appear to be very clear in the above passage, it is simply the fact that things are sometimes the way they seem to be. Taken as a whole, the meaning of this argument is that even the qualia of a nonveridical perception are given by the veridical case, i.e. by what the qualia would be if the perception were veridical. Dretske applies this result to the case of an imagined parasite with an acute thermal sensory system that attaches itself

to a host when, and only when, it senses the host as being 18 C (1995: pp82-84). He asks "what is it like to be this parasite when it senses a receptive host?" His own reply is that, if you know what it is to be 18 C, you know how the host "feels" to the parasite; what the parasite's experience is like (1995: p83).

In line with what I have said about qualia, I would reply that this is wrong. Knowing that this parasite senses hosts as being 18 C and knowing what it is to be 18 C *does not* tell us how the host feels to the parasite. Provided that there *is* some way that the host feels to the parasite, this information is not sufficient to settle the question of *how* the host feels. Maybe the way that the parasite's phenomenal representing of 18 C phenomenologically represents an 18 C host is similar to the way that our hands would feel if they were immersed in a vessel of water of this temperature immediately after they have been immersed in water of 5 C. Maybe it is similar to the way that our hands would feel if they were immersed in this same vessel of water immediately after they had been immersed in water of 35 C (cf. Berkeley: 1996, pp115-116). Or maybe, like Churchland's creatures, this parasite phenomenologically represents 18 C in the same way that we visually represent things that are coloured grey. Perhaps, even, it phenomenologically represents 18 C in a way that is not comparable with *any* of our experiences. In other words, we may know that the parasite has states with *temperature* qualia* but we do not know what the asterisk signifies in this case, and this is a defining feature of its qualia. The parasite's experiences may be characterised as having *ph-hot temperature qualia* or as having *ph-cold temperature qualia* or as having *ph-grey temperature qualia*, etc.

Dretske takes A to be independent of his RT (1995:pp83-84). He says that if the result of A (i.e. what it has to say about such cases as that of the parasite

above) is absurd, then it is one of the two facts that led to it that is to blame (1995:p84). While I do not see anything wrong with either of these two facts I believe, however, that A gives us an incomplete account of qualia precisely because it does not succeed in drawing out the full detail which makes the first fact widely acceptable by those who believe that there are qualia. Consequently, A gives us an incomplete account of qualia. Nevertheless, it is the account of qualia which this argument gives us that Dretske's RT seems most suited to. If we are after a theory of mind that could deal with this account of qualia, then RT can certainly do the job; but RT cannot deal with the more usual account of qualia. If I am right in this claim, then RT will prove to be an incomplete theory of mind.

The problem with A resides, as I have said, in Dretske's interpretation of the first fact. It is true that most defenders of the idea that there is such a thing as qualia and that it is what constitutes the qualitative character of experience, use "qualia" or "the qualitative character of experience" to denote the way things seem or appear to S in modality M. But "seems" and "appears" are used in this context in a way that is not fully captured by Dretske's account of qualia as represented properties (=property qualia).

While a tomato might look-p to S the way tomatoes normally look-p to S, and look-p different to S from other fruits and vegetables, the manifest properties of tomatoes (e.g. redness and roundness) do not (alone) constitute the qualia of the visual experience of a tomato. The specification of these properties as the property qualia of the experience only provides a very coarse-grained description of the qualia of the experience; it tells us which properties are being represented by the qualia of the representational act, but it does not tell us

anything about the way that the redness and the roundness of the tomato are represented by the qualia of the phenomenal act of representing. In other words, it tells us that the experience is constituted by red qualia and round qualia, but it does not tell us anything about that feature of the qualia which is characterised by the asterisks in *red* qualia* and *round* qualia.* This much finer-grained (and conceptually distinct) detail is missing.

This feature of a visual phenomenal event is *responsible for* the redness of the tomato looking-p to S the way red objects normally look-p to S and looking-p different to S from other differently coloured objects. It is, therefore, this feature that is required to be specified *if* the way that the tomato visually seems or appears is to be fully specified. It is this feature that I have referred to with the expression "the way that those properties that are phenomenally represented are phenomenologically represented". Likewise, I contend, it is this feature that philosophers generally intend to pick out (or, more precisely, this feature is intended to be included in that which is picked out) when they say (as Dretske did above) that qualia are supposed to be the way things seem or appear in the sense modality in question. *A represented object such as a tomato not only seems to be red, it also seems to be red in the same way that red is phenomenologically represented by the experience of the tomato.* However, this is a feature of a representing, rather than of the object or its particular colour trope.

In order to see that this is the case, it is worth considering that the following statements make sense (with regard to an object O that is red) to such people:

1. "O (veridically) appears phenomenally-red to subject S1 in M at time T1 when viewed under standard conditions"; and this is not incompatible with

2."O (veridically) appears phenomenally-green to S2 in M at T1 when viewed under standard conditions" (intersubjective qualia inversion); nor is it incompatible with

3."O (veridically) appears phenomenally-green to S1 in M at time T2 when viewed under standard conditions" (intrasubjective qualia inversion).

On my account of qualia we can make sense of these statements by supposing that "phenomenally-red" and "phenomenally-green" are just ways of characterising a feature of the phenomenal representing of objective red, whereby different creatures can both phenomenally represent objective red veridically, but in different ways. Both individuals can, while standing in causal/contextual relation C to a red object, have nonconceptual sensory representings that have the function of representing red, but which phenomenologically represent red differently.

On the other hand, the above statements just do not make sense if the qualia of an experience are (as Dretske contends) the properties the object being perceived has when the perception is veridical. On this account, both phenomenal-red and phenomenal-green would have to be properties of one and the same object (a red object) that can only be (veridically) perceived by a subject as being one homogeneous colour. Likewise, for all other cases of qualia inversion. As such, we must either impute contradictory properties to objects (e.g. the property of being *both phenomenally red and not phenomenally red*, i.e. phenomenally green); or we must deny that cases of inverted qualia such as those above can ever be veridical. *Dretske (in conversation) has, in fact, made the latter claim.* This manoeuvre can enable him to make sense of the above statements so long as the word "veridically" is dropped from 2 and 3, and the subjects in 2 and

3 are taken to be *misrepresenting* red. Of course, if the word is not dropped, the statements will (under this interpretation) still make sense. Unlike the first interpretation which imputed contradictory properties to objects, these claims will not appear incoherent; they will just appear to be mistaken.

11. DRETSKE'S CLAIM OF MISREPRESENTATION

Dretske contends that we cannot have cases of inverted qualia of the type that I have posited above in 2 and 3 (chapter IV. 10). That *it is not the case* that at least one creature *must* be misperceiving in inversion scenarios such as that in 2 is, I believe, clear from what I have said about the possibility that ph-green states could be recruited by selection processes and given the function of representing red tropes (chapter IV.9). At least, this is the case with regard to intersubjective inversion scenarios. In the case of intrasubjective inversion scenarios (number 3 above), I would have to produce further arguments to establish this same conclusion. For instance, there could be a creature which tokened ph-red states when it looked at red objects and ph-green states when it looked at green objects. Now, if these states had the systemic function of representing red and green, respectively, then if it started tokening ph-green states when it looked at red objects and ph-red states when it looked at green objects, it would be tokening states which represented different properties than those which were actually instantiated. It would, therefore, be misrepresenting. In order to get a situation in which such an intrasubjective inversion could occur without the creature misrepresenting, I would have to *either* argue against Dretske's contention that these states have functions which are fixed by historical

factors, *or* argue that the ph-red and ph-green features of representational states can be inverted without inverting the functions of the states which currently possess these features. I am, however, happy to accept Dretske's contention about states having functions which are fixed by historical factors. I would, therefore, be more inclined to take the other option. It is this option which was implied in chapter IV.4 when I suggested that a genetic or other biological abnormality among a portion of the population might enable us to characterise that portion of the population as having *ph-green* *red qualia*, etc. while the normal half could be characterised as having *ph-red* *red qualia*, etc. However, lack of space prevents me from exploring this option. Consequently, I will make no claims here as to the possibility of intrasubjective inversion scenarios which do not involve misrepresentation. I am content to leave this issue unexplored as I do not need to establish such possibilities in order to successfully refute RT. Intersubjective inversion cases such as those which I have already discussed are enough for this purpose. They are enough because they provide examples of situations in which there are mental facts that cannot be equated with facts about informational functions.

CHAPTER V

CONCLUSION

I set out to refute RT by refuting the claim that all mental facts are facts about informational functions. I believe that the arguments that I have provided are enough to achieve this result. The case of Dretskean Mary highlighted mental facts which are not facts about informational functions. One set of such facts (i.e. facts about phenomenal similarity relations between experiences of a specific range of objective properties) was found, upon further examination, to potentially be facts about informational functions. Whether or not they are is an open question which can only be settled empirically. It depends upon whether our quality spaces are Quinean or Tyean spaces. The other set of facts, however, were shown to be facts that are not facts about informational functions. Rather, such facts are facts about the qualia of mental acts of phenomenal representation.

They are facts about the way that experiences phenomenologically represent those external properties that they represent. In other words, they are facts about qualitative features of mental events which cannot be captured by an account of informational functions alone.

This was shown by way of a number of examples of situations in which different creatures were shown to have mental states that could be characterised by features which eluded RT. Furthermore, these were features that could be understood precisely because they are features that we are all familiar with from the first-person perspective of conscious experience. That RT could not account

for these features is because RT suffers from the same problem as functionalism does. RT cannot provide a satisfactory definition of an individual quale which would enable it to deal with inverted qualia scenarios of the type that I showed were problematic for it. Dretske's proposed definition of a quale as a represented property was consistent with RT but it did not enable him to do the job that he intended. It did not enable him to deal satisfactorily with traditional philosophical problems of qualia. These problems are connected with a notion of qualia that is better captured by my proposed notion of a quale as *an intrinsically mental property that represents a token trope of a certain type in a certain way rather than another possible way.*

ACKNOWLEDGMENTS

I am grateful for the support of the Philosophy Department of the University of Canterbury and, in particular, for that of my thesis supervisor Derek Browne. I am also grateful to Fred Dretske for the opportunity to discuss some of the ideas contained in this thesis.

REFERENCES

Bach, K. 1997. 'Engineering the Mind', Philosophy and Phenomenological Research, 62, no.2: pp. 459-468.

Bennett, J. 1965. 'Substance, Reality and Primary Qualities', American Philosophical Quarterly, 2, no.1: pp. 1-17.

Berkeley, G. 1996. Principles of Human Knowledge and Three Dialogues, Robinson, H. (ed.), Oxford: Oxford University Press.

Block, N. & Fodor, J. 1972. 'What Psychological States Are Not', The Philosophical Review, 81, pp. 159-181.

Block, N. 1980. 'Are Absent Qualia Impossible?', Philosophical Review, 89, pp. 257-274.

Block, N. 1990. 'Inverted Earth', in Tomberlin, J. E. 1990.

Block, N., Flanagan, O., & Guzeldere, G. (eds) 1997. The Nature of Consciousness: Philosophical Debates, Cambridge, MA: MIT Press.

Browne, D. 1996. 'Phenomenology Ain't in the Head', unpublished.

Chalmers, D. J. 1996. Consciousness and Experience, Cambridge, MA: MIT Press.

Chisholm, R. M. 1957. Perceiving: A Philosophical Study, New York: Cornell University Press.

Churchland, P. M. 1987. Scientific Realism and the Plasticity of Mind, Cambridge: Cambridge University Press.

Dennett, D. 1988. 'Quining Qualia', in Marcel, A. J. & Bisiach, E. (eds), Consciousness in Contemporary Science, Oxford: Clarendon Press.

Dennett, D. 1991. Consciousness Explained, London: Penguin Books.

Dretske, F. 1981. Knowledge and the Flow of Information, London: Basil Blackwell.

Dretske, F. 1986. 'Misrepresentation', in Bogdan, R. J. (ed.), Belief, Oxford: Oxford University Press.

Dretske, F. 1988. Explaining Behavior, Cambridge, MA: MIT Press.

Dretske, F. 1990. 'Seeing, Believing and Knowing', in Osherson, D., Kosslyn, S., & Hollerbach, J. (eds) An Introduction to Cognitive Science, Volume 2, Visual Cognition and Action, Cambridge, MA: MIT Press.

Dretske, F. 1993. 'Conscious Experience', Mind, 102, no.406, pp. 263-283.

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

Dretske, F. forthcoming. 'The Mind's Awareness of Itself', Philosophical Studies.

Evans, G. 1982. The Varieties of Reference, Oxford: Clarendon Press.

Frege, G. 1993. 'On Sense and Reference', in Moore, A. (ed.), 1993.

Godfrey-Smith, P. 1994. 'A Modern History Theory of Functions', Nous, 28, no.3: pp. 344-362.

Hardin, C. L. 1996. Review of 'Colour Vision: A Study in Cognitive Science and the Philosophy of Perception' by Evan Thompson, The British Journal for the Philosophy of Science, 47, pp. 339-343.

Harman, G. 1990. 'The Intrinsic Quality of Experience', in Tomberlin, J. E. 1990.

Hume, D. 1985. A Treatise of Human Nature, London: Penguin Books.

Jackson, F. 1982. 'Epiphenomenal Qualia', Philosophical Quarterly, 32, pp. 127-136.

Jackson, F. 1997. 'What Mary Didn't Know', in Block, N., et al 1997.

Kripke, S. 1980. Naming and Necessity, Oxford: Basil Blackwell.

Levin, J. 1997. 'Consciousness Disputed', The British Journal for the Philosophy of Science, 48, no.1: pp. 91-107.

Lycan, W. G. Consciousness and Experience, Cambridge, MA: MIT Press.

McGinn, C. 1997. 'Missing the Mind: Consciousness in the Swamps', Nous, 31, no.4: pp. 528-537.

Moore, A. (ed.) 1993. Meaning and Reference, Oxford: Oxford University Press.

Moore, G. E. 1948. 'The Refutation of Idealism' in his Philosophical Studies, London: Routledge & Kegan Paul Ltd.

Nagel, T. 1997. 'What Is It Like to Be a Bat?', in Block, N., et al 1997.

Peacocke, C. 1983. Sense and Content, Oxford: Clarendon Press.

Peacocke, C. 1992. A Study of Concepts, Cambridge, MA: MIT Press.

Proudfoot, D. 1998. 'On Wittgenstein On Representation, Part I', Philosophy Research Papers, no.3, University of Canterbury.

Putnam, H. 1993. 'Meaning and Reference', in Moore, A. (ed.), 1993.

Quine, W. V. O. 1960. Word and Object, Cambridge, MA: MIT.

Robinson, H. 1994. Perception, London: Routledge.

Seager, W. 1997. 'Critical Notice: 'Naturalizing the Mind' by Fred Dretske', Canadian Journal of Philosophy, 27, no.1: pp. 83-109.

Searle, J. R. 1983. Intentionality: An Essay in the Philosophy of Mind, Cambridge, MA: Cambridge University Press.

Searle, J. R. 1992. The Rediscovery of the Mind, Cambridge, MA: MIT Press.

Shoemaker, S. 1975. 'Functionalism and Qualia', Philosophical Studies, 27, pp. 292-315.

Shoemaker, S. 1981. 'Absent Qualia Are Impossible -A Reply to Block', Philosophical Review, 90, pp. 581-599.

Shoemaker, S. 1990. 'Qualities and Qualia: What's in the Mind?', Philosophy and Phenomenological Research, 50, supplement: pp. 109-131.

Shoemaker, S. 1991. 'Qualia and Consciousness', Mind, 100, no.1: pp. 507-524.

Shoemaker, S. 1994a. 'Phenomenal Character', Nous, 28, no.1: pp. 21-38.

Shoemaker, S. 1994b. 'Self Knowledge and 'Inner Sense', The Royce Lectures' Philosophy and Phenomenological Research, 54, no.2: pp. 249-314.

Shoemaker, S. 1997. 'The Inverted Spectrum', in Block, N., et al 1997.

Tomberlin, J. E. (ed.) 1990. Philosophical Perspectives, 4: Action Theory and Philosophy of Mind, Atascadero, CA: Ridgeview Publishing Company.

Tye, M. 1995. Ten Problems of Consciousness, Cambridge, MA: MIT Press.

Wittgenstein, L. 1994. The Wittgenstein Reader, Kenny, A. (ed.), Oxford: Basil Blackwell.