

# Phenomenality, What It Does, and What That Means

Sander Voerman 2006-04-28

In this essay I will present and motivate a number of ideas about phenomenality. In particular, I will focus on two questions: what is its function, and how can we account for it metaphysically in the light of that function? In short, what does phenomenality do and what does that mean? And of course, by asking these questions I already assume that it *has* a function in the first place. But does it? Could we not plausibly defend that phenomenality does not do anything at all? I will discuss this matter as well.

Before we can begin to answer these questions, I believe we should distinguish two dimensions or orders of phenomenality. The first is that of *phenomenal quality*. This dimension involves many further dimensions, corresponding to different sense modalities and different features of experience within those modalities. Thus it involves a multitude of properties, for which I shall not be afraid to use the dreaded word “qualia”.

Qualia are the “qualitative or phenomenal features of sense-experience, in virtue of having which they resemble and differ from each other, qualitatively, in the ways they do” (Shoemaker 1982 [1997: 648]), or as Brian Loar puts it, echoing a famous phrase from Thomas Nagel, the “ways experiences differ and resemble each other with respect to *what it is like to have them*” (Loar 1990 [1997: 598]). My understanding of qualia involves submitting to *internalism*: the view that there are qualities of experience that are *not* qualities of the *objects* of experience. Qualia, then, are the qualities of experience. For example, colour qualia are not surface reflectance properties of objects, but the ways colours look to the experiencer.

The second dimension, or order, of phenomenality is that of *phenomenal affectivity*. Affectivity does not comprise *what* an experience is like, but *how much we like or dislike* what it is like. In this sense affectivity is *second order* phenomenality because it operates on first order phenomenal quality. Affectivity is not or not just about making judgements however, it is really about *how it feels* to like or dislike, to be drawn towards something or to be repelled by it. Examples of affective states are desire, pleasure, pain, fear and disgust.

The distinction between quality and affectivity is brought out neatly in an example from Dennett, despite his subsequent arguments *against* such a distinction. Dennett

introduces Chase and Sanborn, two coffee tasters employed by Maxwell House in order to verify that Maxwell coffee continues to have the right taste (Dennett 1988 [1997: 625]). Initially, both love the coffee, but after a while, both start to dislike it. According to Chase, the coffee still tastes the same to him, but he no longer *likes* that taste. His own taste has changed because he has become a more sophisticated coffee drinker. Sanborn however claims the coffee tastes different to him, and that he still would like the way it tasted before. Since all other coffee tasters agree the taste hasn't changed, Sanborn then concludes something must be wrong with his "tasters". In other words, Chase reports a change in phenomenal affectivity whereas Sanborn reports a change in phenomenal quality.

Let us assume for now, *pace* Dennett, that this distinction makes sense. In section 1 I will address the question of the function of phenomenality with respect to its qualitative dimension. Then, in section 2, we will look at the function of phenomenal affectivity. In sections 3 and 4 I will address two important lines of criticism against functionalism. In section 3 we return to Dennett and his eliminativism. And in section 4 I will discuss epiphenomenalism, the view that phenomenality exists but that it does not have a function. Specifically, I will argue against the epiphenomenalist dualism of David Chalmers. However, as we shall see, my argument has consequences for reductionism as well, and reductionism and dualism will turn out to be equally problematic. In the next two sections, we are going to turn to the philosophy of *pragmatism* to find a way out of this dilemma. In section 5 we are going to do some philosophy of science and discuss the problem of *underdetermination*. In section 6 we shall see that this problem is related to the problem of phenomenality, and we will evaluate the idealism of the classical pragmatist Charles Sanders Peirce and the reductionism of the contemporary pragmatist Herman de Regt. Finally, I will present my own metaphysical view, which I have christened *affective monism*, in section 7.

## 1 The Function of Qualia

The concept of qualia has provoked much discussion because many have held it to be incompatible with functionalism. From their analysis of functionalism they conclude that any functionalist account cannot help but leave out qualia. This approach starts out with the question "what is functionalism?" in order to answer the question "do qualia have a function?" I would like to adopt the reverse approach. I will simply start with the question

“what is the function of qualia?” and give the most intuitive answer I can think of. This answer will be the starting point for my version of functionalism.

### *What Is the Function of Qualia?*

To me the function of qualia is fairly obvious. Qualia help us recognise and classify things. When I encounter a wasp, I experience the wasp in a certain way, namely as a small flying thing that has black and yellow stripes on its back and that generates a buzzing sound. Every time I encounter a wasp I am treated to a similar kind of experience, a similar combination of visual and auditory qualia. It is by recognising these qualia that I recognise wasps.

The externalist might object that this has nothing to do with qualia because the wasp really *is* a small flying thing with black and yellow stripes making a buzzing sound, so that I am only recognising properties of the wasp. However, the internalist can make a good case out of this example. The wasp looks small to me, but we may expect it to look big to animals that are even smaller than wasps, should there be such animals that enjoy consciousness. The wasp does have the objective property of measuring certain dimensions of size in millimetres (when measured from the perspective of someone not at relativistic speeds relative to the wasp), but to offer this as an argument against the notion of size qualia would beg the question against the internalist, because the latter can maintain that it is only through size qualia that we can relate to measurements expressed in millimetres in the first place.

The same goes for the colours of the wasp. The wasp looks yellow and black to me, but it doesn't look like that to animals that cannot see colours or see colours differently. Of course, the wasp really is yellow and black if by that you mean to refer to its surface reflectance properties, and if it wasn't for these properties I wouldn't be able to recognise wasps by their colours, but the point is that it is because I recognise the *colour qualia* that I recognise the wasp.

There is an important difference between this function of qualia and what epiphenomenalists have traditionally required of functionalists in order to functionalise qualia. The function I just carved out for qualia is compatible with the inverted spectrum thought experiment. Suppose wasps would have looked blue and black instead of yellow and black from the first time I encountered them and consistently ever since. In that case I would have recognised wasps by recognising the way they look blue to me in a way that was

functionally similar to the way I actually recognise them as looking yellow. Whether it is the blue quale or the yellow quale that fulfils the role of giving wasps a recognisable colour in the way I experience them seems entirely arbitrary from a functional perspective. All that is needed for recognition is that the colour of wasps in my experience is consistent through time.

Does this arbitrariness constitute a reason for rejecting functionalism or for claiming that qualia are epiphenomenal? In my view, the fact that the blue quale *could* have fulfilled a function that is *actually* being fulfilled by the yellow quale hardly disproves the fact that it is actually being fulfilled by the yellow quale. What it does show is that the roles that qualia fulfil are *multiply realisable*: within different qualitative spectra, different qualia can fulfil the same role. But multiple realizability is compatible with functionalism, so we can happily accept this implication. We shall return to the matter of epiphenomenalism in section 4.

### *Overview of the Theory*

Intuitively we have established that qualia fulfil a role in the way we recognise the things we are familiar with in every day life. I shall now introduce my terminology and present an overview of the theory I wish to defend.

We are acquainted with all sorts of objects – people, tools, furniture, wasps - through practical interaction with those objects. Of these objects we have *practical concepts*. A practical concept is a concept that captures its content in terms of *practical bearing*, an idea that I borrow from the classic pragmatists. *Recognition* of an object as belonging to a type of objects *T* involves establishing a relation between a *representation* of the object and a practical concept of the property that constitutes *T*.

Conscious agents are capable of having *phenomenal experiences* of objects, which are representations of those objects *in terms of qualia*. Qualia realise the role of being the *syntactical properties* of these representations. Conscious agents possess *phenomenal concepts* of the qualia they experience. These concepts are related to their *practical* concepts in such a way that they can translate the practical bearing of a practical concept in terms of the qualia that feature in the way the agent would experience that practical bearing.

The upshot of all this is that conscious agents can recognise and classify objects by relating their practical concepts to their experiences *on the basis of the way those practical*

*concepts are related to their phenomenal concepts* which map onto the qualia of those experiences. Let us now examine the various aspects of this theory in more detail.

### *Phenomenal Concepts*

A concept is a semantic entity whose referent is a property. In order to believe that an object has a property one has to have a concept of that property. That pretty much fixes the *concept of concept* for my present purposes. Qualia are those properties of experience that make up its phenomenal quality. Phenomenal concepts are concepts *about* those properties. In order to *know what a quale is like*, one has to have a phenomenal concept of that quale.

In view of our distinction between phenomenal quality and phenomenal affectivity, it makes sense to make a similar distinction between phenomenal *qualitative* concepts and phenomenal *affective* concepts. The former refer to qualia whereas the latter refer to affective features. However, when it is unlikely to produce confusion I will simply speak of phenomenal concepts even when I only speak about one of the two categories.

The concept of phenomenal concepts raises issues of reductionism. Slightly rephrasing Loar we should distinguish between two reductive claims: (1) that qualia can be identified with properties of which we have nonphenomenal concepts, and (2) that the meaning of phenomenal concepts can be explained in terms of nonphenomenal concepts. This yields three positions: *strong reductionism* accepts both claims, *weak reductionism* accepts (1) and denies (2), and *property dualism* denies both. These positions can vary further in terms of what one considers to be the relevant nonphenomenal concepts. The usual candidates are *physical* concepts and *functional* concepts. Loar takes these together into his notion of “physical-functional concepts”. Thus Loar’s physicalism, which holds that phenomenal concepts are irreducible to physical-functional concepts yet refer to physical-functional properties of the brain, is a version of weak reductionism.

Loar accepts the falsity of (2) on the grounds of intuition. I share this intuition. I think the most vivid way of bringing it out is still that of Thomas Nagel. No matter how many and accurate nonphenomenal concepts we gather on the cognition of bats, we cannot on the basis of those concepts know what it is like to be a bat. Bat experiences may feature qualia that are totally alien to human experience, and since phenomenal concepts are “formed from one’s

own case" (Loar 1990 [1997: 598]), we cannot attain phenomenal concepts about qualia that we are incapable of experiencing.

What remains unresolved is the dispute between weak reductionism and property dualism on the truth of (1). We will deal with that issue later on in this essay.

### *Practical Concepts*

Instead of with physical or functional concepts, I shall contrast phenomenal concepts with what I call *practical concepts*. Practical concepts capture the behaviour of objects from a third person perspective and individuate those objects in terms of their behaviour: "if it walks like a duck, quacks like a duck, and swims like a duck, it must be a duck". My concept of practical concepts is similar to the concept that the classical American pragmatists had of concepts in general. To understand the meaning of a practical concept is to understand its *practical bearing*: to know what *effects* its referent has in *practice*.

I take this concept of *practice* to refer to the practice of everyday life as well as that of scientific investigation. It is the public context within which we agree on evidence we observed from the third person perspective. This is where practical concepts differ from phenomenal concepts, which operate from the first person perspective. If phenomenal concepts were practical concepts, that would mean that I could form the phenomenal concept of what it is like for a bat to observe a prey through his radar-like senses from studying whatever effects it has on the bats behaviour as can be studied from the third person perspective. And it would mean that a doctor who never suffers from headaches would be able to form a phenomenal concept of my tension headaches by understanding the effects these headaches have on my behaviour in practice. But these inferences are impossible. And suppose my doctor does suffer from similar headaches, recognises the effects they have on me as similar to the effects they have on him and correctly concludes that my experience resembles his own. In this case, the recognition of the effects in practice serve to link my situation to his own situation, but the phenomenal concept the doctor applies to me crucially involves a reference to what he himself experienced from the first person perspective and so involves more than just understanding the effects in practice.

Obviously, observations made from the third person perspective are of course also made from the first person perspective. If I observe that you scream in pain, I cannot observe

the pain qualia you experience, but I can only observe effects of your pain in practice. In this sense, I have third person access to what is happening to you. However, these effects I observe by experiencing them myself in a certain way, i.e. I experience auditory qualia when I hear you scream. Third person observation is *mediated* by first person experience.

According to Peirce, “our idea of anything *is* our idea of its sensible effects” (Peirce 1878 [1992: 132]). My concept of a practical concept is indeed that of a concept, which captures observable effects in practice. Or to put it in Fregean terms, we might say that while the *referent* of a practical concept is a property, its *sense* is something like “whatever produces such and such observable effects in practice”.

However, we have to distinguish sharply between *practically observable effects* and *observations*. A practically observable effect is an effect in practice that we can observe in a certain way, whereas the observation is that way in which we observe it. In other words, I understand observable effects in an externalist sense and observations in an internalist sense. When I encounter a wasp, one of the practically observable effects is that the wasp reflects light of certain wavelengths in my direction. My observation of the wasp has as one of its aspects that it features certain colour qualia. Now suppose that evil scientists achieved spectrum inversion in my phenomenal colour vision. My phenomenal concept of *what wasps look like to me* will be a concept of what qualia formations I experience when I observe wasps, and this concept will differ from your concept of what wasps look like to you if the scientists left you alone. Nevertheless, you and I can agree on what wasps are, even if we are aware that we experience them in different ways. This is because our experiences of what wasps look like correspond with our observations of wasps in virtue of an effect wasps produce in practice: the reflection of light in a certain way, among other things. So next to our different phenomenal concepts of what wasps *look like*, we share a practical concept of *wasps* that captures effects externalistically by abstracting away from the qualia.

We happen to live in a world where all members of the scientific community are also members of the same species, *Homo sapiens*, who experience the world in similar ways. But this is a contingent fact. Evolution could have produced two different species of comparable high level of intelligence but equipped with substantially different sensory organs and brain architecture. Suppose, for instance, that evolution produced *überbats*. These bats have huge brains, they have claws capable of constructing and operating tools, and they have the

capacity to communicate using sounds. If humans and überbats learned to communicate and started to do science together, I don't see why they couldn't agree on how to classify all sorts of things. Which means they would share practical concepts.

I would like to see how far we can go if we assume that practical concepts and phenomenal concepts jointly exhaust the realm of all concepts. This will commit me to some version of pragmatism: it means that, with the exception of phenomenal concepts, all scientific concepts, including those that describe the universe at the level of microphysics, are ultimately explicable in terms of practical bearing. I will return to this enterprise in section 5.

### *Phenomenal Representation*

We represent things phenomenally, that is to say, my representation of a wasp has a way it is like for me to have it. This phenomenality is not epiphenomenal but part of the representational system. If every representation is a point in a representational space, then at least some of the dimensions of that space are phenomenal. Let us stick to the visual for a moment. Suppose I am looking at a wasp flying in front of me. This causes me to have an experience of the wasp. The experience *represents* the wasp. This means that the representation has certain properties that jointly measure properties of the wasp in terms of the measurement scales defined within the representational system. My claim is that at least some of these scales range over qualia. This means that qualia instantiations are syntactic elements of phenomenal representations, just like characters are syntactic elements of sentences and numbers and symbols are syntactic elements of mathematical formulae. And just like numbers and symbols, qualia are not representations themselves. Instead, only within a *representational system that defines phenomenal scales of measurement*, and as part of some *compositional structure* (the experience) do qualia make a semantic contribution.

Just like we use knowledge of a language to attach meaning to a syntactic expression composed of symbols, so do we use knowledge of the way we experience things in order to attach meaning to our experiences. Knowledge of a language requires concepts of the various symbols, so that we can recognise them correctly, and knowledge of how this language represents objects in terms of symbol structures. By analogy, knowledge of the way we experience things involves phenomenal concepts, so that we recognise qualia, and knowledge that relates our phenomenal concepts to our practical concepts. This latter



knowledge causes me to trigger the practical concept of wasps whenever a certain pattern of phenomenal concepts is triggered that matches the constellations of qualia I experience when I observe wasps.

Does this mean I am postulating a *language of consciousness* along the lines of the well-known *language of thought* hypothesis? The answer is no. My claim is only that consciousness provides syntax for the human representational system, and I have used the example of language as an analogy for the sake of clarity. Given the large number of dimensions involved in phenomenal experience (brightness, two colour dimensions, three dimensions for spatial location, the dimensions of smell and taste, auditory dimensions such as loudness, pitch, and the dimension for timbre, and so forth), the high resolutions of discrimination we are capable of along these dimensions, and the complex arrangements in which qualia can be combined (different colours at different locations, etc.), it seems to me that we are dealing with a syntactical system whose expressive power is at least in degree far beyond that of spoken and written language, which already follows from the fact that we recognise linguistic expressions through our visual and auditory phenomenal representations whereas even when we share phenomenal concepts we have a hard time agreeing on how to express them in our language.

#### *'Upward' and 'Downward' Functionalism*

Functionalism is about identifying *functional roles* and explaining how those roles are *realised*. This basic idea already yields two approaches in order to be a functionalist about some phenomenon X: (1) to give a conceptual analysis of X as a functional role, find a realiser for this role and explain how the realiser realises X; (2) to find a functional role and explain how X realises that role. Let us call approach (1) 'downward functionalism' (because it analyses 'downwards' towards an underlying realiser) and approach (2) 'upward functionalism' (because it analyses 'upwards' towards an overarching role). These two approaches are not incompatible; they simply seek to answer different questions. Let us take the heart for example. Downward functionalism would analyse the heart as a blood pump, and explain how a certain muscled organ in our chest is able by means of its structure of ventricles and valves and its rhythm of contractions, to pump the blood. Upward functionalism would explain that, given that blood transports chemicals to and from all sorts of cells, it has to flow

towards and back from cells, and that the heart makes this happen because it pumps the blood around.

This example shows that the distinction between roles and realisers is a relative one. In relation to blood circulation, the heart is a realiser, but in relation to the ventricles and valves it is a role. Essentially, every practical concept specifies a functional role, because it identifies its referent in terms of its behaviour. The way I see it its roles all the way down to microphysics, and the lowest concept layer is the one which refers to roles for which no deeper roles can be found in order to realise them. We explain the occurrence of these unrealised roles by postulating laws of physics.

What about phenomenal concepts? My account of qualia has been a form of *upward functionalism*. I have identified a functional role – syntactical property of a representational system – and explained, albeit in a rather sketchy form, how qualia realise this role in humans. Nevertheless, I think qualia can also be thought of as roles themselves. What qualia *do* is that they *are like something*, and for certain qualia I think something can be said about *how that is done*. For example, how is it that a purple dress looks to me the way it does? I can explain that my phenomenal concept of purple is a concept of a colour halfway between red and blue, that human colour experiences are realised by combining a phenomenal value in a dimension which ranges from red to green with a phenomenal value in a dimension which ranges from yellow to blue, and that purple qualia are realised by combining a value towards the red with a value towards the blue. In my view, this is a functional explanation in the sense of downward functionalism. But of course, it only explains phenomenal concepts in terms of other phenomenal concepts. Let us call this *intraphenomenal downward functionalism*. The intraphenomenal downward functionalist allows that qualia are structural, interrelated properties about which a lot can be explained and understood. My example of colour combination is only a silly example of this. However, down the road every downward explanation leaves some role unrealised, and the same goes for intraphenomenal downward functionalism. For example, the difference between the way green looks and the way red looks is not something about which I can explain anything further. Our phenomenal concepts of red, green, yellow and blue seem phenomenally *basic*.

What I cannot do either is analyse *those* concepts further in terms of *practical* concepts. That would be strong reductionism, a position we rejected earlier on. We know that,

corresponding to the two-dimensional phenomenal colour spectrum, there is neural architecture in our brain. For example, we know how cones that pick up green light and those that pick up red light map onto the same neurons in such a way that registration of green light inhibits the neuronal activity whereas that of red light excites it. This seems to explain at least *something* about our phenomenal colour spectrum. And so it does. What it explains is that we cannot discriminate between “more of what looks red to us” and “less of what looks green to us”, because they are coded in the same way and therefore measured along the same axis. But what it does not explain is why red qualia look the way they do and not the way green qualia look.

Summarising, we can be *upward functionalists* as well as *intraphenomenal downward functionalists* about qualia. We have seen that qualia can fulfil a syntactical task that is required in order to recognise objects. Note that the concept of recognising objects is a practical concept. In other words, our *upward* functionalism is *not* intraphenomenal.

## 2 The Function of Phenomenal Affectivity

The function of phenomenal affectivity is even more obvious than that of qualia. If you have been working too long and too hard behind the computer, your fingers, wrists or arms may start to feel painful. You decide to quit and when I ask you why you stop, you might answer, “Because my wrist hurts”. When you walk away and I ask where you are going, you might answer that you are going to get some food because you are hungry.

Negatively affective features such as pain make us *quit* certain behaviours, and positively affective features such as hunger make us *start* certain behaviours. That is what affectivity *does*. From a biological or *design stance* perspective affectivity has a function in many occasions. Hunger, for example, is triggered when we are in need of food and it plays the role of engaging in behaviour aimed at finding and eating food. And pain obviously serves us when it makes us avoid or end a state of affairs that is damaging our bodies.

### *Upward Functionalism about Phenomenal Affectivity*

In section 1 we have seen that human beings represent things phenomenally. These representations present us with information about the state of our environment and about

the state of ourselves. On the basis of this information we take action. Since the information is *made available* in a phenomenal way, it also has to be *received* in a phenomenal way in order to guide action. This phenomenal way of guiding action I call *motivation*.

*Phenomenal affectivity* is the motivational dimension of phenomenality. It does not involve a plurality of dimensions itself, but ranges simply from positive to negative, from attraction to repulsion. This means for example that our concepts of pain, disgust and regret do not refer to different affective properties, but that they refer to classes of experience which all involve a negative value in the same affective dimension, but which differ with respect to the kind of phenomenal representations they operate on. For example, our concept of pain applies to those affectively negative experiences which are (1) about our own bodies and (2) involve certain qualia, such as the feeling of stinging and burning, but not others, for which we would have used other concepts ('nausea' for example).

Strictly speaking, *motivation* or even *action guiding* does not qualify as the practical role which is realised by affectivity, because I take motivation to *mean* affective guiding and I take *action* to *mean* motivated behaviour, with the consequence that my concepts of motivation and action are not entirely nonphenomenal themselves. On a practical level, what affectivity realises is that, given that representation is realised by phenomenal quality, the represented information can be used by a deliberation process according to some system of preferences in order to generate a certain kind of behaviour. Thus we can be upward functionalists about phenomenal affectivity in a way that complements our upward functionalism about phenomenal quality.

### *The Role of Taste*

Consider again how phenomenal quality plays a role in recognition of objects. Let us take coffee as the object for our example. When I drink coffee, my experience triggers a phenomenal concept of how it tastes, and on the basis of my memory that connects phenomenal concepts to practical concepts, my practical concept of coffee is activated and I recognise the drink as coffee. We can portray this analysis as postulating a relation between two subsystems: the phenomenal representation system (PRS henceforth) which activates a phenomenal concept on the basis of current experience, and the phenomenal-practical

memory (PPM henceforth) which serves as a function or lookup-table connecting phenomenal concepts to practical concepts.

Phenomenal affectivity is the result of a third subsystem which interacts with both PRS and PPM, and that is the system of *taste*. Taste is a function that relates phenomenal qualitative concepts to phenomenal affective values and to something that I shall call *mood*, for lack of a better term. In different moods, activation of the same phenomenal qualitative concept can result in different phenomenal affectivity without this signifying a change in taste. So taste is a double argument function that produces phenomenal affectivity on the basis of phenomenal concepts and mood.

In relation to PRS, taste gives our current experience of the actual its phenomenal affective feel, resulting in the experience, for example, that the coffee *tastes good*. Let us call this kind of affectivity *reward*.

In relation to PPM, taste attaches affective value to our memory of practical situations and events that *have* generated positive or negative experiences in the past and/or *could* generate positive or negative experiences in the future. The relation between taste and PPM is what underlies the experience of *desire*. Thus, when I wake up in the morning, I remember the taste of coffee and on the basis of my attraction towards the memory of the taste I start making coffee. Note that on this account, taste is understood in an internalistic sense, because it operates on phenomenal concepts, whereas the content of desires is understood externalistically, since desires are constituted in relation to PPM, which provides practical concepts that refer to the properties of objects we encounter in practice.

Since *reward* is constituted by the relation between taste and PRS, and *desire* by the relation between taste and PPM, one might argue that we have established downward functionalism about desire and reward, featuring PRS, PPM and taste as realiser components. Note however that this kind of realisation, if one thinks the notion applies in this case, remains strictly intraphenomenal. PRS is a phenomenal module, and so is taste. Even the concept of PPM presupposes phenomenal concepts, since it is defined as a phenomenal-practical concept associative table. Given that the whole system does something practically, and that its basic components are defined using phenomenal concepts, including the concept of phenomenal affectivity presupposed by the taste module, we must conclude

that our 'broader' functionalism of phenomenal affectivity itself is still a version of upward functionalism. Let us now consider some criticism that this view might be vulnerable to.

### 3 Against Eliminativism

Recall Dennett's example of Chase and Sanborn. According to the view that Dennett attacks, which I shall call the *target view*, there are such things as qualia, and consequently we can distinguish 'pre-qualia' and 'post-qualia' processing. Pre-qualia processing includes low-level sense perception mechanisms – similar to PRS in my proposal – whereas post qualia includes memory of past experiences – PPM in my terminology – and what Dennett calls one's "canons of aesthetic judgement", to which I refer as one's *taste*. The target view takes the reports of Chase and Sanborn at face value. Chase's growing dislike for Maxwell Coffee is explained by a change in his taste, whereas Sanborn's dislike is explained by a change in PRS. Since the former is post-qualia whereas the latter is pre-qualia, the target view holds that Chase's qualia have stayed the same while Sanborn's must indeed have changed.

Against this view, Dennett argues that Chase and Sanborn could be mistaken. Although the target view explains Chase's report, this report could equally well be explained by a shift in both PRS and PPM, while Chase's taste remained constant. In that case, his qualia changed over time, resulting in a growing dislike for the coffee, but because his memories were changed along with it, he misremembers coffee in the past as having tasted the same. Conversely, Sanborn's report also allows an alternative explanation. It could be that there is nothing wrong with Sanborn's PRS, but that both his taste and his PPM changed. As Dennett puts it, this would be a kind of "nostalgia effect" (Dennett 1988: 626), which altered Sanborn's memory of the past in terms of his changing taste so as to keep representing the past experiences as positive.

So the upshot of Dennett's argument, recast in my terminology, is this. We have three subsystems – PRS, PPM and taste – and since our qualia reports are based on relations between these subsystems, each report can be explained in different ways by postulating changes in different subsystems. Every explanation based on a change in one system allows for a rivalling explanation that postulates a change in the other two. Hence, we would never know whether there was 'pre-qualia' change or not, and thus we would never really know

whether qualia reports are accurate. This leads Dennett to adopt eliminativism, the view that there are no qualia at all.

### *Reporting Phenomenality*

Dennett claims he does “not deny the reality of conscious experience”, and that he grants “that conscious experience has properties” (Dennett 1988: 619). He claims only to be denying that there are qualia. However, this has generated a lot of confusion, as the term “qualia” has been taken to mean a lot of things. For Dennett, qualia are supposed to “loom as a challenge to functionalism or materialism or third-person objective science” (Dennett 1988: 627), and that is why he wants to quine them. In his view, all claims about consciousness should be made, or be verifiable, from the third person perspective. If we brush terminological issues aside, I think his eliminativism is very close to what I have called *strong reductionism*, the view that all meaningful propositions about properties of consciousness should be explainable in terms of third person perspective concepts. What distinguishes Dennett’s position from that of weak reductionists like Loar is that Dennett wants to show that every concept must be interpreted as being about *behaviour* that we can agree on from the third person perspective, and that if we cannot, we should get rid of the concept. We can clearly see this view at work when Dennett discusses an objection to his arguments about coffee tasting:

There is a strong temptation, I have found, to respond to my claims in this chapter more or less as follows: “But after all is said and done, there is still something I know in a special way: I know *how it is with me right now*.” But if absolutely nothing follows from this presumed knowledge – nothing, for instance, that would shed any light on the different psychological claims that might be true of Chase and Sanborn – what is the point of asserting that one has it? (Dennett 1988: 626)

Now let us focus on the objection first. In Dennett’s example, Chase and Sanborn have not really reported their qualia at all. They only made claims about their qualia. The first claim they made was about whether the qualia they experienced were similar to qualia they

experienced in the past. But even if they were mistaken on this account, the objection runs, it does not follow that they were mistaken about their qualia of the present. Note that both Chase and Sanborn assume their PPMs are working well and that both Dennett's alternative explanations assume that their PPMs are failing. Thus we can blame all the mistakes on memory. The second claims they made did not report the qualia either, but only the affectivity with regard to whatever qualia they were experiencing. And Dennett does not dispute that Chase and Sanborn dislike the coffee when they report that they dislike it.

In a crucial way, the knowledge of qualia of the present seems to have been left out of the reports. Dennett seems to be thinking that this counts against the very idea of knowledge of qualia, and thus of irreducible phenomenal concepts. But we must not forget that it was Dennett who designed the example. So why did he let Chase and Sanborn report only on the relation between present and past qualia, and on the affectivity towards the present qualia, but not on the present qualia themselves?

Let us therefore expand the example. Chase reports not only that the coffee tastes the same to him as it always did, and that he has grown to dislike it, but he also reports what that taste is like, using metaphors and subtle terminology that coffee tasters use in order to communicate their phenomenal concepts about coffee qualia. Because Sanborn is also a coffee taster familiar with this terminology, he responds to Chase's report in despair, explaining how much he would still like that taste, but that to him the coffee tastes different now, and using coffee taster's metaphors, he reports the qualia he experienced when he tasted the coffee. Chase reacts surprised, together with other Maxwell House coffee tasters, who all agreed that Chase's description of his qualia was very similar to their own, and very different from Sanborn's. In fact, Sanborn's PRS might be so screwed up that he cannot even find the right words anymore to describe his qualia, which is a surprise to his colleagues in itself, because, let us say, he was always one of the best in describing the intricacies of coffee taste.

Now suppose you are a psychologist and your topic of interest is taste. Why would you *not* use these additional reports as data? Why could these data not shed light on the psychological condition of Chase and Sanborn? And suppose you are a psychologist who never tasted coffee in your life, let alone paid attention to the world of different coffee tastes and subtleties. Doesn't it make sense to say you would in that case be unable to really



understand what the coffee tasters are talking about, no matter how much you would read about coffee and the relation between the metaphors in the taster's reports and, say, the states of their brains? I would say that a bit of introduction into the world of coffee by tasting it yourself would shed a *lot* of light on the minds of coffee tasters for a psychologist. In fact, without it, you'd be totally in the dark. Therefore, the 'expanded' case of Chase and Sanborn no longer gives us reason to doubt the empirical relevance of phenomenal concepts.

### *The Dynamics of Mood and Taste*

We can strengthen this line of argument against Dennett if we consider the fluctuations of mood. Enter Jane, the wife of Chase. Jane loves Maxwell coffee, and because she is the wife of Chase, she acquired the language used by professional coffee tasters. However, she would never become a Maxwell House coffee taster herself. Even though she is able to recognise Chase's description of the taste, and agree with Sanborn's praise of *that* taste, she simply does not fancy tasting it *all day long*. She is not always *in the mood* for coffee. Among people who spend most of their working hours in universities this may sound like blasphemy, but in fact it is very normal for people to like certain drinks and yet only appreciate them under certain circumstances. Personally I like coffee after dinner, for example, but not during dinner or just before it, when I am hungry and anticipating the meal.

Now consider Dennett's position again. For Dennett, reference to taste qualia is nonsensical because one cannot in principle distinguish between disliking a taste one used to like at an earlier instant and encountering a different taste. When we apply this reasoning to short-term affective changes due to mood fluctuation it becomes absurd. It would mean that whenever Jane is not in the mood for coffee she would be unjustified in saying that she knows what Maxwell coffee tastes like yet does not feel like it (or that she dislikes it when she drinks it anyway), because it might just as well be that during such a period of time her memory of the taste of coffee is malfunctioning, whereas it functions correctly when she does feel like coffee. In my view, that is ridiculous.

Conversely, if we *do* allow the very intuitive idea that Jane knows the taste of coffee also in the moments when she does not feel like it, and that she *can* recognise the way coffee tastes to her as being the same in the cases she likes it as in those cases she drank it while not

being really in the mood for it, I don't see how we could possibly avoid introducing such a concept as phenomenal quality and distinguishing it from phenomenal affectivity.

My suggestion is not that we experience quality and affectivity separately. We only experience the totality of the outcome, so to speak, which means we experience everything affectively. Note that in this respect I differ from the target view. Dennett presents the target view as a view that understands phenomenality in terms of qualia, and allows that people have 'reactive attitudes' to these qualia which involve both descriptive and aesthetic judgements. The target view does not seem to understand these aesthetic judgements as being a dimension of phenomenality themselves. And maybe *judgements* aren't phenomenal experiences indeed, but in that case my point is that a view that only allows for affectivity in the form of judgements has left out something vital, namely the *experience* of affectivity that *leads* to the judgments.

### *Fallibility*

Despite the fact that all our experiences are thoroughly affective, we have the capacity to compare our experiences for qualitative similarities even when affectivity differs. Dennett relies heavily on the fact that human beings are fallible in their judgements, but my suggestion is not that we are *infallible* in making such comparisons either. We have to look 'through' the affectivity as it were, and notice what it is that we like in one mood and dislike in the other, and this might be a cognitive task that involves fallibility. I do not have to deny that people can sometimes mistake affective difference or similarity for qualitative difference or similarity in order to maintain that affectivity and quality are different concepts.

We are also not infallible in making claims about our taste, and about whether it changed or not. Taste produces affectivity as a function of phenomenal quality and mood, we have seen, and to say that one's taste has changed is to say that this function was altered rather than that the function got different input. For example, Chase's claim that his taste has changed does not only involve a belief that his qualia have remained constant, but also that his dislike for the coffee he is tasting is not just because he is not in the mood for coffee, or for that kind of coffee, should coffee tasters happen to have different moods for different kinds of coffee. In fact, the first time he unexpectedly disliked the coffee he might well have ascribed it to the mood of the moment, and only once he discovered that over time he

consistently started disliking Maxwell Coffee in circumstances in which he used to like it before might he have adopted the interpretation that his own taste has changed.

Summarising, the Chase and Sanborn case does not establish that phenomenal qualitative concepts lack empirical relevance; Dennett's reasoning becomes absurd in the face of mood fluctuation; and the use of phenomenal concepts is perfectly compatible with ordinary human fallibility.

#### 4 Against Epiphenomenalism

Consider the following version of the *zombie* argument (Chalmers 1996: 94-99). There is a logically possible world that is indiscriminable from the actual world in terms of practical concepts, yet features no phenomenality at all. In zombie world, my *zombie twin* is sitting behind a computer, and its brain is processing incoming information by means of neural mechanisms just like my brain, and produces the same kind of behaviour as my behaviour. In zombie world, these mechanisms are apparently sufficient to realise the very functional roles that I claim are being realised by phenomenality in the actual world. But if they are sufficient in zombie world, they must be sufficient in the actual world as well, and phenomenality must be seen as an *epiphenomenon*,<sup>1</sup> an addition in the actual world in comparison to zombie world that does not serve any function. Hence, my upward functionalism about phenomenality is false.

Of course, we may wonder whether zombies are logically possible. However, let us not forget that we already rejected *strong reductionism*, which means that we cannot hold that it follows from the *meaning* of our practical concept descriptions of the actual world that phenomenality must exist. Is not the absence of such conceptual entailment precisely what

---

<sup>1</sup> Chalmers himself denies that the zombie argument implies epiphenomenalism. He understands epiphenomenalism as the view that phenomenality does not have a *causal role*, and speculates that this might not follow from the view that it does not have a *functional role* (Chalmers, 1996, 150-160). Nevertheless, if we understand epiphenomenalism as the view that it has no functional role, then Chalmers is definitely an epiphenomenalist. As he concludes elsewhere: "I think it best to accept instead that phenomenal consciousness is distinct from any physical or functional property, and that it does not need to have a function to be central to our mental lives." (Chalmers, 1997, 423)

logical possibility means? From the absence of logical necessity, and the premise that phenomenality does necessarily arise from the referents of practical concepts, it follows that the actual world differs from zombie world in terms of laws of nature. Laws of nature connect *distinct* properties. The zombie argument seems to leap from the rejection of strong reductionism to the conclusion of property dualism without any room for the intermediate position of weak reductionism.

Following Chalmers, we shall refer to this dualism of phenomenal and nonphenomenal properties connected by natural law as *naturalistic dualism*. From naturalistic dualism about the actual world and the functional sufficiency of zombie world the conclusion follows that phenomenal properties have no function. This is a serious threat to upward functionalism.

### *Supermasochism*

If naturalistic dualism is true, then not only zombie world is logically possible, but *spectrum inversion* must be logically possible as well. For if the correlations we find between practical concepts and phenomenal concepts are not based on conceptual entailment but on natural laws connecting distinct underlying properties, then it is no less logically possible that the laws could have been *different* than that the laws would be *absent*. In other words, *any* systematic lawful connection between the practical and the phenomenal would have to be logically possible.

Now with respect to an inversion of colour qualia this shouldn't be much of a bullet to bite after one has already swallowed zombie world. Colour inversion seems much weaker, and as I have argued in section 1, it is even compatible with my brand of upward functionalism.

However, there is another phenomenal spectrum inversion which I find much more difficult to accept. Enter the *supermasochist*. A supermasochist is practically indiscriminable from a human being and has the same qualia as a human being, but his *phenomenal affectivity is inverted*. In supermasochist world, my supermasochist twin enjoys a quale whenever I am suffering from the very same quale, and vice versa. However, since supermasochists are identical to their actual twins in terms of practical concepts, this means that supermasochists typically work towards negative experience and attempt to avoid positive experience.

Supermasochists are not like masochists in the actual world. A masochist in the actual world also wilfully submits himself to negative experience, such as pain, but it can be argued that for a masochist the pain is, even though affectively negative in itself, in a wider context stimulating in a positive way. Under appropriate conditions, such as those that obtain in a consensual sadomasochistic relationship, the pain will trigger further emotions in the masochist which include intense pleasure and comfort. Not so for the supermasochist. A supermasochist does not submit himself to pain because all things considered this generates a positive experience. A supermasochist submits himself to pain because he is hard wired to engage in behaviour that in supermasochist world results in painful experience. To further illustrate the difference between masochism and supermasochism, consider the supermasochist twin of a masochist in the actual world. Whereas the masochist gets stimulated in an ultimately rewarding way when he is administered pain by his dominatrix, the *supermasochist twin masochist* actually feels a 'pleasurable' stinging or burning sensation when he is beaten, and under conditions which mirror sadomasochism in the actual world, this initially pleasurable experience will generate a further state of mind that makes him feel *utterly miserable*. But he cannot help it, because supermasochists are hard wired to engage in behaviour that makes them feel miserable.

The idea of supermasochism is conceptually absurd. Even though I just succeeded in writing down what it would be like, I do not claim to understand really what I am talking about when I consider an organism which experiences that which he attempts to realise as affectively negative, all things considered. Is a supermasochist motivated by negative affectivity - pain and misery? No, because motivation is already implicit in the meaning of positive and negative affectivity. It is part of the very idea of affectivity that it has a positive and negative direction, and that what is positive about the positive direction is that it *motivates*, whereas the negative *deters*. But then it means that a supermasochist aims at what deters him, and avoids that which he is motivated to achieve. However, that is nonsense too. It is also part of the very idea of motivation that one attempts to realise what one is motivated to realise, and that one attempts to avoid that which deters you.

I think we have stumbled on something fundamental here. Apparently, the concept of *motivation* has phenomenal as well as practical implications that cannot be separated from each other on the penalty of conceptual absurdity. Yet this does not make us strong

reductionists either. It is one thing to say that I engage in a certain kind of behaviour, and it is something else to say that I feel desire or pain, but nevertheless there is a conceptual connection between the two that rules out supermasochism.

However, since naturalistic dualism would imply the logical possibility of supermasochism, the logical *impossibility* of supermasochism implies that naturalistic dualism must be false. If phenomenal affectivity has conceptual implications on behaviour that rule out supermasochism, then the connection between the phenomenal and the practical is not merely a matter of natural law, and phenomenal affectivity is no longer an epiphenomenon. And since phenomenal affectivity operates on phenomenal quality, this means that qualia cannot be epiphenomenal either. Given that I like the qualitative taste of coffee, and given that coffee produces this taste in me, I am motivated to drink coffee.

### *The Ensuing Dilemma*

We must note two things now. First, phenomenal affectivity inversion is logically impossible in relation to the practical, but not in relation to phenomenal quality. Rather, if qualia inversion were possible it would imply a corresponding affective inversion *with respect to the inverted qualia* in order to maintain practical indiscriminability. The result would be one in which the relation between affectivity and practicality remains uninverted. Given that the role of qualia is syntactical, and given that practical concepts do not logically have phenomenal qualitative implications, I think I have to admit that qualia inversion *is* logically possible.

Second, even though we have argued against naturalistic dualism from the logical impossibility of supermasochism, it is not at all clear whether this renders *zombie world* logically impossible as well. It *should* if the reasoning from zombie world to naturalistic dualism is compelling, but in that case there must be *something* logically impossible about zombie world, and we haven't established such a thing.

These two notes put us in an awkward position. Within the current debate, it seems one either sides with the property dualist and admits that all inversions of the phenomenal vis-à-vis nonphenomenal, as well as the zombie scenario, must be at least logically possible because the properties involved are distinct, or one sides with the reductionist and claims that zombies as well as *all* inversions are logically impossible because the properties are

identical. But now we have intuitively arrived at the conclusion that *some* inversions are possible and others are not. The obvious option, to be a reductionist about uninvertible phenomenal properties and a dualist about invertible ones is not available either, because dualism implies the logical possibility of zombies, and one cannot be a qualitative zombie without being an affective zombie as well, since affectivity presupposes phenomenal quality. Furthermore, we seem to have reasons to suppose that zombie world is logically impossible in virtue of its relation to other thought experiments, but we have found nothing within the zombie scenario itself to account for this.

## 5 Pragmatism and the Problem of Underdetermination

Rather than confronting the dilemma immediately, I am first going to digress a bit towards the philosophy of science and the philosophy of language. In particular, I will look at a proposal from Herman de Regt to use Peirce's pragmatic maxim in order to solve the problem of underdetermination (De Regt *forthcoming*). Within my framework, this proposal amounts to the view that all scientific concepts are *practical concepts*. The problem of underdetermination provides a good context for further development of the idea of practical concepts and to see how this idea may fit into contemporary philosophy of science.

After that I will return to the dilemma between weak reductionism and property dualism in section 6. I will criticise remarks about phenomenal experience made by De Regt, who applies the pragmatic maxim in a way that resembles strong reductionism. Interestingly Peirce, in his later work, steered pretty much in the opposite direction, and developed a rather speculative idealism, which is different from the three positions of strong reductionism, weak reductionism and property dualism discussed so far. Inspired by Peirce, in section 7 I will finally urge my own position, which I shall call *affective monism*, the view that the metaphysics of our world is fundamentally affective in nature. This view will provide a way out of our dilemma. Furthermore, affective monism accepts some kind of underdetermination, but not in such a way that it becomes rational to deny that entities postulated by our best scientific theories exist. Thus the problem of underdetermination will be solved as well.

## *The Problem of Underdetermination*

A scientific theory that explains certain observations is *underdetermined* by those observations if it is possible to explain the observations equally well with different theories. With respect to this difference, we can distinguish two kinds of underdetermination. Following De Regt, let us speak of *historical underdetermination* when two theories explain past observations equally well, but yield different predictions for future observations, which might involve new experiments that create conditions under which the theories have not yet been tested (De Regt *forthcoming*). In this case, the theories are different because they yield different predictions.

In contrast, *logical underdetermination* involves two theories that not only explain the same past observations, but also yield the same predictions for future observations (De Regt *forthcoming*). Nevertheless, if two theories are logically underdetermined by the same observations, they must be different in the sense that if one of them is true, the other must be false. In this essay, we will deal with logical underdetermination only, and I will from now on refer to it simply as underdetermination.

Bas van Fraassen claims that all scientific theories that postulate unobservable entities, such as photons, electrons, and even viruses, are underdetermined (Van Fraassen 1980). If a scientific theory explains all our observations by postulating unobservable entities, and would furthermore explain all future observations, then this makes the theory *empirically adequate*, but it remains an open question whether the theory is *true*.

The argument for this view, which Van Fraassen calls *constructive empiricism*, runs as follows. Let us take the theory of viruses as our example. The theory explains why people get ill under certain circumstances by assuming that there exist very tiny organisms which have structural properties that allow them to interact with human bodies in such a way as to make us ill. Furthermore, if we combine the theory of viruses with our physical theories, we can predict that, although viruses are too small to see with the naked eye, we can verify our claims about virus anatomy using electron microscopes that magnify an image of the virus on a screen. However, observation of this image only verifies our theory of viruses on the assumption that our theory of electron microscopes is correct. But this theory is itself based on assumptions about what unobservables underlie our observations. The only thing we know is that together, they are adequate in capturing the relation between conditions under



which people get ill and conditions under which electron microscopes produce certain images. If our physical theory is false, then the images no longer verify the virus theory, and in fact, on the truth of some alternative physical theory, might be explained by an alternative to virus theory which together with the alternative physical theory would predict the same relation between illness and microscope observations, without assuming the existence of viruses, and perhaps by assuming the existence of something else instead. The general, underlying claim is that, when we take all of our best scientific theories as a whole, they might be adequate in capturing all systematic relations between observations, but insofar the different interlocking theories relate to each other in terms of postulated unobservables, it is possible in principle to construct alternative theories that derive the same observational relations from a different underlying unobservable structure.

This poses a problem for the philosophy of science. The problem is simply that if this is true, then it may no longer be rational to believe in our best scientific theory. More specifically, it would no longer be rational to believe in viruses. Van Fraassen has proposed to solve this problem by replacing the 'Prussian' concept of rationality, which holds that it is rational to believe a proposition only if there are compelling reasons in favour of the proposition, by the 'English' concept which holds that it is rational to believe a proposition as long as there are no compelling reasons *against* it. Because the virus hypothesis is compatible with current evidence, there are no compelling reasons against it, and therefore it would be rational to believe in viruses. However, a consequence of the English concept, in combination with Van Fraassen's view that virus theory is underdetermined, is that it would *also* become rational to *deny* that viruses exist. In my view, this is not a very intuitive result.

A second problem for Van Fraassen's view is that, despite his own opinion, I believe it can easily be extended into the realm of the observable. As Churchland has argued, our own senses and the (low level) neural networks that process information from those senses are not in any fundamental sense different from scientific instruments, and involve background theory in the *way* they process the information. Thus, the distinction between what is observable and what is unobservable loses its epistemological significance (Churchland 1982). If the theory-ladenness of the instruments we use to observe viruses would establish the rationality of virus anti-realism, then the theory-ladenness of our visual cortex might just

as well make it rational to disbelieve in anything we see in every day life. And that would be radical scepticism indeed.

### *A Pragmatic Solution*

According to De Regt, we can solve the problem of underdetermination by adopting Peirce's method of conceptual elucidation, which is summarised by the pragmatic maxim: to understand the meaning of an idea is to understand its sensible effects, its observable effects in practice. De Regt uses this method to argue against the claim that equally empirically adequate theories could be logically different. He quotes the following passage from Peirce:

What the pragmatist has his pragmatism for is to be able to say:  
"Here is a definition and it does not differ at all from your confusedly apprehended conception because there is no *practical* difference"  
(Peirce 1903 [1998: 141]).

The pragmatic maxim now serves as a principle of individuation: two concepts are not different if they are not practically different. Since practical bearing is defined in terms of observable effects, practical bearing becomes just as coarse-grained as empirical adequacy. Thus, the upshot is that two theories cannot differ in meaning if they do not differ in empirical adequacy. But if they do not differ in meaning their truth-values cannot differ either, and so the gap between empirical adequacy and truth disappears.

Since the very idea of underdetermination is based on a dissociation between truth and empirical adequacy, the pragmatic verdict must be that the concept of underdetermination that gave rise to the problem was unclear, and that the discussion between scientific realism and constructive empiricism that was based on this problem was ill formed. As De Regt puts it:

And now we can see that the concept of underdetermination which intuitively undermines the objectivity and veridicality of science is unclear. Indeed, the pragmatist will argue that if we think that even when all empirical data are in there will still be an arbitrary choice for either this or that "ontology of unobservables", we actually have an

unclear concept of reality, existence, truth and science (De Regt *forthcoming*).

Suddenly the entire problem has been brushed aside. What has happened? Is this a cheap trick or a genuine philosophical improvement? Let us analyse the argument in terms of the framework I developed earlier on. We have defined those concepts that are individuated by means of the pragmatic maxim as *practical concepts*. What De Regt is proposing is that we understand all scientific concepts as practical concepts, and that we get rid of philosophical concepts that cannot be understood in this fashion, such as the concept of underdetermination.

Let's apply this strategy to the virus example. If our concept of a virus is a practical concept, then the meaning of this concept should amount to a conjunction of predicates of the form 'that which makes organisms ill in such and such ways under such and such circumstances', 'that which produces such and such images on the screens of electron microscopes', and so forth. These predicates do not include any information that goes beyond the scope of empirical adequacy. Therefore, it becomes impossible to construct a theory that is equally adequate but logically incompatible with it. And this makes it impossible to explain what underdetermination means.

As a result of this, the concept of *realism* changes its meaning. It is no longer the *substantial* claim that the entities postulated by empirically adequate scientific theories really exist, but it becomes the trivial claim that if we observe certain effects, then *whatever it is that those effects are effects of must exist*. Such a claim no longer requires argument, nor should it prove to be very useful in, for example, an explanation of the success of science. Maybe we can accept realism as a tautology. At most, it involves only the assumption of an 'external permanency', as Peirce called it.

Conversely, the concept of constructive empiricism turns out to be incoherent. Van Fraassen has sometimes suggested that, regardless of the technical arguments on underdetermination, we can simply establish the distinction between empirical adequacy and truth by considering that, rather than viruses, there might be 'as if viruses', entities that behave as if they are viruses while they aren't. But once we understand the concept of a virus as a practical concept, it becomes equivalent to the concept of 'whatever behaves like a

virus'. Hence, the concept of something that behaves like a virus but is no virus turns out to be self-contradictory.

However, this solution leaves two problems. First of all, if underdetermination really is an unclear concept, then what should we say about the reasons that gave rise to the idea of underdetermination in the first place? What about the insight that any unpredicted result can be accounted for in different ways by revising different hypotheses? Secondly, what about phenomenality and phenomenal concepts? If any truth there is to be had in science is to be cast in terms of practical concepts, then doesn't that commit us to strong reductionism, a view that I already rejected? Doesn't the practical story leave facts about consciousness underdetermined? Shouldn't we leave room for the facts about what it is like to be a bat, for example, to be underdetermined by the third person accessible evidence for our scientific theories? I will deal with the first problem presently. We shall return to the second problem in section 6.

### *Two Modes of Practical Concept Individuation*

Let us focus on what I shall call the *Choice Thesis* (CT), the claim that recalcitrant evidence can always be accounted for in different ways by revising different hypotheses. In order to maintain that underdetermination is unclear, at least in the shape that gives rise to constructive empiricism, we must either show that CT does not imply underdetermination, or we would have to show that CT is false. Since I believe that CT is correct, I am going to take the former approach.

CT is usually understood in the context of a 'network' view of scientific theories. According to this view, a hypothesis is part of a network of many interrelated hypotheses, and it is really the *entire network* that is tested against experience, rather than the individual hypothesis. In the words of Quine:

[T]otal science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field. (Quine 1953: 42)

The argument from CT to underdetermination of this 'total science' is straightforward: if any piece of evidence can be accounted for by more than one way of adjusting the network, then

apparently different network configurations can satisfy the same empirical boundaries. This is not a historical but a logical point, because for every future observation CT will apply again. Quine himself explicitly regards CT as a sign of underdetermination:

[T]he total field is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to re-evaluate in the light of any single contrary experience. (Quine 1953: 42)

However, we must not forget that Quine is here talking about *statements*, about the question of how to assign truth-values to *sentences* within the *language* of the scientific theories in question. With respect to the meaning of such sentences, Quine defended a holistic view. The meaning of a sentence depends on the way that sentence is related to other sentences in the network. But this has an important consequence. Suppose we have to choose between revising statement  $S_1$  and  $S_2$  in the light of contrary experience. We may decide to save  $S_1$  by rejecting  $S_2$ . However, since the choice was between  $S_1$  and  $S_2$ , there must be some significant relation between these statements in the network. Since the meaning of  $S_1$  is a function of the way in which it is related to other statements, revising the truth value of  $S_2$  may very well *change the meaning of  $S_1$* ! Quine may have shown that in principle we can decide to sacrifice or save any statement we like in the light of any observation we encounter, but it is not at all clear that we can always save the *original meaning* of any statement in the light of new observations.

Some philosophers have argued for even less realistic views of science on the ground of this kind of meaning shifts. But I think we should reason in the opposite direction. In my view, underdetermination of statements and the holism of meaning *cancel each other out* in such a way that realism becomes reasonable again. If we choose to save a hypothesis, and we take into account further development of our theory on the basis of this decision in the light of further evidence, then in the long run, our theory will *amount to the same thing* as the theory we might have ended up with had we initially decided to reject the hypothesis. And it is this “same thing” that both theories will have in common that we can be realists about.

In order to explicate this view we need to distinguish two modes of concept individuation, which I shall call the *mode of expression* (Peirce 1878 [1992: 130]) and the *mode of*

*practical bearing*. In the mode of expression, practical concepts are individuated in terms of whatever framework, language, scheme or other system of communication some community uses in order to share these concepts. I am not here thinking of conceptual schemes in the strong sense of the word, but rather of a Davidsonian picture where language use consists in the tuning of a *passing theory* by the interpreter to the *idiolect* of a speaker (Davidson, 1986). Whatever it is that this idiolect draws on that the interpretation can tune into is what constitutes the mode of expression (this includes what the speakers believes to be the interpreters *prior theory*, the information the speaker conveys to the interpreter about how to adapt this theory to his utterance, by whatever means, and the general context in which the utterance takes place). This mode, then, is the level of Quinian/Davidsonian holism, involving concepts as *predicates* that derive their meaning from their role in the various interrelated *statements* that the person who has those concepts believes to be true. Within the mode of expression, any configuration of interrelated statements that explains a set of observations is underdetermined by those observations.

The mode of practical bearing individuates concepts in terms of their practically observable effects. This may sound confusing. We are concerned with practical concepts, and I already defined those as capturing observable effects, so how is it that this now turns out to be just one of two modes of individuation? The answer is that both modes involve the capturing of observable effects, but that the mode of expression also involves a 'way' in which these effects are captured. The mode of expression is the mode of expression of observable effects. Given that the same observable effects can be expressed in different ways, we can construe the mode of practical bearing as a coarse-grained version of the mode of expression which groups together different possible expressions of the same observable effects.

Obviously, this is where we have departed from the Davidsonian view. Davidson found it nonsensical to refer to content in a mode independent of our way of expressing that content. Put differently, for Davidson the mode of expression is the only mode of individuating concepts and statements. This is directly related to his nominalist view on ontology, which allows only *events*, and no *properties* as ontologically basic, in contrast with *property realism*, the view that events are property instantiations (Kim 1976). The relation between these two issues is as follows: if one accepts property realism, then one gains a tool

for individuating concepts in a less fine grained manner than mode of expression. One can simply individuate concepts in terms of the properties they refer to. Conversely, if we would reject any mode of individuation beyond the mode of expression, we would be forced to reject property realism as well. In section 0 I presupposed property realism in my very definition of concepts, and I am not willing to give it up, for reasons beyond the scope of this essay (see for example De Muijnck 2002: especially chapters 10 and 11, 126-149).

We can now define the mode of practical bearing as the mode that individuates practical concepts in terms of the properties they refer to, where the properties in question are essentially functional roles, as I argued in section 1. Let us now call these *practical properties*. A practical property is the property of behaving in a certain way. Two practical concepts are equal in their mode of practical bearing if they refer to the same practical property.

With the two modes of individuation distinguished, we are in a position to account for CT and show that it does not imply underdetermination in any problematic way. Let us again take viruses as our example. Suppose that when the existence of viruses was first postulated in order to explain certain observations, a rivalling theory was developed as well which insisted that the virus hypothesis was false. Suppose further that proponents of this theory, in the long run, with all sorts of drastic revisions everywhere in the network of interrelated statements, are able to explain all observations that we will in the long run be able to explain. To use a phrase from Peirce, in the *ideal end of inquiry* there would be two communities of scientists, one sharing a network of beliefs including the belief that the virus hypothesis is true, and the other sharing a network of beliefs which involves regarding the virus hypothesis as one of the many ideas in the history of science that had to give way to other ideas.

In their modes of expression, the two total theories held by these communities are different, and in so far as they are different, they are underdetermined by the observations that both are able to explain. In particular, the statement 'viruses exist' when evaluated within the language of the one community will be true, while it will be false within the other community. However, the meaning of the statement differs in the two communities, because the second community has made all kinds of revisions in their network of beliefs in order to explain observations without having to accept the virus hypothesis. This means that on the

mode of practical bearing, the statement 'viruses exist' yields different propositions in the two communities. What it amounts to in the first community is that there are objects that instantiate the practical property  $P_1$  of producing certain effects. Which effects is fixed by other statements about viruses in the community, statements relating viruses to illness, electron microscopes and so forth. However, the second community has a different theory of microscopes and illness, at least different in the mode of expression. Therefore, in the second community the statement 'viruses exist' postulates the existence of a property  $P_2$  unidentical to  $P_1$  and in fact not existing at all. Furthermore, because of the different theories of the two communities, and because they explain the same observations, in the second community there might very well be a statement different from 'viruses exist' that, given its relations to other statements, would yield the practical bearing that there must be objects instantiating  $P_1$ . Such a statement would involve a concept that is different in its mode of expression from the concept of a virus in the first community, but equal to it in its mode of practical bearing because it refers to the same practical property.

Furthermore, this bizarre scenario would only work if the communities were kept separate, at least long enough to develop their theories independently from each other. If they were to make contact after that, interpretation would simply reveal that the word 'virus' as used by speakers from the first community could no longer be translated to the word 'virus' in the language of the second community. If, on the other hand, the communities were never separate at all practical considerations would have forced collapse of one of the two languages and standardisation of the other.

The conclusion we can draw from this scenario is that *if* the two communities would operate separately for a while and were then to meet again, they would, through interpretation, be able to understand each other (for their languages are not impenetrable conceptual schemes) and reach agreement that instantiations of the property  $P_1$  exist. Members of the second community, after having mastered the language of the first community in its matured form, would agree that within that language, the belief in the existence of  $P_1$  instantiations would be expressed as the belief that viruses exist. In this sense, pragmatists can be realists about viruses and reject the idea that this belief would remain underdetermined by our observations even in the long run.



## 6 Underdetermination and Phenomenality

The pragmatist response to the problem of underdetermination sketched above can be interpreted in two ways. According to what I shall call the *radical solution*, the very idea of underdetermination is nonsensical, there is no deeper nature or structure of the world beyond the reach of scientific observation and experiment, and thus neither can there be meaningful propositions about this structure or nature of which we would be unable to know the truth values. The radical solution is advocated by De Regt, and it may seem to follow naturally from the account developed in the previous section.

It is however also possible to give a weaker interpretation. According to what I shall call the *modest solution*, the pragmatist response to the problem of underdetermination only applies to the realm of practical concepts, but this realm need not be exhaustive of all concepts. On the modest solution, there may very well be fundamental propositions about the world that are underdetermined by scientific evidence. The modest solution only denies that statements in the microsciences *express such propositions*. According to the modest solution, underdetermination is not an unclear concept as such, but it becomes an unclear concept when applied to *practical concepts*, and the concepts of the microsciences, such as the concept of a photon or the concept of a virus *are* practical concepts. As shown by the line of argument from the previous section, propositions expressed by statements that employ practical concepts exclusively cannot be underdetermined by scientific observation.

Both solutions solve the problem of underdetermination, which was raised, not simply because of a worry about underdetermination as such, but because of a worry that we might no longer be rationally compelled to believe in the existence of entities postulated by our best scientific theories, even in the long run. On both solutions, the theories in question only state that certain practical properties are instantiated, and for this empirical evidence suffices.

Let us now confront both interpretations with the puzzle of phenomenality. I shall first examine and criticise the radical stance taken by De Regt. After that, I will argue in favour of the modest solution, and use it as a starting point in order to deal with the dilemma from section 4.

### *Against the Radical Solution*

De Regt believes that we should use the pragmatic maxim not only to get rid of the problem of underdetermination, but also to get rid of the hard problem of consciousness. In fact, he explicitly regards the problem of consciousness as a *case* of the problem of underdetermination:

Take a modern example from the philosophy of mind as an illustration (the zombie case). If there is no practical difference related to thinking that some creature is a human being with feelings and thinking that this creature is exactly like a human being with feelings except that it has no feelings, then the concept of having feelings is unclear. The intuition might be that there could nevertheless *be* a difference (the difference between having feelings and not having feelings), even if all scientific information is in. On what grounds would one like to argue that there is this “difference”? There are no clues in behavior or material structure to claim an absence of feelings. We treat the creature as a human being in all circumstances (De Regt *forthcoming*).

The argument seems to be as follows. Premise 1: let us accept the pragmatic maxim as a valid principle of individuation even in the study of phenomenality (the radical solution). Premise 2: the pragmatic maxim rules out any difference logically if there is no difference in behaviour or material structure. Premise 3: zombie world does not differ from the actual world in terms of behaviour or material structure. Conclusion: a zombie world that is different from the actual world is logically impossible.

The argument is valid, but it begs the question against Chalmers because he would never accept premise 1. This is easy to see if we translate the views again into my terminology of practical concepts. According to premise 1, our talk of phenomenality must obey the pragmatic maxim. Concepts that obey this maxim are practical concepts. Hence, premise 1 is a version of strong reductionism, which reduces phenomenal concepts to practical concepts. I have rejected strong reductionism and maintained that this rejection is largely an *intuitive* matter involving the idea that what something feels like simply *means*

something different from the way it behaves. But it is a *fact* that Chalmers *shares* this intuition and thus that the rejection of this intuition can never be a *premise* in an argument against his position.

I myself am not ready to sacrifice this intuition without separate argument to the conclusion that it is misguided, and De Regt does not offer such argument. The observation, that without the intuition the hard problem of consciousness would not have been raised, does not by itself constitute a good reason for rejecting the intuition. If we start thinking like that, we are no longer taking consciousness *seriously* (Chalmers 1996: xi). Furthermore, as noted before, there are some examples that bring out the intuition forcefully. If the pragmatic maxim were valid with respect to phenomenality, then knowledge of the behaviour and physical makeup of a bat would in principle allow us to know what it is like for the bat to be a bat; then Mary the superscientist who never experienced colours herself would be able to derive from her study of colour perception what it is like to see red (Jackson 1982); and then the very *idea* of qualia inversion would have to be logically ruled out once we have fixed all practical properties.

In each of these examples and thought experiments, and in many others, my intuition runs against strong reductionism. If there were no alternative to the radical solution, this would pose a serious problem for me, because that solution presupposes strong reductionism. But I have already hinted at a more modest solution, and as we shall see, this solution does not commit us to strong reductionism. In the remainder of this essay I will assume that the radical solution is mistaken, and develop a theory of phenomenality based on the modest solution.

### *The Modest Solution*

According to the modest solution, the statements in the microsciences that have been the focus of the debate between scientific realism and constructive empiricism are not underdetermined by scientific evidence, because these statements express propositions which refer solely to the instantiation of *practical properties*. However, the modest solution leaves room for other kinds of inquiry to refer to other kinds of properties, whose existence or nature may be underdetermined by scientific evidence. I propose to use this room to house phenomenality and avoid strong reductionism.

For example, the difference between the red quale and the green quale cannot be revealed to Mary the superscientist by reports of scientific observation and cannot be expressed in terms of practical concepts. For all Mary knows, the world might be spectrum inverted with respect to the way it actually is. Of course, the difference between this kind of underdetermination and the kind that Van Fraassen was after is that Mary is not able in principle to formulate different theories that are consistent with her evidence. In fact, she is not even able to formulate *one* such theory because she lacks the relevant phenomenal concepts. However, *after* Mary has first seen red and green, she might, and I think *should* agree that the *actual* colour spectrum is no more consistent with the knowledge she had prior to her first colour experience, as would have been a red-green *inverted* spectrum. In this sense, the matter of what colour spectrum is the actual human colour spectrum is underdetermined by all the scientific evidence about colour perception that Mary had access to before she had seen colours herself.

In the words of Chalmers, Mary's increase of knowledge upon her first experience of red "*narrows down the way the world is*" (Chalmers 1996: 373-374), and because it does, it involves factual knowledge: knowledge of new facts rather than just a new way of knowing facts already known in terms of practical concepts, as Churchland would have it (Churchland 1985, 1989b). Hence, there are *facts* underdetermined by evidence couched in practical concepts. And of course, just like Mary didn't know all the facts about human colour experience, so do we lack knowledge of facts about the experience of bats, and of many other animals, presumably. But whereas we might brush aside underdetermination in the case of Mary as merely historical, since she can resolve the matter when she experiences colour herself, the underdetermination of facts about bat experience by our evidence is *logical*: we can never determine these facts because we are in principle unable to acquire the phenomenal concepts of bat experience.

#### *Peirce on the Limits of Pragmatism*

The idea that phenomenality escapes the reach of the pragmatic maxim was already suggested by Peirce himself, notably using the spectrum inversion thought experiment to strengthen the underlying intuition:

I understand pragmatism to be a method of ascertaining the meanings, not of all ideas, but only of what I call “intellectual concepts,” that is to say, of those upon the structure of which arguments concerning objective fact may hinge. Had the light which, as things are, excites in us the sensation of blue, always excited the sensation of red, and *vice versa*, however great a difference that might have made in our feelings, it could have made none in the force of any argument. [...] My pragmatism, having nothing to do with qualities of feeling, permits me to hold that the predication of such a quality is just what it seems, and has nothing to do with anything else. Hence, could two qualities of feeling everywhere be interchanged, nothing but feelings could be affected (Peirce 1907 [1998: 401]).

Peirce's *intellectual concepts* are similar to my practical concepts. Peirce anticipates the modest interpretation of the pragmatic maxim in his claim that it does not apply to all ideas – apparently there are ideas that are not practical concepts, and these include, or are perhaps exhausted by, feelings. This brings us back to our dilemma. Phenomenal properties cannot be reduced to practical properties, and inversion of qualia would leave the practical situation unaltered. But dualism, as we have seen, leads to problems of its own. Let us now consider a passage where Peirce addresses the question of the relation between what he calls the ‘physical’ and the ‘psychical’ in terms of *laws*:

Then the question arises whether physical laws on the one hand, and the psychical law on the other are to be taken –

(A) as independent, a doctrine often called *monism*, but which I would name *neutralism*; or,

(B) the psychical law as derived and special, the physical law alone as primordial, which is *materialism*; or,

(C) the physical law as derived and special, the psychical law alone as primordial, which is *idealism* (Peirce 1891 [1992: 292]).

Option (A) in the passage above, which Peirce calls *neutralism*, is presupposed by Chalmers' naturalistic dualism. According to naturalistic dualism, we need to postulate phenomenal law *in addition to* physical law in order to be able to account for phenomenal experience. Naturalistic dualism is thus not only a property dualism but also a *nomological* dualism: it distinguishes two kinds of law, neither of which depends logically on the other. The view that physical law does not imply phenomenal law follows from the zombie thought experiment, and the view that phenomenal law does not imply physical law follows from Chalmers' thesis that phenomenality does not have a function.

Option (B) implies reductionism, because if phenomenal and physical properties were distinct, there would be no way in which a body of laws operating on the former kind of properties could be logically dependent on a body operating on the latter. However, if this rule applies to (B) it must apply to (C) as well, for the relation between (B) and (C) is clearly one of symmetry. So (B) and (C) are both species of reductionism, reducing in opposite directions. However, *so far* we have taken reductionism to imply the (B) type direction of reduction. We can thus consider (A) and (B) as the two lemma's of our dilemma. What is interesting is that Peirce has rejected both (A) and (B) and argued in favour of (C):

The materialistic doctrine seems to me quite as repugnant to scientific logic as to common sense; since it requires us to suppose that a certain kind of mechanism will feel, which would be a hypothesis absolutely irreducible to reason, – an ultimate, inexplicable regularity; while the only possible justification of any theory is that it should make things clear and reasonable.

Neutralism is sufficiently condemned by the logical maxim known as Ockham's razor, i.e., that not more independent elements are to be supposed than necessary. By placing the inward and outward aspects of substance on a par, it seems to render both primordial.

The one intelligible theory of the universe is that of objective idealism, that matter is effete mind, inveterate habits becoming physical laws (Peirce, 1891 [1992: 292-293]).

The basic idea is quite simple. If we cannot hold that phenomenality depends on the practical without getting into trouble, and we also cannot hold that phenomenality and the practical are mutually independent without getting into trouble, we should consider the view that the practical depends on the phenomenal.

Peirce tried to make sense of such a view using the concept of *habits*. He understood psychology as the study of habits and of the habit of contracting habits. From his idealist perspective, physics is not fundamentally different from this; it simply studies the habits of *the universe*, which turns the entire universe into something psychological.

Today we frown upon such a view, and for good reasons. There is a great deal involved in human habits that psychologists explain with reference to the cognitive, functional organisation of the human mind, which can be accounted for reductively with reference to the neural architecture that *implements* this functional organisation. Furthermore, human habits can be interpreted by ascribing *propositional attitudes* to human beings. Those propositional attitude ascriptions measure *intentional states* that are realised by the aforementioned organisation. In contrast, there doesn't seem to be a scientific basis for ascribing propositional attitudes to the *universe*, or for believing that there is something mental underlying the laws of physics that enjoys intentional states. To put it bluntly, we do not want to adopt the idea that the universe is *thinking*.

However, such far-fetched idealism is not at all implied by the general suggestion that the practical might depend on the phenomenal instead of the other way around. Rather than following the Peircean trail based on the concept of *habits*, we shall explore the idea that the practical depends on the phenomenal in terms of the concept of *supervenience*.

## 7 Affective Monism

In this section, I will focus on the concept of supervenience, restate some of the aforementioned views and problems in terms of supervenience, and present my view of affective monism.

The concept of supervenience expresses a dependency between two variables or groups of variables, where those variables can be facts, properties, events, laws or statements, depending on metaphysical taste and philosophical purpose. Chalmers defines the general idea in terms of properties:

B-properties *supervene* on A-properties if no two possible situations are identical with respect to their A-properties while differing in their B-properties. (Chalmers 1996: 33)

He allows for different varieties of supervenience by distinguishing different readings of 'possible' and of 'situations'. Situations can be 'individuals' or 'entire worlds', yielding *local* and *global* supervenience respectively. In what follows, when I talk of supervenience without qualification in this respect, I will be speaking of global supervenience. The second distinction is between *natural* and *logical* possibility, yielding natural and logical supervenience respectively. This distinction plays a central role in Chalmers' discussion of reductionism and dualism. Natural possibility requires that a situation does not violate the laws of nature, whereas logical possibility requires only that a situation can be conceptualised without contradiction.

#### *Natural Supervenience and Naturalistic Dualism*

In section 4 we have seen that the role of natural laws is crucial in the debate on phenomenality, because if we need to postulate natural laws in order to account for the correlation between phenomenal and practical properties, then the properties must be distinct and naturalistic dualism must be true. Thus, the only thing that is required to establish naturalistic dualism is an argument to the conclusion that the relation between the phenomenal and the practical involves laws of nature. I have loosely presented Chalmers' argument to this effect, based on the logical possibility of zombies and inverted qualia. Let us now look a bit closer at the argument, and see how it involves the concepts of natural and logical supervenience. As I understand it, the argument can be summed up as follows:

- (1) phenomenal properties supervene on practical properties;
- (2) supervenience is either logical or natural;
- (3) a world which differs from ours in terms of phenomenal properties without differing in terms of practical properties is logically possible (follows from the zombie and inverted qualia thought experiments);



- (4) hence, phenomenal properties do not supervene *logically* on practical properties (follows from (3) and the definition of supervenience);
- (5) hence, phenomenal properties must supervene *naturally* on practical properties (follows from (1), (2) and (4));
- (6) thus, there must be natural laws that connect phenomenal properties to practical properties (follows from (5) and the definition of natural supervenience).

The argument is valid: conclusion (6) follows from the premises (1), (2) and (3) and the definitions of logical and natural supervenience. Premise (1) is generally accepted within contemporary philosophy of mind, except perhaps for its implicit property realism. However, I have discussed property realism earlier on and decided to accept it for independent reasons.

Premise (2) has drawn some criticism. To deny (2) is to hold that there is a third variety of supervenience besides logical and natural supervenience, and thereby, that there is a third meaningful notion of possibility aside from logical possibility and possibility in the light of the laws of nature. In particular, some have tried to construe a notion of *metaphysical* possibility that was supposed to be stronger than logical, but weaker than natural possibility. Chalmers argues forcefully against this idea on the basis of his 'two dimensional' view of intensions (Chalmers 1996: 131-140). Without further discussing or fully endorsing his view at this point, let me say that I cannot really make sense of any notion of possibility 'halfway' between natural and logical possibility, and that I believe Chalmers has come a long way, and maybe even all the way, to showing why. Let us assume that premise (2) is true.

I have discussed premise (3) in length already. In section 4, I have concluded that not all phenomenally different situations are logically possible if they do not differ practically as well, but that *some* are. In particular, a situation different in terms of affectivity without practical and qualitative difference is logically *impossible*, but mere qualitative difference without practical difference, and with the affective compensated vis-à-vis the qualitative but unchanged in relation to the practical, is logically *possible*. Thus, I must admit that *some* worlds differing in phenomenal properties without differing in practical properties are logically possible. And that is enough to establish proposition (4).

We can reformulate our dilemma as follows. Proposition (4) must be true because some phenomenal inversions are logically possible. On the other hand, conclusion (6) must be false because some other phenomenal inversions are logically impossible. But (6) follows from (4) on the premises (1) and (2). I have accepted (2) and I have accepted the property realism implicit in (1).

I see only one way out. We must simply reject (1) altogether. If property realism is true, premise (2) is true, proposition (4) is true, and yet conclusion (6) must be false, then we must simply *deny that the phenomenal supervenes on the practical*. On a first glance, this may seem to lead to an even more drastic, perhaps interactionistic, form of dualism than that of Chalmers. But dualism is not the only option. Instead, I am going to develop the view that *the practical supervenes on the phenomenal*. In particular, my proposal is that practical properties supervene logically on phenomenal properties. Thus, I will be defending something reminiscent of Peirce's idealism, cast in terms of the framework of supervenience. I realise that nowadays, this is a very unorthodox move to make. But as we shall see shortly, the idea has some great advantages. In particular, it will solve the dilemma from section 4. And although it may sound highly counterintuitive, perhaps downright absurd, at first, we can explain the absurdity away and avoid the features that have made classic versions of idealism, such as that of Peirce, implausible. In order to get there, we are going to turn the established framework of supervenience upside down. Literally.

### *Sketch of the Theory*

Let me start by giving a brief overview of my theory of affective monism. The central hypothesis is that all properties, their instantiations, and any laws governing those instantiations supervene logically on phenomenal properties, their instantiations and laws governing those instantiations.

Furthermore, just like supervenience relations obtain between practical properties at different layers of organisation, so do supervenience relations obtain between phenomenal properties at different layers. These layers are not collectively situated hierarchically 'below' the layers of practical properties, but are rather 'on par' with practical properties in the sense that any layer of practical properties supervenes on a corresponding yet richer layer of phenomenal properties. A practical layer consists of logical consequences of a phenomenal

layer that can be accessed from the third person perspective. But there is more to the phenomenal layer than its practical consequences, and part of what is beyond these consequences in the 'upper' layers can be accessed from a first person perspective, namely when a certain organisation of phenomenal property instantiations realises a cognitive system that implements such a perspective. Human beings are such systems, as are many other animals, presumably.

However, phenomenal properties at lower levels of organisation are beyond such access, and phenomenal properties at the lowest levels are in principle precluded from being captured by our concepts. For if we could, we would in principle be able to derive, for any organism, what it is like to be that organism if we had a suitably detailed description of its micro-organisation in phenomenal terms. In fact, the belief in the existence of phenomenal properties on which microphysical practical properties supervene is based solely on an analysis of philosophical issues concerning phenomenal quality, phenomenal affectivity and mental causation.

This analysis is at the heart quite simple. If we cannot give up the belief that phenomenal affectivity plays a functional role by *motivating* us to action, and we cannot give up either the belief that some of the practical effects of phenomenal affectivity supervene logically on microphysical practical events, then, given that phenomenality does not itself supervene logically on practical events, we must adopt the belief that phenomenal affectivity is somehow already at work on the microlevel, and that practical microphysical properties supervene *on it*. This leads to a metaphysical picture of the world as universally governed – in a sense, *motivated* – by phenomenal affectivity. Which is why I call this position *affective monism*.

One might also wish to label this idealism, or panpsychism, but it is important to understand that affective monism does not postulate *ideas, thoughts* or *representations* at the microlevels of properties constituting our universe. For these would all require implementation in cognitive systems, and subatomic particles are not cognitive systems. In other words, I am not turning the *mental* into an unexplainable, because fundamental, phenomenon; I am not a believer in a cosmic conscious agent. The universe is not thinking.

I am also not questioning functionalism, or cognitive psychology, about mental capacities. The only thing I am claiming is that whatever it is that gives intentional states

realised in a cognitive system their *phenomenal* dimension must already be present in the structure that realises such a system, yet must escape our description of that structure in terms of practical concepts.

Finally, postulating phenomenal properties at the 'ground level' as a supervenience base for everything in the universe allows us to explain the logical possibility of qualia inversion and of zombie world in terms of *multiple realizability*. If phenomenal properties are realisers of practical properties, then we can simultaneously be monists about the ontological relation between these properties in our world, and yet maintain that it is logically possible for the same practical properties to be realised differently, by a supervenience base featuring some qualia inversions, or by a supervenience base featuring nonphenomenal properties, or for them to not be realised by anything more fundamental at all. However, because affectivity is involved in the realisation of causal relations between practical events in a way that crucially involves the direction of its range from positive to negative, we can also explain that changes in affectivity in the supervenience base logically imply changes on the level of realised practical properties. This will rule out the case of supermasochism. Thus, affective monism harbours all our intuitions about these matters and thereby frees us from the dilemma from section 4.

#### *Affectivity, Motivation, Causation*

While I am writing this, I am listening to an album by a band called *Pain of Salvation*. Just before I started on this paragraph, I had paused briefly and walked to my stereo to tune up the volume in order to fully enjoy a brilliant guitar solo that I knew was coming up. My action was caused by a combination of the two ways in which taste can play a role as distinguished in section 2. Both my actual experience of the onset of the solo as well as my awareness of what was going to come, since I know the music by heart, were accompanied by a high level of positive affectivity. It is in virtue of this affectivity that I was motivated to interrupt my work for a minute, walk to the stereo, and tune up the volume.

If supermasochism were logically possible, then we would have to deny that it was in virtue of the affectivity. For then it would make sense, at least conceptually, to say that my supermasochist twin experienced both the onset of the solo as well as his memory of how the solo continues in a highly negative way affectively, and then was moved on somehow, with

no *other* differences compared to the actual situation, to interrupt work, walk to the stereo, tune up the volume, and experience an increase in *agony* as a result (as in my case the pause from work and the higher volume increased the *pleasure* I got from the music). Affectivity, or at least the *sign* of the affective value, would then be conceptually independent from the fact that I was moved to act.

In section 4 I have made it clear that, in contrast to the naturalistic dualist, I am not going to bite this bullet. I want to maintain that it is really in virtue of positive affectivity that I was moved to action. As we have seen, this connection between affectivity and action is locked inside the concept of *motivation*. Let us now see how affective monism can make sense of this concept, and how the supervenience picture of affective monism rules out supermasochism.

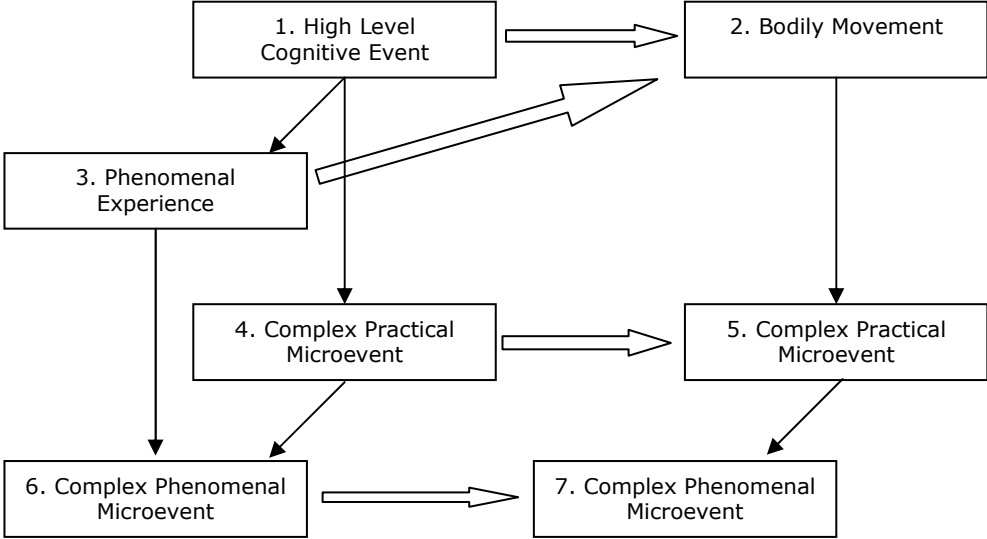


Figure 2-1  
 Bodily movement caused by phenomenal experience according to affective monism.  
 Downward arrows depict logical supervenience, rightward arrows causation.

Have a look at figure 2-1. The boxes have abstract labels, but can be read so as to apply to the *Pain of Salvation* example. The boxes refer to *events*, property instantiations at times. I will refer to these events using the numbers of the boxes. The events (1), (3), (4) and (6) occur at the same time, and (2), (5) and (7) occur at a same, later time. (3) is my phenomenal experience of the guitar solo, which contained an affective experience of the first notes of the solo as I was actually hearing it, an affective experience of my memory of the entire solo, and

as a result of this, a further affective experience of desire to enjoy the piece of music fully, and to hear it louder than I was playing the CD at the time.

(1) is a cognitive event that corresponds to (3) in the sense that it consists of what *logical consequences* of (3) can in principle be captured in *practical concepts*. Thus, whereas (3) instantiates a phenomenal property, (1) instantiates a practical property. I shall refer to phenomenal property instantiations as *phenomenal events* and to practical property instantiations as *practical events*. We can describe the property instantiated by (1) as a property of being in a state of desire, but in contrast to the phenomenal property, this time we would be using the notion of desire in a purely practical sense, namely as a state which tends to cause me to manipulate my environment in such a way as to cause in me the belief that the volume of the music has reached a certain level. For the sake of clarity, I will be speaking of *phenomenal desires* and *practical desires* in order to honour this distinction.

Note that the practical desire instantiated by (1) is very abstract, as it does not carry implications on the specifics of its biological realization. This is as it should be, because (3) does not carry such implications either. Nevertheless, (1) contains enough to explain (2), the event of my walking to the stereo and increasing the volume.

At this point, I am going to make three assumptions about *causation*. The first is that this kind of explanation signifies a causal relation. In other words, from the fact that (1) explains (2), I will infer that (1) *causes* (2). The second is that for any cause *X* of an effect *Y*, if *X* supervenes logically on underlying property instantiations, then we can in principle refer to some complex configuration *Z* of such instantiations underlying *X* that is narrow enough to be considered a cause of *Y* as well. The third is that if *Z* is a cause of *Y* in this way, this does not negate *X*'s being a cause of *Y* too.

These assumptions can, and have been, disputed. Showing that logical supervenience obtains might not be enough to remove all philosophical worries about conscious causation. However, remaining worries have to wait for another occasion. In our discussion, the problem has been that we could not establish logical supervenience in the first place. Chalmers has maintained that even in the absence of logical supervenience, phenomenality might have a causal role, but we have seen that this would be a peculiar causal role, one that after all would not deliver a *functional role* for phenomenality, and would therefore not fully dispel the image of epiphenomenalism. In the context of *that* issue, it would already be an

enormous improvement if we could show that the phenomenal and the practical are tied to each other by means of logical supervenience after all. Any further worries beyond this improvement are beyond the scope of this essay. So let us make these assumptions for the sake of the argument.

According to affective monism, (1) supervenes logically on (3). I cannot phenomenally desire to turn up the volume of the music without also *practically* desiring to do so. In terms of our definition of supervenience, there is no logically possible world that is different with respect to (1) without also being different with respect to (3). Note that this view follows directly from the most intuitive conception of phenomenal affectivity: that it can have a positive and a negative value, and that the positive motivates the agent to realise or sustain its content whereas the negative motivates him to prevent or undo the content of the experience.

This intuitive idea has a less intuitive consequence however, and that is that there must be something like *microphenomenality*, a conclusion that I think we should nevertheless embrace. Figure 2-1 illustrates how this consequence follows. The bodily movement of (2) obviously supervenes on microlevel property instantiations in virtue of the mereological organisation of the body. Hence, there must be a complex microevent consisting of the instantiation of a specific pattern of the practical properties that low level science, let us say microphysics, talks about. This is event (5) in the figure.

If we assume that the domain of microphysical practical properties is causally closed, then there must be a cause of (5) on the level of microphysics underlying (1). This is (4) in the figure. The relation between (4) and (1) is one of functional realisation, which guarantees logical supervenience. So (1) supervenes logically on (3) as well as (4). However, as I have argued at length now, phenomenal properties cannot supervene logically on practical properties. Thus, the fact that (3) and (4) both underlie (1) cannot be explained by invoking the idea that (3) supervenes on (4). And the reverse, that (4) would supervene on (3) would obviously violate the relations of mereology and abstraction in the figure. As we have seen, both (3) and (1) abstract away from the specifics of their implementation at the microlevel.

Affective monism provides a way out by introducing (6), a *phenomenal microevent*. On this event, both (3) and (4) supervene logically. In my view, this is the only possible road left to us to explain how (1) could supervene logically on (3) as well as (4). Of course, the idea of

microphenomenality raises many questions that demand further philosophical inquiry. But for now, let us assume that the idea makes sense, and see how it fits into the model.

We can simply define (6) as the instantiation of a property that contains everything it needs to contain in order to make it logically necessary for both (3) and (4) to occur. In particular, it contains the ‘motivational force’ that underlies both the phenomenal experience of affectivity as well as the causal power of a practical desire. The best way to understand (6) is to focus on (4) and adopt the view that (4) is not ontologically primitive, but rather that it is a *causal signature* of that which *is* ontologically primitive, namely (6). This is the big difference between affective monism and naturalistic dualism. On naturalistic dualism, the universe contains, at the most fundamental level, both practical and phenomenal property instantiations, which are equally primitive and logically independent of each other. On affective monism, practical property instantiations at the microlevel are not primitive and independent; they are rather a *part* of the phenomenal microevents.

So what I am defending is that the microstructure of the universe as revealed by microphysics is not a primitive ontological structure, but a signature of a richer ontological structure, and that it is out of this richer structure that phenomenality arises.

In the figure, this structure consists of (6) and (7). Between these, there is a relation of causation, which is a result from the motivational character of (6). Note that (6) and (7) are not simple events, in the sense that they consist of single primitive property instantiations. Rather, they are immensely complex formations of primitive property instantiations, organised in a wonderful way so as to realise (1) and (2). According to affective monism, when organised so as to yield certain cognitive events, such as (1), the phenomenal properties in an event such as (6) will add up so as to produce a lively experience for a conscious being. Ontologically speaking, this experience – (3) in the figure – is nothing over and above (6), and since it captures everything from (6) that is needed to establish (1), a cause of (2), it makes perfect sense to say that (3) is also a cause of (2). That is how phenomenal affectivity causes action.

#### *Some Questions and Suggestions on How to Answer Them*

The obvious problem for any version of panpsychism is that it seems to violate the commonsense idea that some things are conscious whereas others aren’t. A philosopher who



is ready to accept that inanimate objects such as tables and chairs might be conscious deserves a nomination for the bullet swallowing awards. Even *people* are sometimes unconscious, after all. Thus, the panpsychist is forced to make a distinction between the kind of phenomenality that is omnipresent and the kind that obtains only in truly conscious agents. The first kind would then figure as a necessary, but not sufficient condition for the second. This idea is sometimes advertised as ‘panprotopsychism’.

But is this a valid move? For what does it mean to say that something is a necessary condition for consciousness in a way that goes beyond the boundaries of materialism if you cannot explicate this idea any further than saying that it is necessary? Isn’t that just an *ad hoc* postulation of something that exactly fits the description of our explanandum without adding any substantial idea as to *how* the explanandum – consciousness – is realised?

These questions might also be directed to me. Given that my claim is that every event supervenes logically on a microphenomenal event, and that microphenomenality is pretty much defined as ‘the thing that does what materialism couldn’t explain’, then I can see why you might think my proposal is rather cheap.

In response to this critique, I would need to show you that you are judging my position in the wrong way. The purpose of affective monism is not to explain macrophenomenality in terms of microphenomenality, and if you aren’t in the business of explanation you cannot be properly criticised for being *ad hoc*. In fact, I am not even trying to explain microphenomenality itself. Rather, I am claiming that it is in principle impossible to have phenomenal concepts of microphenomenal properties because they lay beyond the boundaries of experience.

Of course, you might object that it doesn’t make sense to call a property phenomenal if it cannot be experienced. We should avoid a terminological debate on this matter, and if it makes you feel better I am willing to restate my view as the claim that there are ‘microprotophenomenal’ properties or something to that effect. Before you roll your eyes, review the arguments from the previous sections again. The point of affective monism is not that it offers microphenomenal – or microprotophenomenal – concepts, the point is that only by assuming that there exist microphenomenal properties of which we can have *no* concepts can we arrive at a consistent supervenience framework that harbours our intuitions about mental causation. In other words, the claim is that the *concept of microphenomenality* makes

sense even though we cannot have *microphenomenal concepts*. It might be interesting for further inquiry to build a more elaborate version of this argument and relate it to 'mysterian' positions such as that of Colin McGinn.

If microphenomenality is not a sufficient condition for consciousness, then what is? This is where I think more popular theories of consciousness can be of use. Suppose, for the sake of the argument, that some version of the global workspace model would yield accurate correlates for consciousness. In that case, the affective monist might hold that only when microphenomenal properties are instantiated in a complex structure such that it realises a global workspace system on the level of macropractical properties do these microphenomenal properties 'add up' in such a way as to realise the *phenomenal functional* modules of PRS, PPM and taste. It would be interesting to see how (and if) these three modules can be squared with different actual third person perspective functional models of the practical correlates of consciousness.

Even without microphenomenal concepts, it might be possible to make the proposal that there is affectivity behind microphysics more concrete by trying to connect the philosophical idea of the 'motivational pull' of affectivity to physical concepts. If all physical processes are governed by some kind of proto-affectivity, then perhaps there must be some dynamic that different physical processes that obey different physical laws have in common.

In this context, I find the physical concept of *entropy* particularly interesting. According to the second law of thermodynamics, the degree of 'dispersion' of energy, measured using the concept of entropy, tends to increase if not hindered to do so. In fact, entropy will never decrease unless this decrease is only local and related to a larger entropy increase elsewhere. Thermodynamics was originally applicable to 'ideal gases' only, but the concept of entropy and the second law of thermodynamics have been redefined so as to apply to physical processes *in general*. According to some, this law should not be understood as *a* law of physics but as *the* law of physics, as a structural dynamic that works *across* the other physical laws.

Another interesting feature of this law is that it is *time-irreversible* whereas other physical equations are *time-reversible*. The second law of thermodynamics introduces the *direction of time* into the physical vocabulary. Needless to say, the idea that time has a direction is implicit in our phenomenology and our concepts of affectivity, motivation and

causation. Affectivity is future-directed, it can motivate what *will* happen. Causes *precede* their effects. Conceptual analysis of entropy and affectivity might offer ways to make the supervenience relation between microphenomenality and microphysics more concrete by specifying how the ‘motivational pull’ of microaffectivity might map onto entropy-increase relations between events at different points in time.

And let us also not forget that the interpretation of microphysics at its most basic levels – those of quantum field theory and, perhaps, string theory – is still very much in the open. Some interpretations already involve references to conscious activity, and some theories in the consciousness debate explicitly draw on quantum mechanics. Chalmers has explored some such proposals in his 1996 book, including some varieties of panpsychism. It would be interesting to investigate how affective monism might figure in this fascinating area of philosophical debate.

## 8 Summary

What does phenomenality do, and what does that mean? Phenomenality motivates our behaviour ‘from the inside’ in a way that to some degree escapes description ‘from the outside’. And what that means is that we cannot be materialists or dualists and that we have to be some kind of idealists. In this essay, I have discussed contemporary arguments about the function of phenomenality, trying to steer clear of the rocks of strong reductionist materialism on one side, and epiphenomenalist dualism on the other, because either sits at odds with important intuitions that are brought out by the thought experiments of spectrum inversion, Mary the colour-deprived neuroscientist, zombies, and supermasochism. In search of a third alternative, I have reviewed the idealism of C.S. Peirce, and finally proposed the view of *affective monism*, a reconstruction of idealism in terms of the supervenience framework of Chalmers. I have argued how this view can solve our problems, and finished with some suggestions about how to approach the new problems that this position might raise.

## References

- Block, N., Flanagan, O., and Güzeldere, G. (eds.) 1997, *The Nature of Consciousness: Philosophical Debates*, Cambridge MA, MIT Press.
- Chalmers, D.J. 1996, *The Conscious Mind*, Oxford, Oxford University Press.
- 1997, “Availability: The Cognitive Basis of Experience?”, in Block, Flanagan, and Güzeldere, 421-424.
- Churchland, P.M. 1982, “The Ontological Status of Observables: In Praise of the Superempirical Virtues”, reprinted in id. 1989a, 139-151.
- 1985, “Reduction, Qualia, and the Direct Introspection of Brain States”, reprinted in id. 1989a, 47-66.
- 1989a, *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA, MIT Press.
- 1989b, “Knowing Qualia: A Reply to Jackson”, in id. 1989a, 67-76.
- Davidson, D. 1986, “A Nice Derangement of Epitaphs”, in E. LePore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford, Basil Blackwell, 433-446.
- Dennett, D.C. 1988, “Quining Qualia”, reprinted in Block, Flanagan, and Güzeldere 1997, 619-642.
- De Muijnck, W. 2002, *Dependencies, Connections, and Other Relations: A Theory of Mental Causation*, Ph.D. Thesis, Tilburg University.
- De Regt, H.C.D.G. 2002, “Implications of Inquiry: A Pragmatist View of the Scientific and Manifest Image”, in S. Clarke & T.D. Lyons (eds.), *Recent Themes in the Philosophy of Science*, Kluwer Academic Publishers, 175-192.
- forthcoming, “The Functional View of Science: Nozick and the American Pragmatist Tradition”.
- Jackson, F. 1982, “Epiphenomenal Qualia”, *Philosophical Quarterly*, 32, 127, 127-136.
- Kim, J. 1967, “Events as Property Exemplifications”, reprinted in id. 1993, *Supervenience and Mind: Selected Philosophical Essays*, New York, Cambridge University Press, 33-52.
- Loar, B. 1990, “Phenomenal States”, reprinted in Block, Flanagan, and Güzeldere 1997, 597-616.

- McGinn, C. 1989, "Can We Solve the Mind-Body Problem?", reprinted in Block, Flanagan, and Güzeldere 1997, 529-542.
- Nagel, T. 1974, "What Is It Like to Be a Bat?", reprinted in Block, Flanagan, and Güzeldere 1997, 519-527.
- Peirce, C.S. 1878, "How to Make Our Ideas Clear", reprinted in id. 1992, 124-141.
- 1891, "The Architecture of Theories", reprinted in id. 1992, 285-297.
- 1903, "The Maxim of Pragmatism", reprinted in id. 1998, 133-144.
- 1907, "Pragmatism", reprinted in id. 1998, 398-433.
- 1992, *The Essential Peirce, Vol. 1*, Bloomington, Indiana University Press.
- 1998, *The Essential Peirce, Vol. 2*, Bloomington, Indiana University Press.
- Quine, W.V.O. 1953, "Two Dogmas of Empiricism", in id., *From a Logical Point of View*, 2nd. rev. ed., New York, Harper & Row, 20-46.
- Shoemaker, S. 1982, "The Inverted Spectrum", reprinted in Block, Flanagan, and Güzeldere 1997, 643-662.
- Van Fraassen, B.C. 1980, *The Scientific Image*, New York, Oxford University Press.