On the Explanation of Mind

John Mc Geever

Ph.D.

Department of Computing Science and Mathematics

University of Stirling

Submitted in partial fulfillment of

the degree of Doctor of Philosophy

October 1999

Acknowledgements

Thanks to Leslie Smith, the rest of the Computing Science dept, and the ex–members of the CCCN. The University and the Computing Science department provided funding for this project. M. Elton deserves thanks for showing me how difficult it is to go head–to–head with contemporary philosophers who have abandonded ‘old fashioned’ notions of empirical validity, epistemological justification, and ontological grounding.

Thanks most of all to friends who aren’t in any way associated with the university, the department, or what I did here. They provided a nicely separate world away from the invented world of academia to which I could escape.

M S and R deserve special mention, as does L who showed me something real at a point where I thought it wasn’t. I ought to point out that this is all Darragh Smyths fault.

And I’d like to thank my parents, our parents, those beings that crawled, in the spirit of scientific enterprise and discovery, from the sea into the trees (and later that descended from the trees) all those years ago. Without you, I wouldn’t be here, I’d be in the sea, like the dolphins, having a whale of a time.

Declaration

This thesis is my, John Mc Geever's, work except where indicated in the referring to the work of others.

Abstract

Open any book on philosophy of mind or cognitive science and there will be found a statement of functionalism. It will say that function, functional state, or functional role are all that is necessary to determine, fix, or explain something else (the 'something else' can be many things: mental content or phenomenal experience, for instance; I focus on the latter, but the distinction isn't clear). What does this mean? The meaning it can be given is subtle. And the only meaning it can have indicates that there isn't *just* function. Function may be all that is necessary, but 'function' will have to mean more than *just* functional role. But functionalism wishes 'functionalism' to mean *just* functional role, so there is a difficulty for functionalism and computationalism. This thesis examines that difficulty.

In philosophy of mind and cognitive science there is a view in which a certain explanation explains all function and all behaviour, but it is acknowledged that there are things that such explanations do not explain. So those things must be non behavioural, and have no functional role. They do nothing at all, but we refer to them in explanations. There are difficulties which arise from this.

And certain explanations are seen to explain everything, yet leave something out. But that which is left out is seen as being 'nothing but' something else, or 'merely derived' from something else; thus is avoided the problem of things which are not explained, because they are not 'real'. If they are merely derived, then their status isn't 'real', and so they are not really counterinstances to the complete explanations. But this has a problem; and the problem is this: if they are 'nothing but', they are as real as what they are 'nothing but', and if they are declared 'unreal', they must be considered as separate and further facts, and so are real. So what meaning does 'merely derived facts' have? Why does this problem occur? It occurs because functionalism is accepted in the sense of 'just' function, and this has no meaning. So we cannot talk about 'merely derived' or 'higher level' facts as somehow lesser than 'privileged' facts.

What is the difficulty? The difficulty is this: functionalism is a strong abstract Platonistic view. But functionalism would not admit to this, and so it requires something in addition to just functional role. So there is that which has a function, and functionalism implicitly refers to it; and that is ontological, and so ontologies cannot be so easily eliminated or ignored (and cannot be declared 'merely derived'), while considering 'just' the function, computation, or behaviour.

Contents

Chapter 1

Introduction

1.1 Overview

The introduction chapter orients this work within the terminology of cognitive science and philosophy of mind, with special relevance to functionalist views.

The second chapter describes the important distinctions between functionalism and physicalist functionalism. It describes what will be called 'radical functionalism', a form of functionalism common to cognitive science more so than philosophy of mind. Radical functionalism is a 'pure' form of functionalism without the further ontological constraints that a physicalist functionalism may have. It also has no strong constraints on what realises functional organisation. It is a liberal functionalism. The difficulties with regard to the issue of realising base and the difficulties with assigning functional role are examined.

The third chapter assumes that a form of radical functionalism is the case. The argument is a reductio, which follows radical functionalism to some contradictions. It demonstrates the need for a physicalist functionalism. The particular necessary requirements of such a physicalist functionalist view are considered. The way is then open to consider physicalist functionalism.

Chapter four considers an inessentialist view that follows from an acceptance of functionalism in any form, including the physicalist functionalism arrived at through the difficulties of radical functionalism in the third chapter. The essential difficulty with inessentialism is described, and some ways of dealing with this difficulty are suggested. Inessentialism is assumed to be the case, and a contradiction follows. The argument is a reductio, and it points to the precise difficulty with inessentialist views. It is noted that none of these difficulties result if a strongly eliminativist view is held.

The difficulties with inessentialist views are compatible with a form of monism, if a non-hierarchical view of levels of explanation is accepted. This requires that a 'complete' view of behaviour and functional explanation be dropped. The monist view resulting is akin to the anomalous monism of Davidson (see (Davidson 1980)) in one respect, namely the lack of strict psycho-physical laws. The various criteria for a view such as this, which avoids the inessentialist difficulties, are described in chapter five.

If eliminativism is not opted for, and the conclusions of the previous chapters are accepted, then there remains one difficulty. This difficulty is an aspect of all phenomenal realist views, but it is argued that the difficultly results from a semantic confusion. The difficulty assumes a certain question to be answerable in two ways, which conflict. It is argued that the two ways in which this question can be answered are not comparable, and thus no conflict results.

The contributions of this thesis are firstly to show that radical functionalism is incoherent and secondly to show that inessentialism is incoherent. Physicalist functionalism is preferable over 'radical' functionalism, but inessentialism is an aspect of physicalist functionalist views. It is argued that there are no problematic implications following from this, as long as a monist view, without a privileged level of complete description, and without strict psycho-physical laws, is held.

In practical terms, Dennett's Cog[1] can never be what it is intended to be eventually. Cog is the practical side of a radical functionalist view.

1.2 An issue

1.2.1 The question

I see a red rose. You see a red rose. At least we know that the terms we use to refer to redness is the same. So we assume that the experience is the same: same term, same referent. We are also assuming that there is something that is a referent. You and I both say 'red', and we both believe that redness, the experience, is the same (or suitably similar) in both cases[2].

An explanation of the brain ought to explain both how I see and what I see when I gaze upon a red rose. It ought to explain this: that 'I' 'see' a 'red' rose, which appears, to me, to be out there and bright red. That is my experience, my personal point of view. It may be illusory, it may be non existent, it may be a side effect of something else. Nevertheless,

[1] Cog is a robot in the continual process of being built at 'the Cog shop', MIT. It has basic visual motion detection abilities and tends to grasp for moving objects with its arm. It also has other sensory motor abilities. It attracts a significant amount of popular science media hype: "renegade conscious robot at MIT" (Popular Science, June, 1995, p 88); "A rudimentary 'pain' mechanism is built in to stop Cog punching itself in the eye" (The Times Higher Education Supplement, June 3, 1994. p. 16); "birth of a human robot...the robot that wants to be human" (New Scientist, May 14, 1994, p 26–30). See (Dennett 1995). As an engineering project it is intriguing. It is to be noted that the Cog team acknowledges this hype to be hype.

[2] On The question of 'sameness': there are some that search for a fact of sameness, and others that deny that there is a fact of sameness, or deny that there is the specific 'thing' or 'fact' to which 'sameness' is applied.

this illusion, side effect, or non-existent thing needs to be explained.

This is the issue of phenomenal experience, or consciousness, or qualia. It is because there are different views on what ‘redness’ means, that there are differing views on how redness is explained. Philosophy of mind and philosophy of language are related in this regard. We talk about philosophy of mind in language, and our use of language arises from the mind.

I experience red, I say ‘red’, and you do the same, but your experiences are yours, and not mine. We can share experiences in the sense of having co-occurrent experiences. Still, I assume that red for you is red for me, because your word ‘red’ is the same word I use. In Wittgenstein’s box there may or may not be a beetle[3]. He never opens it to anyone, so we only have his word on it. So too for my beetle box. We both use the same words to refer to our beetles, but the beetles may not be the same, or there may not be any beetles.

There may not be a beetle in the box. Whether the beetle is representative of meaning in language, or our personal experiences, the issue is the same. We open our own box. So how are we sure about the contents of the boxes of others? Private language, personal experiences, other minds: these issues will come to the fore once explanation starts.

What is to be explained? Seeing a red rose? The claim that I see a red rose, or the belief or claim that the rose has redness? Alternatively, perhaps the *claim* that I experience redness when viewing a rose is to be explained. The explanation could take many forms. It could describe how the rose interacts with me to produce an experience of redness. It could state how the rose interacts with me to produce dispositions to behave in a way that results in the verbal claim, “I see a red rose”.

Redness cannot be denied. It is part of the collection of phenomenal qualities, of experiential qualities, that are our existence. Those qualities are an important aspect of what we seem to know. There is little point in ‘eliminating’ these qualities. (If someone attempts to, replace his or her colour television with an old black and white model). However, what we think they are can be eliminated, if what we think these qualities are is not, in fact, what they are. There is ‘redness’, but it may not be separable from objects that are red, or persons who see red. It may not be separate from the claim of redness that persons make. There is ‘redness’, but what this term refers to may be quite different from our intuitive understanding.

Using the term ‘redness’, in the basic sense that we all understand that term, has an epistemic aspect. It supposes that there is meaning in the term ‘redness’, in the sense that we all understand, even if that term refers to something which may in actuality be something quite different from our basic understanding. There is also an ontological

[3] See (Wittgenstien 1953, 293) . I use the beetle-in-the-box analogy as Wittgenstein argued that, “if we construe the grammer of the expression of sensation on the model of ‘object and name’ the object drops out of consideration as irrelevant”. Our intuitive concepts of phenomenal experience may similarly fall out of consideration.

aspect. This is not concerned with how we know that referring to 'redness' is justified. It is concerned with the question of what 'redness' is. The ontological and epistemological aspects are not so easily distinguished, however, and creating a clean distinction without due care can cause unnecessary difficulty.

Saying that there are phenomenal properties to be explained assumes that we are justified in taking phenomenal properties as something to be explained. Nevertheless, this is not the assumption that there is something of a specific type to be explained. This may not be the case. Referring to phenomenal properties is not begging the question, as they may be something quite different from our intuitive understanding.

There are accounts of mind that argue that there is something to which phenomenal experience refers. There are phenomenal realist accounts that argue that the term refers to that to which it seems to refer: phenomenal experience is actually phenomenal experience (for instance, the views of Searle and Chalmers (see (Searle 1992), (Chalmers 1996a)), who are explicit in claiming that phenomenal experience is what it seems to be, to our intuitive understanding). These accounts argue against the possibility that phenomenal experience refers to something quite different from our basic understanding. In these accounts, phenomenal experience is treated as fundamental and irreducible.

To our basic understanding, the phenomenal experience we refer to is not a theoretical entity. It seems unlike the phlogiston of old; it seems to be an explanandum. It does not seem a hypothetical construct introduced to explain something else. However, this point is open to debate. There are accounts which declare that phenomenal experience is a construct (for example (Churchland 1996)). These accounts declare that phenomenal experience does not refer, or that it does not refer to what our basic understanding would suggest. These accounts treat phenomenal experience as a theoretical entity like phlogiston, which is 'eliminated' (declared as being non-existent).

What are thought of as secondary properties are also thought of as qualia, raw feels, phenomenal character, and subjective experience. To Locke, these were considered as what they seem to be to our intuitive understanding. Locke used the term 'secondary properties' (or secondary qualities) to refer to our immediate experiences. Redness is a secondary property, as is painfulness. Locke distinguished these properties from 'primary properties', which would be called 'objective' or 'scientific' properties today[4]. Primary properties are the properties of objects of which we, in loose terms, are not directly aware. Primary properties of objects have a disposition to cause secondary properties. In Locke's view, weight and bulk are primary properties, but heaviness is secondary. Weight, the primary property (weight and mass being essentially synonymous in Locke's time), has a disposition to cause heaviness, the secondary property.

[4] The distinction between primary and secondary qualities was examined by Locke in his *Essay concerning human understanding*, published December 1689. See (Niddith 1975) for a contemporary edition of this work.

According to Locke, we were directly in touch with our experiences; there was no question of doubting them; we have direct knowledge of our redness experience for example. Knowledge that came from the senses, Locke called 'sensitive' knowledge. Secondary properties we have a direct knowledge of, but not primary properties, in Locke's view. Primary properties, being distinct from secondary properties, (but which have a disposition to cause secondary properties), are known indirectly. To Locke, heaviness we have direct knowledge of, but not bulk. (These would translate to weight and mass, respectively, in todays terminology).

The reason for creating a special place for our immediate phenomenal experience (for the secondary properties of Locke) is that there are aspects of our phenomenal experience and they seem intrinsic; they appear as they appear. They do not seem to require definition in terms of something else. Redness appears, and seems to be, redness. It may be part of a colour spectrum, but it is still redness, and that appears as redness, regardless of greenness. To refer to 'redness' seems coherent, and 'redness' seems irreducible. It does not appear that redness could be, or is, anything else.

For the claim of qualia to be meaningful, there must be something to which 'qualia' refers, and there must be some way in which we have epistemic justification for knowledge of qualia. In the past, the justification for qualia was not an issue. Experience was considered immune to doubt, and thus there was little concern for the question of knowledge of direct experience. Currently, there is an issue of the epistemic justification for the claim of qualia. This is because the claim of qualia is no longer considered self-evident. Thus, the claim of qualia must be supported by an epistemic appeal, even if that epistemic appeal states that qualia are just qualia, and our epistemic certainty is without doubt.

The epistemic certainty of qualia is the supposed core-epistemic fact that allows us to claim that qualia are what they seem to be. This is to be distinguished from knowledge about what we experience, as this is not direct knowledge of experiences. One could distinguish between direct knowledge of qualia, and knowledge derivative on, or about, qualia. The 'direct knowledge of qualia' is that which provides a degree of immunity to doubt. There is a need for this distinction in the term 'consciousness'. There is the question of direct experience, and the question of a judgement, or knowledge about experience. This can be termed the distinction between phenomenal consciousness and access consciousness (Block 1990). There are similar distinctions made elsewhere (for example, (Bisiach 1988)).

The question of 'phenomenal consciousness' rather than 'access consciousness' is the question of qualia. It is the issue of 'what it's like' to be an experiencing thing. The distinction can be made in these terms. The issue of 'what it's like' has been raised by Nagel (Nagel 1974), and this has lead to the issue of phenomenal consciousness being stated in terms of 'Nagel consciousness' (Nelkin 1989). The distinction can also be stated in terms of 'first order' knowledge and 'second order' knowledge: first order knowledge is knowledge of experience, and second order knowledge is knowledge about experience.

Having a distinct 'direct' knowledge of qualia, which is distinguished from judgement knowledge has problematic implications. A problem is that such 'direct' knowledge can be considered independent of issues of behaviour and function. This knowledge is usually considered in these terms, because it must be a type of knowledge immune to doubt. If we can doubt our direct knowledge of experiences, then there is a question over the claim of phenomenal experience being what it seems to be.

There are accounts that do not have this 'direct' knowledge of qualia. Qualia, in these views, are equated (if they are equated with anything at all) with the second type of knowledge. Thus, qualia are the result of second order knowledge about, or a judgement about, something else. In such views it is the judgements about something that is what qualia are.

The 'higher order thought' hypothesis is one such view (Rosenthal 1990). In this view, it is a higher order thought about a state that is the condition for consciousness. This is not to say that a higher order thought about a state makes the state a conscious state. Nor is it to say that it is the higher order thought about a state that is a conscious. These interpretations are expressly discounted by Rosenthal. It is the relation between the higher order thought and the state that provides consciousness. Unfortunately, Rosenthal does not say what a 'thought' is, in his view. In the absence of a clear definition, it can be said that it is a particular judgement (higher order thought) about a state that is the condition for consciousness.

There are complications in these distinctions between direct knowledge and judgement knowledge, depending on how they are used. For instance, Rosenthal's higher order thought view allows for sensory qualities of which we are unaware. By sensory quality he means phenomenal experience, but without judgements about this phenomenal experience. Therefore, there is room for cases where there is no higher order thought directed at the state providing sensory quality; and so there is no possibility for judgements about sensory quality in such a case. He makes a distinction between phenomenal experience and 'access' consciousness, or judgement. Thus, we can be unaware of (in that we make no judgement about) phenomenal experiences. Others may find that the notion of a phenomenal experience of which we are unaware to be incoherent, in that what it is to be a phenomenal experience includes awareness of the experience. This is a statement that phenomenal experience is always accompanied by judgements about that experience, or that judgement is a criterion for phenomenal experience.

Our understanding of qualia is that we are aware of qualia. How much coherence is there in an unfelt pain? Views, which allude to such, are difficult and intricate to understand. There may be some coherence to felt pains which are not painful *qua* noxious, however[5]. Nevertheless, where such subtle distinctions are drawn, subtle problems arise.

[5] A friend who suffered terrible migraines as a child would on occasion be brought into hospital when

The distinctions result from three assumptions. (1), there are phenomenal experiences; (2), we have direct knowledge of them; and (3), we can make judgements about our experiences. The problem is that (3) seems to be necessary for us to know we are experiencing. The question can be asked, what is the difference, if there is a difference, between real and ersatz pain? Perhaps judgement is all that is necessary, and elimination is possible. If (3) is not necessary, we still have direct knowledge of experiences, as in (2), yet while not having necessarily having judgement knowledge of (being able to make a judgement about) them. The requirement for (2) arises because a strong phenomenal realist claim must argue that we have direct epistemology of experiences, which is immune to the vagaries of judgement.

The direct knowledge of phenomenal experiences claim, as opposed to judgements we can make about experiences, is necessary if phenomenal realist claims are made. However, the cases in which judgement is absent, and where there can be experiences of which we are unaware (such as is possible in Rosenthal's view), there is a difficulty. This may lead to arguing that judgement knowledge always obtains in the presence of phenomenal experiences. Thus, judging one has experiences reveals one as having experiences, and having experiences allows for judgement of experiences. This is the claim that (3) always obtains in the presence of (2) above.

Ultimately, these situations point back to asking what the issue is in the first place. Our claim of qualia is an epistemic claim about ontology. It is a judgement we make about our experiential situation. The judgement must be justified, and the judgement must reveal ontological phenomenal experiences, in some manner. Again, this leads to the question of what, exactly, the claim of qualia is. Is the claim of phenomenal experience the claim of ontological experiences, or the claim of judgement of experiences, or a claim of judgement of experiences that reveals a direct epistemic link to ontological experiences? Is redness an ontological fact, a judgement of what seems to be an ontological fact, or an epistemologically sound judgement of an ontological fact?

There are arguments that claim that redness is simply not an ontological fact, that claim there is no ontological redness. Yet the proponents of such arguments still see and experience red. Redness experience *qua* ontology is denied, because 'red experience' is not about ontology, it is epistemic, it may be purely a judgement, and in that way, 'red experiences' are not denied as 'red experiences'. They are not denied as 'red experiences', but they are denied as 'red experience stuff (or natural kind, or ontological item)'.

There are arguments that state that all there is to the knowledge of redness is judgement knowledge, without 'direct' epistemic knowledge. Thus, in these views, the experience of redness is the judgement that we are having an experience of redness. Redness

all pain killers failed, to be given an intravenous narcotic. I asked whether this got rid of the pain, and the surprising answer was that it did not. However, she said, "the pain didn't bother me any more". The pain was still there, but in some way divorced from its feature of awfulness.

qua ontology is denied, but red experiences remain. Phenomenal properties, subjective experience and qualia, may or may not refer to ontological kinds in their own right. Our knowledge of qualia may have one or both aspects of direct epistemology and judgement.

1.2.2 A distinction

Immediacy is an aspect of phenomenal experience, whatever phenomenal experience is. The immediacy of phenomenal experience is one reason to claim that it is what it seems to be. It seems that if we are indeed certain of anything, we are certain of our experiences.

If phenomenal properties are eliminated or reduced to something else, there is the implication that we are more certain of ‘something else’ than we are of the phenomenal properties in question. When Dennett says that yellowness is just the judgement of occurrent yellow (Dennett 1981), he means to say that he is surer of the mechanism underlying the judgement of occurrent yellow, than he is of the phenomenal experience of yellowness. That he is surer of the latter means that this certainty can override his immediate thoughts regarding the former. Yellowness appears to be yellowness, in itself, but as Dennett is more certain of something else being this yellowness, he revises his original understanding. Dennett’s *a posteriori* knowledge overrides his intuitive understanding, or so it seems.

If we accept phenomenal (secondary) properties, some form of inference to items of an objective ontology (primary properties) may be required if idealism or solipsism is to be avoided. By means of transcendent inference, hypotheses are formed about what is unobservable, often by postulating the existence of unobservables. The justification of transcendent inference itself is a separate issue.

There are arguments out of solipsism that are not classifiable as using transcendent inference. An example is Moore’s action of looking at his hands, described in his “a proof of the external world” (Moore 1962). This does provide a route out of solipsism based on the assumptions inherent in “here is a hand, and here is another hand”, but it does not directly provide a transcendent inference.

Similarly, there is an anecdote (Boswell 1791) about Johnson kicking a rock, and taking this to be a refutation of Berkeley’s idealism. It is not a refutation. Berkeley’s immaterialism concerns transcendent matter, not the apparent solidity of experienced matter.

The idealist label, however, is controversial. An idealist is variously classified as someone with a solipsistic bias who does not see the need for transcendent inference (to transcend the contents of their minds, or their experiences), or one who wishes to find such an inference. Descartes, Berkeley, Kant, and Schopenhauer are called idealists. Yet, each used a type of transcendent inference.

There are difficulties with the issue of transcendent inference. It presupposes that there is a clean distinction between observation and inference. With the distinction clean, then there is the observable, and the unobservable. Current empirical scientific and psychological knowledge cast doubt on the unquestioning acceptance of such a distinction. The issue of transcendent inference, and its justification becomes problematic if there is no clean distinction between observation and inference. The emphasis can shift to thinking of all as inference, rather than all as observation. This can lead to eliminativism with regard to qualia.

If it is argued that there is no clean distinction between observation and inference, there is less justification to a claim of direct awareness of phenomenal experience. There are arguments to suggest that there is no clean distinction between knowledge about (or judgements of) phenomenal experience and direct knowledge of phenomenal experience, and so there is no 'direct' knowledge of qualia. There is also a difficulty with transcendent inference, as there is no clean 'direct' realm of phenomenal properties to transcend. There may be no inference to noumena, in the sense of a strict transcendent inference from phenomena to noumena. There is an inference to scientific ontologies, to functional and behavioural explanations. However, these may not be considered as transcendent ontologies. As there is no privilege given to phenomenal properties, these inferred ontologies, facts, and explanations are seen merely as the extension of whatever it is that we think of phenomenal properties.

In such a view, facts about atoms are no different in from facts about experienced redness. Thus, claims of 'specialness' for facts about experienced redness are difficult to justify.

1.2.3 Objective and transcendent

Our claims of an 'objective' ontologies or facts, are necessarily dependent upon our own makeup. Our makeup defines our epistemic capacities. Our claims about anything will reflect us as it reveals the world.

How could one conceive of an objective world? A world as it exists objectively, without subjective colouring, is an interesting idea. The truly objective world is not a subjectively described world. The 'objective' world about which we all agree can be said to be built upon intersubjective agreement. If the subjective is taken to include privacy and ontological phenomenal experience, there is a possibility for the concept of a truly objective world, objective ontologies, and objective facts. However, nothing can be said about such an objective world. The items of the objective ontology cannot be described; any comment on this objective world is subjective. In the absence of any ontological status being given to subjective experience, or subjective knowledge, the truly objective concept loses ground.

Kant's term, 'noumenon', is applied to an item of this truly objective world[6]. As an objective item, its inherent nature is unknowable: the 'thing in itself' is unknowable. Everything about it is unknowable, apart from its existence: that we know of noumena (the plural of noumenon) can be argued. A transcendent inference from phenomena can be argued. This type of inference provides knowledge of the existence of noumena. The transcendent inference is the bridge that provides knowledge of the existence of the unknowable, in this instance. Generally, a transcendent inference is an inference to an unobservable. Noumena, however, is a stricter concept than an unobservable. In physics, there are what are classed as unobservables, but specific knowledge about these unobservables is claimed.

Regardless of the status of phenomenal experience, and of our knowledge, there is the fact that what we know is dependent on ourselves. There is a world as it appears, whether or not phenomenal experience is eliminated, reduced, or embraced in a dualist manner. Whether this is justification for a strong concept of 'subjectivity' is another matter.

Regardless of the status of phenomenal experience, be it reduced, eliminated, or an irreducible ontological kind, there is the sense in which we have a particular point of view. We are located in some place, at some time, and we are dependent on our makeup for our sensory and epistemic apparatus. Using the term, 'view from somewhere' allows some of the issues regarding 'phenomenal experience' and 'subjective knowledge' to be raised without using either of those terms. Thus, it does not seem to beg the question regarding the ontological status of phenomenal experience, or a particular school of epistemology implied in subjective knowledge (see (Nagel 1979) and (Nagel 1986)). That we have a view from somewhere does not imply anything about the status of our phenomenal experiences, but it does indicate that there is an indexical issue. Nagel can deal with experience via dealing with our point of view, our subjective point of view.

If the nature of objective views is considered as independent of the meanings of the terms relating to our own point of view, including 'phenomenal experience', it becomes a truism that there are no red things, for example. When a tree falls in the forest with nobody in earshot, it does not make a sound. Look at someone tasting coffee and you will find there are no tastes to be found. It is meaningless in the context of the objective concept that is divorced from the concepts coming under the blanket term 'subjective'. There are no red things because 'red things' is meaningless. There are redness-experiences, but that seems not 'objective', and if made 'objective', seems not 'subjective'.

Introductory student texts on epistemology usually have a section on 'Colour Skepticism'. This is usually the lead into Skepticism regarding the senses, the external world, the self, and so on. The basic statement will be the same: looking for purely objective subjective colours is tricky. The colour of something is dependent on we who look at it. Dogs do not see redness in the world. It is difficult to build a purely objective view: the

[6] See Kant's *Critique of Pure Reason*, first published 1781. There are a number of contemporary translations and e-texts; see (Kemp Smith 1965) for a popular translation.

subjective will always be implicitly included. If the subjective is not included, it will seem that the accounts, which leave it out, have exactly that difficulty: there is something left out.

1.2.4 An explanatory gap

It comes down to what 'redness experience' refers to. It may rigidly designate, or it may not[7]. It may designate ontological experiences in tune with our intuitive understanding, or it may not.

Consider what 'water' refers to. Does it refer to watery-stuff, or $H_2 0$? Consider what 'red experience' refers to. Does it refer to 'redness experiences', or to something else? Something else may be neural firing, or physical states, or functional states, perhaps. Now, water is watery-stuff, and water is $H_2 0$. Not all watery-stuff is $H_2 0$, of course. The intension of water, and of 'red experience', depends on what they are, and what we know. The prior intension[8] of water is watery-stuff. The posterior intension[9] of water is $H_2 0$.

Phenomenal experience terms similarly have two aspects, a prior and posterior intention. What is water? Is water watery-stuff, or $H_2 0$? It is both. The posterior intension rigidifies the prior intension, as not all watery-stuff is $H_2 0$. What is experiential redness? It is 'yellow sensation', or is it whatever 'yellow sensation' is? It can be both. Phenomenal experiences can be raw-feels, and they may be neural firing.

Phenomenal realist claims, however, argue that the posterior intention of 'yellow sensation' will turn out to be the same, or similar, to the prior intension: phenomenal experiences are what we intuitively think they are. This would be an argument for ontological experiences. Perhaps 'red experiences' are 'redness experiences', and that is the whole story. Or perhaps 'redness experiences' are something else. On the other hand, perhaps 'redness experiences' are eliminated, for other reasons: perhaps 'redness experiences' are argued to be merely the judgement of 'redness experiences'.

Is experiential yellowness 'yellow experience', or is it whatever 'yellow experience' is? And, can we conceptually understand that whatever 'yellow sensation' is, that it is experiential yellowness, as we know that $H_2 0$ is a watery-stuff which is water?

Consider that the question of phenomenal properties, of qualia, is an 'X' to be explained. There are many ways of dealing with 'X'. One approach, the phenomenal realist approach, is to state that there is an 'X', and it is what it seems to be. A way to justify

[7] "The silliest philosopher in the world" is not a rigid designator as it picks out different people in different logically possible worlds. "The smallest prime number" is a rigid designator, as to our understanding it is not a logical possibility that this number can be anything other than 2. This rests on modal reasoning, which rests on the concept of logical possibility, which rests on logic. A logically possible scenario is one which is not logically contradictory.

[8] This is what Kaplan termed the 'character' of a term (Kaplan 1978)

[9] Kaplan calls this the 'content' of a term (Kaplan 1978)

this statement is to argue that ‘Y’, an explanation or view that encompasses or could encompass everything, leaves something out, and what it leaves out is ‘X’.

The reductionist approach argues that there is a ‘Y’, quite different from ‘X’. It argues that the explanatory domain of ‘Y’ is such that it can be shown that ‘Y’ encompasses all that ‘X’ could be, and ‘X’ is reduced to ‘Y’. ‘X’ is not denied, but ‘Y’ is seen to encompass what there is, and so ‘X’ can be described in terms within ‘Y’. The specifics of the reduction are a separate issue. ‘X’ is considered something, but something that can be reduced. Molecular motion is seen to encompass all that heat could be, so heat was reduced to molecular motion, but there is still heat. Constellations are arrangements of stars, but there are constellations: they are just arrangements of stars.

An eliminativist would argue, similarly to the reductionist, that there is a ‘Y’, and that it encompasses enough to question the status of ‘X’. However ‘X’ is not considered reducible to ‘Y’, nor is ‘X’ considered something in its own right. Thus, ‘X’ is just denied; it is considered something that has no referent. Phlogiston was considered a substance released by heating, causing heated items to lose weight. But molecular facts were seen to encompass and explain this resultant weight loss. Thus, phlogiston was not reduced to these other facts, it was denied; there is no phlogiston, it simply does not exist.

Generally, there is a relation between ‘X’ and ‘Y’. In the eliminativist case there is no ‘X’ to be related to ‘Y’. The apparent immediacy of qualia (the ‘X’) is the same regardless of what opinion one has. Reducing pain to neural firing, or eliminating it, does not help the crying child or the child’s parents.

Relating phenomenal experience to something else, be it neural firing, microphysics, or behavioural dispositions, is a real problem. Accepting the relation when we arrive at one is a more difficult problem. Eliminativists still feel pain the same way that dualists do.

This difficulty is called, among other things, the explanatory gap (Levine 1983). The explanatory gap is the apparent gap there is between our immediate experiences and our explanatory constructs. The gap may be metaphysical: there may be an actual existing difference between phenomenal experience and other more objective ontologies. If it is not metaphysical, it is still epistemological. It may be that there are no ontological qualia, it may be that eliminativism holds, but it does not seem that way; experiences will always seem to us as experiences themselves.

For this reason, Levine argues that an epistemic gap will always be present in explanations of phenomenal experience, regardless of whether there is a metaphysical gap. The prior intension of the terms of phenomenal experience will never be related to the posterior intentions of the terms of phenomenal experience in a way that seems acceptable.

An explanatory gap for water would be that we all consider water as watery-stuff, but cannot convince ourselves that it is H_20: our understanding that water is H_2O may not remove our lingering feeling that there is still a gap.

The epistemic gap concerns what we can know. There are arguments that claim non trivial and insurmountable limits to our epistemic capabilities, and that these limitations entail that there will always be an explanatory gap, but that there is no metaphysical gap (McGinn 1989).

In the claim that there is no metaphysical gap, but there shall always be an epistemic gap, there is a difficulty. If there is no way to bridge the epistemic gap, how is the claim that there is no metaphysical gap justified? It is difficult to argue for no metaphysical gap, yet for an epistemic one, because it can seem that, given an epistemic gap, there is no justification for claiming there is no metaphysical one.

The explanatory gap is the statement of a gulf between some form of objective explanation, 'Y', and phenomenal experience, 'X'. 'Y' could be physics, or behavioural dispositions, or some form of explanation that does not refer to items of phenomenal experience directly. Schrödinger contrasted "the two general facts (a) that all scientific knowledge is based on sense perceptions, and (b) that nonetheless the scientific views of natural processes formed in this way lack all sensual qualities and therefore cannot account for the latter" (Schrödinger (1958, 103)). Summing up the embarrassment of the explanatory gap, Eddington noted the attitude of the materialist who "regards consciousness as something which unfortunately has to be admitted but which it is scarcely polite to mention" (Eddington 1928, 384).

A way to argue for an epistemic explanatory gap is to argue that there is no *a priori*[10] entailment from 'Y' to 'X' (phenomenal experience). An *a priori* entailment would make it seem to us that there is no gap. In the absence of this, the link will seem unclear. However, there may be an entailment *a posteriori*, and this would rule against a metaphysical gap. It is not obvious that water is anything but watery stuff. Nevertheless, *a posteriori*, water is H_2O. It requires a little work for us to understand this. In this example water and H_2O are related by identity, whereas a physical or objective explanation 'Y' and phenomenal experience 'X' may be related in some other way. The stronger claim of a metaphysical explanatory gap would need to argue (or if not argue, declare) that there is no *a posteriori* entailment from 'Y' to 'X'.

To argue against an explanatory gap requires arguing for an *a priori* or an *a posteriori* connection between 'Y' and 'X'. However, it is particularly difficult to argue a case for an *a priori* connection. There are arguments against an explanatory gap, which do not provide an *a priori* connection between 'Y' and 'X', though they do argue for an *a posteriori* connection. An instance of such an argument is one which states that first person phenomena can be translated into third person terms (Hardcastle 1993). This is akin

[10] I use *a priori* in the manner of Kant. The two essential aspects of *a priori* knowledge, according to Kant, are necessity and universality. Necessity, as experience does not show us that things could not have been otherwise; and universality, because experience confers only a judgement of comparative universality through induction.

to making the claim that, *a posteriori*, first person phenomena are merely third person phenomena. Nevertheless, it does not seem that way to us. These views argue against a metaphysical gap. However, they remain with an epistemic gap.

Whereas it may be argued that redness is, or is related to, something else, it still does not seem that way to us. Thus, the argument has not addressed an epistemological explanatory gap. The explanation may relate redness and something else metaphysically, but not epistemologically. It is true that an epistemic gap does not necessarily provide reason for believing the gap to be metaphysical. However, neither does arguing that since there is no metaphysical gap, there is no epistemological one. There is a difference between bridging the epistemic gap and convincing oneself that there is not one because a particular explanation suggests a possible *a posteriori* entailment.

Another way of describing the issue of the explanatory gap comes from Chalmers. He introduced the now popular classifications of 'hard' and 'easy' problems (Chalmers 1996a). Depending on the context, the hard problem is used to indicate the problem of phenomenal experience or the problem of bridging the epistemic explanatory gap between something else (a form of objective explanation) and phenomenal experience. The easy problems are seen to be the problems of something else. Something else, which does not appear to be related to experience, includes the problems of explaining behaviour, function, and so on. The behaviour of an experiencing thing is considered a relatively easy problem, but relating this to experience, or explaining experience itself, is hard.

The hard/easy distinction is dualistic, as is the concept of an explanatory gap. The degree to which this is so depends on the view. It assumes, to varying degrees, that issues of reportability, of function, of behaviour, are separate, and can be dealt with separately, from the issue of phenomenal experience. This is not a minor assumption. Perhaps attention, or thinking, or speaking, or behaving have phenomenal experiential aspects (Lowe 1995). Perhaps redness is not cleanly separable from the report of redness, or the behaviour that redness can induce, or our feelings that redness is a warm colour, and the mood changes it can induce.

1.2.5 First person authority

Descartes, in his method of doubt, argued that we could be mistaken about the physical existence of our own bodies, whilst we could not be in error about what is in our minds. Thus, there is something special about our knowledge of our own minds. This concept is not limited to the Cartesian dualism of Descartes, nor is it limited to the view that this self-knowledge is certain. It does not need a 'Cartesian theatre'. The issue is one of the asymmetry between our knowledge of our own minds, and our knowledge of the external world. This latter knowledge includes knowledge of science and of the minds of others.

The issue is one of first-person authority: others, the third persons, have an indirect route to our minds, whereas the first person does not. The first person counts as an authority on their own minds because, it is supposed, they can know about it transparently, or directly, and this is different from the type of knowledge others can have.

'Authority' does not entail certainty. We need not be infallibly correct about our own mental states. It does not necessarily entail incorrigibility, so we may be wrong, but we cannot be corrected by anyone else. Nevertheless, 'authority' does require some form of asymmetry, some uniqueness about our knowledge of our own minds.

1.2.6 Modal concerns

Phenomenal properties appear immediate, their qualities intrinsic. This is apparent in Nagel's pondering of what it is like to be a bat (Nagel 1974). Dennett, who is an eliminativist, admits this point also. Saying that 'what it's like to be a bat' is to act bat-like (Dennett 1991) does not eliminate what it is like to be a bat; it merely equates it with behaviour. The 'what it's likeness' remains, as does an epistemological explanatory gap.

In asking the question, "what are phenomenal properties?" we are presuming that 'to be/experience an X' is different from 'what is X'. Yet, are these issues distinct? What are experiences? Are they not what it is to be (or have) experiences? The question, "what are phenomenal experiences" is more accurately thought of as a question regarding the explanatory gap. It is the question of how phenomenal experiences are related to something else. Something else is the scientific or objective explanatory view that is used.

Where the existence of phenomenal properties is assumed and a distinction is drawn between experience, and knowing what experience is, a difficultly arises. This is the distinction between phenomenal experience, in our intuitive understanding, and any *a posteriori* knowledge we have. A possible situation arises in which we seem not to know what phenomenal properties are, objectively, or scientifically, or from an explanatory viewpoint. The child which falls does not know what pain is. It has, is, or experiences, pain, but it does not know what it is, in the context of explanation. The child may not know what pain is, but it experiences it.

Nagel asks whether one would know about 'what it's like' from everything we knew about 'something else', to which 'what it's likeness' is related; could we know the 'what it's like' aspect of experiences from *a posteriori* knowledge alone? The 'something else' is our explanatory apparatus, our *a posteriori* knowledge. This may be function for the functionalist, behaviour for the behaviourist, or neural firing for the computational neuroscientist. If redness is not immediately apparent (apparent in an *a priori* manner) as an entailment from, an implication of, or an aspect of this 'something else', a gap exists.

The distinction is one between experiences themselves which we experience, and the knowledge of 'something else' that they are, or to which they are related. One can know one, yet not know the other. The child that falls knows pain, but does not know pain, as it relates to (or is eliminated in favour of) something else. There is a more detailed example akin to, but the opposite of, the child's dilemma. It concerns an imaginary colour-blind neuroscientist called Mary.

Jackson pondered on a neuroscientist named Mary who lived in black and white isolation (Jackson 1982). Mary is conceived as knowing everything about colour as it relates to 'something else'. 'Something else' in this case is considered as scientific knowledge. She thus knows that there are three types of cone in the retina and that there are different wavelengths of light, for instance. She knows that these facts combine to produce an intricate space of 'colour'. She knows the functional and behavioural aspects. She knows that red is a 'warm' colour for instance. She knows everything (in the context of 'something else') there is to know about the perception of colour, but she has not experienced colour. In a similar but opposite way to which the crying child does not know about pain, Mary in her monochrome environment knows about colour. She knows everything there is to know about redness (*a posteriori*) as much as the child knows nothing there is to know about pain.

If released from her drab environment, she would experience something she never experienced before. The question is, does she gain additional information, or learn an additional fact? That there is a difference between her black and white existence and her colour existence is not in dispute. Perhaps through some trick of neural stimulation or some deep meditation, she induced a red experience while still in her drab environment, but this does not alter the question the argument raises.

Jackson argued that Mary 'learns'; he argued that experiencing red for the first time counts as an additional fact for Mary. That is where the issue lies. Mary experiences colour for a first time, true. If considered an additional fact, then it is over and above the 'something else' that she knew; thus, it is not encompassed by the *a posteriori* knowledge she has. If this is so, then 'something else' is not a complete view, as it does not encompass all the facts; it leaves out facts about experiential colour. If it is argued that Mary does not 'learn', the fact that she has a new experience does not necessarily count against the fact that 'something else' encompasses all there is to know.

It can be considered a modal concern. If Mary learns, then there are additional facts about colour experience over and above those provided by 'something else'. Thus, the world as known in 'something else' terms is not complete. The facts that Mary learns are facts that place additional constraints on how the world is. This places additional restrictions on the space of possible worlds that have persons that experience redness. It eliminates possible worlds, as 'something else' under specifies with respect to the world: *a posteriori* knowledge under specifies, because it leaves out facts of experience.

If her experience of redness does not count as a fact, however, there are no modal concerns. Yet she experiences a change, so this additional experience of redness, over and above her knowledge of 'something else' as it relates to perception, must still be dealt with.

To counter the ‘extra facts’ view, colour experience, though it is a new experience for her, must not be seen as a fact, or as a separate additional piece of knowledge. It must be contained within ‘something else’. It is possible to argue a distinction between ways of knowing the facts about ‘something else’. She may have known all the facts about ‘something else’ in an academic sense, but not in some other way. Perhaps she was not ‘acquainted’ with the facts about redness, though she knew them all. Perhaps she knew all the facts about redness in some sense other than an ontological sense. Maybe she just experienced old facts in a new way. Maybe she just gained some ability. The arguments hinge on claiming that facts about experience are encompassed by ‘something else’. The argument is that *a posteriori* knowledge is in principle ‘complete’, in that it fully specifies the way the world is.

1.3 Relation

Whatever the status of phenomenal experience, the relation it has to physics, neuroscience, or something else, needs to be shown. Moreover, if there is no relation, this too needs to be argued. The relation of phenomenal experience to something else will allow the justified ascription of phenomenal experience to an instance of something else. The relation between phenomenal experience and something else may not provide a bridge over the explanatory gap. There may still be an epistemic gap in the presence of such a relation, especially if the relation does not show a conceptual link or an *a priori* link between phenomenal experience and something else.

The ways in which phenomenal experience (or other mental items) can be related to physics (or some other explanatory account) generally fall into four categories. It could be denied as something in its own right: this is eliminativism. In such a case, it is not reduced, it is bluntly denied. It can be reduced by showing that an account of phenomenal experience can be recovered from an account of some lower level such as physics. It could be that phenomenal experience (or other mental state) is identical to a physical state. Finally, not knowing the specifics of the relation between phenomenal experience (and other mental items) and physics, one can specify the dependencies between them.

1.3.1 Elimination

Our experiences are not postulated to explain something else, and so they shall not evaporate with a flash of insight. Nevertheless, their status as specific things in their own right can be denied. The term ‘qualia’ may refer to something, but perhaps it does not refer to something in particular, not to something specific and special, or it may just be a theorectical construct.

Eliminativism is the view that does not consider our experiences to be specific ontological kinds in their own right. Eliminativism does not force a difference in how our experiences seem to us. Pains and redness are both still there. Eliminativism states that pains are pains *qua* unpleasant painfulness. Eliminativism is not anaesthesia. There are still 'qualia' to be explained. However, they are not explained by positing specific ontological kinds.

It would seem, because of its name, that eliminativism is the elimination of something. That presupposes there is something which eliminativism eliminates. There are pains; nobody would disagree. Eliminativism does not eliminate these. Eliminativism does not eliminate anything. It merely states that there is nothing to eliminate, and there never was anything to eliminate. Eliminativism is a view that has a disadvantage, in that it is difficult, in any meaningful sense, to deny what is not there. Eliminativism says that a non-existent thing does not exist. There is nothing that eliminativism denies; it declares that theories, which do not include certain elements (such as irreducible 'qualia') are true. Thus, eliminativism does not leave out 'experiences'. Eliminative accounts still have all these wonderful phenomenal experiences intact.

When Dennett states that yellowness is just the report of yellow, he is not denying yellowness (Dennett 1981). But to someone who believes that yellowness is more than the judgement of yellowness, someone who has the view that reportability is not phenomenology, Dennett is denying yellowness (Cam 1985). Eliminativists do not feel they eliminate anything at all, but non-elminativists feel that they do.

Some eliminativist views suggest that our concepts of phenomenal experience would undergo some change in the face of sufficient knowledge and empirical evidence. It is suggested that we could eventually understand that phenomenal experience is not what we thought it was. Authors of such views give the impression of having attained this state already (see (Churchland 1983), and (Hardcastle 1996)). These arguments bear some resemblance to the 'category mistake' described by Ryle (Ryle 1949). Ryle's example tourist did not know that there was nothing to Oxford University but the colleges. Similarly, it is suggested, we do not know that there is nothing to our brains but neural firing.

There is a difficulty, however. It can seem that some eliminativist arguments sound like reductionist arguments. If it is said that there are no qualia, just reports of qualia, there is an ambiguity. It could be that qualia are eliminated, in that there are no qualia. Alternatively, it could be meant that qualia are nothing but reports of qualia, in which case qualia have been reduced, not eliminated. There is no fixed fact as to the essential quality of phenomenal experience of yellowness in qualia eliminativist views, as 'yellowness' is not thought of as separate of the knowledge, belief and claim of yellowness. The way eliminativism is sometimes phrased allows confusion between eliminativism and reductionism: is yellowness eliminated in favour of phenomenal judgement, or reduced to phenomenal judgement?

In some eliminativist views, there is no fixed fact about whether or not something is 'experiencing'. It is indeterminate; it is like asking whether a joke was funny, or wondering if that film was a little better than I thought it was. In Dennett's view, phenomenal experience is like this (Dennett 1988). In his view, experience is the judgement of the experience, and nothing more. That is to say, qualia experience is eliminated, and there is only judgement, rather than saying qualia experience are reduced to qualia judgment.

Elimination usually takes place like this: an account 'Y' is used as a general explanatory tool, and is viewed in such an expansive way to suggest that there is nothing further, and hence no 'X'; thus 'X' is eliminated. In an expansive acceptance of 'Y', which is independent of any concepts of 'X', elimination is the only response. This is especially so in the artificial intelligence view of computationalism. This is akin to functionalism, but much more complete, in that it considers only computations and functions, without further constraint. Such a view can only lead to eliminativism with regard to phenomenal experience. Such a view can be summed up, as, "Y explains everything, and this does not include X, therefore, there is no X". To see how expansive and unconstrained the computationalist view can be, see see (Rey 1986) and (Tienson 1987),

1.3.2 Reduction

Reduction works this way: an account 'Y' is viewed as an explanatory tool. It may be decided a reduction of an 'X' to 'Y' be attempted. It is, ultimately, an appeal to authority. It is an appeal to the authority of the 'Y' which is to be shown to provide a deeper understanding of what 'X' is.

There may be two different ways of describing the same thing. Both descriptions have the same referent. There is no asymmetry in these descriptions: both are considered equal. However, if one description encompasses the other, or if one description is more complete than the other is, then it cannot be said that both descriptions are equal.

If one set of facts or one description can be replaced with a more fine-grained description asymmetry arises. This is especially so when one set of fine grained facts is seen to provide a more complete description of an object. Another set of facts, if they are not as complete, can be replaced by the more complete set of facts. A vague description can be replaced with a more complete description that keeps the essential features of the vague description. The vague description can be derived from the complete description; reduction can be seen as re-description.

These 'privileged' descriptions provide the baseline 'most complete' description or explanation from which other sets of facts can be derived, and to which other sets of facts are reducible. However, the facts that are reducible to other facts do not have the same status as the 'privileged' facts. There are many opinions as to how this point of 'different to privileged facts' can be expressed.

Anything can be described in many different ways. There is no sense in which a description of a chair as comfortable is more ‘privileged’ than a description of it as old. However, when the chair is given what is considered a complete microphysical description, the situation is different. Other descriptions of the chair are seen as reducible to the microphysical description. The microphysical facts are considered ‘fundamental’; the other facts are seen as merely derived, higher level, reducible, or simply as facts which ‘come for free’. This is not elmininativism as the reducible facts are not denied. However, they are not given the same status as the facts to which they are reduced.

Eliminativism says that a phenomenal realist theory is off target, and that items within that theory need to be jettisoned. Reductionism, however, can say that the realist theory is ok, but that it can be shown how certain supposed irreducible items of that theory can be derived from within another theory, or within that theory.

Reduction can seem eliminativist (which it is not) if it reduces facts that have ontological aspects in their common meanings. Given the ontological commitments that are closely tied to the ‘fundamental’ or ‘baseline’ facts, there is little room for ontologies tied to reducible facts. Thus, mental facts being reduced to physical facts can suggest that any mental ontology is eliminated, as physical facts are all there is, and only ontologies tied to physical facts are accepted. There may be an implied assumption that ontological commitments may only be tied to ‘privileged’ facts, the facts to which other facts are reducible. Quine and Rorty opted for eliminativism rather than reduction in this regard, as Churchand has more recently (see for instance (Quine 1966), (Rorty 1965) and (Churchland 1981a)).

1.3.3 Supervenience

The most common form of relating apparently separate sets of facts is through supervenience. This relation can also be used where a more specific relation between sets of facts is known. It can be used to relate two sets of facts in an understandable way, although one set is taken to be reducible to the other. However, that it can be used to relate sets of facts in the absence of a more clear relation is its strength.

Davidson applied the supervenience concept in philosophy of mind (Davidson 1970). The supervenience concept was developed by Kim (Kim 1978) (Kim 1984a). Since then, it has been in frequent use.

There are different varieties of supervenience, each with differing constraints on the relation. A supervenience relation essentially states, “There is a relation between these things and the relation is one of dependency”. The differing forms of supervenience depend on the degree of dependency.

Supervenience relates two sets of facts by dependency: by how one set of facts ‘fixes’ the other set. Thus, supervenience can be seen as a statement of the conditions required for a set of supervenient facts to obtain. Supervenient facts are fixed by the facts upon which they supervene.

This is an example of the use of supervenience: if one particular level, or description, or theory is taken to be complete, or causally closed (or otherwise all encompassing), one can argue that all must supervene upon it. If microphysics is taken to be a complete view, then everything must supervene upon microphysical description. This statement relates everything in the world as supervening on the microphysical.

If one is sure that all supervenes on the physical, or one is sure that physical theory or description is complete, then *everything* must supervene upon it. Not just things and laws, but everything: classes, numbers, universals to name a few (Armstrong argues what may be called a 'total supervenience thesis' in this context (Armstrong 1982)).

If all supervenes on the physical, then the supervenience relation may merely be a placeholder in our ignorance, to be replaced in due course. The relation can still be used, however, but a more accurate relation would supersede it. Kirk suggests, along these lines, that everything that supervenes on the physical should be a strict implication from the physical (Kirk 1996). Pettit's physicalism is a view along these lines (Pettit 1993).

Usually, supervenience relations relate two sets of facts at a particular instant: it is assumed that there is no temporal context to the relation. 'A'-facts supervene on 'B'-facts implies that 'A' facts obtain when the 'B' facts obtain only. The supervenience base need not refer to an instant, or a specific moment. For instance, the facts of what makes me the person I am today may be argued to supervene on the totality of the facts of my past. Or it may be argued to supervene purely on my present state, independent of the past specifically (in so much as the past influenced my present state, and thus does not need to be referred to directly). There is a well-known thought experiment along such lines.

A freak lightning strike hits a log in a swamp, and suddenly, a replica of myself (called swampman) is created (Lycan 1987). Therefore, it should have the same beliefs as I do. However, there are those that argue not. Swampman has not been to London, but I have. Though we are identical, Swampman's belief that he has been to London is in error, while my similar belief is not. There is a difference in our beliefs. If my beliefs supervene on my physical state as it is now, then swampman and I share the same beliefs. If swampman does not share my beliefs, then my beliefs do not supervene merely on my present state. The supervenience base for beliefs extends into the past. There are arguments (Armstrong 1982) for adding a temporal dimension to the supervenience base, of which Lycan's swampman argument is one. There are two meanings to 'belief' here. There is the correctness of the belief, which can be evaluated in terms of my and Swampman's past. In addition, there is the experience of belief. One can argue that one aspect of belief supervenes on the present state of the person, while the other does not.

Supervenience is the statement that there is a relation, without necessarily the knowledge of the specifics of the relation. Thus, it does not entail any particular fact about the specifics of the relation. Supervenience does not entail reduction, or elimination, though it can be used in an argument for such relations.

Certain facts may be said to supervene upon other facts. However, this relation need not be necessary. It may be that certain facts could fail supervene on other facts. The facts upon which other facts supervene could obtain in the absence of those facts.

If a set of facts, which supervene on other facts, do not obtain in the presence of those other facts, modal concerns arise. If there is a contingency in the supervenience relation, then the contingent supervenient facts are modal world fixers: they further restrict the space of possible worlds.

If a set of facts, which supervene on other facts, necessarily obtain in the presence of those facts, there are no modal concerns. The necessarily supervenient facts are not further world fixers. The facts upon which they supervene are modal world fixers, but the supervenient facts, as they supervene necessarily, do not further restrict the space of possible worlds.

A note on symmetry, or lack thereof

Contingency of supervenience relates to cases where certain supervenient facts do not obtain: where certain facts, which supervene on other facts, do not obtain in the presence of these other facts.

If the relation were necessary, then both sets of facts obtain together, or not at all. The supervenience relation is somewhat symmetrical. The statement of the supervenience relation, however, is explicitly asymmetric. If A necessarily supervenes on B, then in the absence of the supervenient facts, A, there is no B. One does not obtain without the other. Yet, the supervenience relation is asymmetric in phrasing: one set facts supervene on the other, not the other way around.

This asymmetry must be based on something. It is usually based on the 'privilege' of a certain set of facts, or the sense that one set of facts is more 'fundamental'. Alternatively, it can simply be that one set of facts is more expansive and more descriptive.

Consider that the mental necessarily supervenes on the physical. If this is so, then mental facts do not obtain in the absence of the required physical facts. However, it holds in the other direction also; the specific physical facts do not obtain in the absence of the specific mental facts, as that would be an instance of physical facts obtaining in the absence of mental facts. In this scenario, the relation which states, "particular mental facts supervene on particular physical facts", is asymmetrical. Nevertheless, this could be stated as, "particular mental facts and particular physical facts are necessarily co-occurrent", and this sounds symmetrical. This depends, however, on how 'mental state' is considered. It may be that a single mental state may supervene on a number of physical states; the relation is then asymmetrical: the mental state cannot change without change in physical state, but that mental state may supervene on a number of physical states.

In this case, physical facts are seen to be more ‘privileged’, as they are more expansive and more ‘fundamental’. This being so, the supervenience relation, with its inherent ‘directionality’, is appropriate. In the case of a contingent supervenience relation, the contingency defines and requires asymmetry. If supervenience is necessary, then the dependency is bi-directional and co-dependent, though the supervenience relation is asymmetrical.

1.3.4 Identity

Another manner in which the phenomenal and the physical (or qualia and function, or one set of facts and another set) is related is by identity. Yet, stating that the phenomenal is identical to the neurological or the physical does not remove an epistemic explanatory gap. It still seems to us that they are different, in the way that, *a priori*, it is not clear that water is H_2O.

That water is H_2O only became evident after some work. The way it is phrased now is that water is *identical* to H_2O. When such *a posteriori* statements of identity were first used, it was believed that they were contingent. Water is H_2O, true, but it may not have been, as there is a possible world in which it is XYZ, it was claimed. The logical non-contradictory nature of ‘water $= XYZ$’ is a non-trivial question. Now, however, we are somewhat sure of the necessary nature of the ‘water $= H_2O$’ relation, but the justification for this statement depends on the theory of reference implied[11].

Statements of identity between the mental and the physical, however, do not have an agreed status as being a necessary relation. Thus, the identity statements can be read as contingent. This non-necessity can be read as stating, for example, that Pain is C-fibre firing, but it may not have been.

The contingent nature of the identity relation is not that a thing is contingently identical to itself. Things are necessarily self-identical. It is the contingency of the *statement* of identity. The contingent identity relation is always a relation between classes, or types of things, rather than singular things. ‘Pain’, the class, or type, is identical to the class of ‘C-fibre firing’; this says nothing of the relation between particular pains and particular neural firings.

An advantage of identity theories is that the notion of identity avoids the notions of epiphenomenalism and emergentism. If the phenomenal is identical with the physical, neither epiphenomenalism nor emergentism is appropriate. Smart, an early proponent of an identity theory argued this point in favour of the identity theory (Smart 1959).

[11] Putnam argued the contingency of water=H_20 to support an externalist view of meaning, in his famous ‘twin earth’ argument (Putnam 1975a). The argument presented stated that what is in our heads does not determine the reference to our thoughts: what is in our heads does not fix the reference to the posterior intension of water, which is H_20.

There is another aspect to the identity relation that is not an aspect of the supervenience relation. If the phenomenal is identical to the physical, then there is spatial and temporal coincidence. The physical state is in the brain, and the phenomenal state of experiencing redness is therefore coincident with the physical state, as they are identical. The coherence of this statement depends on the coherence of arguing that our red experiences are literally and strictly 'in the head'. Identity, therefore, tends towards a narrow view of phenomenal experience.

There may be a sense in which two things related by identity seem contingent. There may be an epistemic separation between two particular things. However, this separation cannot be metaphysical, if identity is so. Kripke argued that the co-occurrence of mental and physical states is contingent, and this contingency cannot be explained away. But as a thing cannot be contingency identical to itself, he concluded against the identity theory (Kripke 1972). However, Kripke's argument depends on whether an identity relation between particular things, or types of things, is considered.

An identity relation ought to state what exactly is identical to what. Kripke's argument requires that a 'mental state', such as 'pain', and a 'physical state', such as 'neural firing', are rigid designators. His argument against identity theory does not hold if these do not rigidly designate. Mental states may be identical to physical states, but this does not say what a particular mental state and physical state designates.

A mental state may not rigidly designate; it may not designate a particular thing. It may designate a type or a class of things. In using a particular mental state term, we import other related terms. Is pain painfulness, or is the latter a property of the former, and is it an essential property of the former? By using 'pain' as a mental state term, are we importing physical state properties? Using 'pain' as a rigid designator may not be so simple to justify.

Rigid designator identity is the identity of particular things. It is the identity of tokens. This was Kripke's argument: if there are specific things, and they are identical, this is not a contingent matter. The more relaxed identity, type identity, relates types of things[12]. If singular terms are rigid designators, all identity statements with singular terms flanking the '=' are necessary, Kripke would say.

Where types of things are related by identity, specific tokens of these types may not be related by identity necessarily. One could describe differing views of identity, depending on how the mental and physical terms designate. As to what a mental state term designates, there are many opinions, and so there are many ways of keeping identity theory in the face

[12] In a brown fruit bowl are two green apples. There is one type of apple, but two apples. One type, two tokens. The bowl itself is a token of the types: brown, on the table and so on. The green of the apples is a type with tokens: the apples, grass, and so on. Historically, the type token distinction came from dealing with language. For instance, considering the answer to the question, "How many letters in the following greeting: 'Hello'." There are five tokens and four types.

of Kripke's comments. It may be essential that a particular (token) pain is C-fibre firing, but not to 'pain' as a type. 'Pain', the type, is not a rigid designator, though particular (token) pains may be. If it is not essential that 'pain', the type, is a particular physical state, then accounts of 'pain' the type, need not be physicalist in the sense of referring to a physical state directly.

'Pain', the type, may not designate particular neural structures by structure or location; it may pick out neural stuff by its causal role. Moreover, many things, many neural structures, could have filled that causal role. Thus, the mental could be related, by identity, to the causal roles that many different physical states could fulfil, rather than physical states directly. Lewis argues that causal roles are definitive of mental states, and that particular mental states are identical to particular physical states because physical states fill these causal roles (Lewis 1966). He considers this a type-type identity theory: it relates types of mental state to types of physical state, via a relation to causal roles that are filled by physical states. Lewis considers the causal roles as states which fill the causal roles of folk psychological roles (Lewis 1980).

Armstrong identifies mental states as states which are apt to bring about behaviour, and builds his view on this (Armstrong 1968). This is somewhat similar to Lewis. Yet, when causal role is invoked, there is controversy over whether this is type-type or token-token. This hinges on whether it is the role occupied, or the occupier of the role that is referred to in the identity relation. If is the role occupied, and not the occupier of that role that is important in causal role, then type identity need not apply.

Horgan considers functionalism, and comes out in favour of calling it a token identity view (Horgan 1984). In functionalism, functional roles are definitive of mental states, and this says nothing of physicalistic constraints. But specific mental states happen to be physical states, so token identity applies. It need not be the case that there is a relation between types of mental state and types of physical state. Horgan argues that the statement of functionalism does not imply type identity, but token identity is clearly the case.

Jackson, on the other hand, argues that functionalism is compatible with type identity (Jackson, Pargetter, and Prior 1982). He considers mental states as designating a state type that fills a functional role. Thus, he argues that the statement of functionalism does imply type identity, via the fact that mental states are related to functional states, which are related to a class of physical states.

Davidson argues for token without type identity (Davidson 1970). He argues that there are no strict psychophysical laws that relate the mental and the physical. The absence of strict laws, in his view, implies there is nothing upon which to argue for type identity, as Jackson does with functionalism. In Davidson's view, there simply is not a relation between types of mental state and types of physical state. For this, his view earned the rather nice name 'Anomalous Monism': anomalous, for there are no strict relating laws, and monist, since the mental is not a separate ontological category in his view.

As one can ask what need there is of supervenience, one can ask what need is there to invoke identity. The committed physicalist takes it that everything ought to be entailed by the physical. In a similar manner in which he argued against the need for a supervenience concept, Kirk argues that strict entailment from the physical to everything else removes the need for a separate identity relation (Kirk 1979).

1.4 Functionalism

Behaviourism states that mental states are behaviours or dispositions to behave. Analytic behaviourism states that the meaning of mental state terms is given by specifying the relevant behaviours or dispositions to behave. The causal theory of mind states that mental states are typical causes of behaviour and dispositions to behave. The causal theory lead to analytic functionalism.

Functionalism states that what matters for the mind are functional roles; they are what matters for having a mind, and what matters for being in one or other mental state. The differences in various types of functionalism are differences in the classification and description of functional roles. Essentially, there are two main categories of functionalism. The first category is common-sense, or analytic, functionalism, which states that it is common knowledge which functional roles matter for the mind: the 'folk' functional roles. This form of functionalism considers the common sense functional roles to give the meaning to the mental state terms. To be in the mental state M is to be in that state which fills the common sense functional role associated with M. The functional roles give the meaning, but do not fix the reference, of the mental state terms.

This is to be contrasted with the second category of functionalist accounts, the empirical functionalisms. These accounts may fix reference on the nature of states that play these roles: neural structures, for instance. This could rule out robots with minds, as there may be constraints on what can support this functional role. This is to say, the nature of that which plays a functional role may matter, and thus multi-realisability may not be so. Thus, there is a difference depending on whether it is the role, alone, or whether the occupier of that role is also deemed important.

Generally, the folk functional roles fix reference on further functional roles, which empirical science uncovers; it is an empirical *a posteriori* matter as to which roles are crucial for possession of mental states. Functionalism of this sort may not speak of folk functional roles at all.

The commonality in functionalist accounts is that it is something somewhat abstract about our internal nature that is essential to having mental states. The variations in functionalist accounts stem from differences regarding how abstract these states are, and who describes their natures. The states may be abstract enough to allow many physically different things to have minds, and thus neurons are not essential. Neither are neuroscientists: that degree of empirical knowledge is not necessary, in such a view, to describe the functional roles. Alternatively, the states may be more specific, restricted to neural structures, perhaps, in which case the neuroscientist can describe the underlying functional roles.

1.4.1 Functionalism and physicalist identity theses.

A physicalistic idea of mind tells how mental phenomena are constituted; it tells us what they physically are. Identity theories are physicalist theories. Functionalist theories tell of what mental phenomena do. What they do can be described in different ways.

Functionalist accounts are different from behaviourist accounts, though both are concerned with action, with what is done. Behaviourist accounts attempt to define mental phenomena in terms of behaviour. Functionalist accounts can be seen to describe mental phenomena in terms of what they do, but this is in a functional, rather than behavioural, way: mental states are not defined in behavioural terms in functionalism. Functionalism can allow for mental phenomena to be considered independent of behaviour, though this is a subtle issue. A point can be made, however, that if functionalism describes mental states being the 'cause' of behaviour; those mental states cannot be defined in terms of behaviour.

Functionalism has fewer restrictions compared to grounded physical notions, and so does not have the difficulties of physicalist accounts. Physicalist accounts may require 'pain' to be a specific type of neural event. Thus, for creatures to experience pain, they must have this type of neural structure. Functionalism can be more liberal, in that many things may realise mental states; this is termed multiple realisability.

Functionalism can be construed in many ways. Each way allows for a varying degree of multiple realisability. It is not chauvinist, and can allow for creatures very different from us having mental states.

Functionalists may consider the mental to supervene on the physical, but this is not required. Functionalism can allow for ghosts made out of ether-stuff, so long as it can support the relevant functional organisation. Functionalism is compatible with dualism, however, as functionalism does not require specific commitments, *per se* on the physical nature of mind. If functionalism relates the mental to the physical, it has something in common with identity theories. Functionalists would agree that any particular mental state is a particular physical state, and this is token identity. Functionalism without any specific physicalist constraints, is what I term 'radical functionalism'. Radical functionalism has the least restrictions on multiple realisability.

Functionalism does not entail reduction, as there can be functionalist dualists. However, there are non-reductionist monists also. As they are monist, they accept a token identity of the mental to the physical, and accept a degree of multiple realisability. However, they do not claim that the mental is reducible to the physical.

A non-reductionist physicalist monist must deny type identity. Type identity relates types of mental state to types of physical state. There are thus relations, psychophysical laws, or bridge laws which relate the mental and the physical. Reduction is at least plausible in this view. Thus, for non-reductionist functionalism, type identity must be denied.

If type identity is denied, then particular mental states are, or are related to, particular physical states, but there are no strict laws relating types of mental state to types of physical state. Thus, all that is acknowledged is a relation between a particular mental state and something physical. There is nothing upon which to base a reductionist account. Eschewing reductionism in this instance could be denying reduction epistemically: denying that we can find a reduction. Alternatively, it could be or denying it metaphysically: denying that there is a reduction.

It is this sense of separateness of 'mental property' in non-reductionist functionalists accounts, which causes some semantic confusion. A monist non-reductionist token-identity functionalism does not see the mental as a different thing from the physical, but does reject reduction. Davidson holds such a view (see (Davidson 1970) and (Davidson 1980)). The 'monist' label indicates Davidson rejects dualism. Yet, in this view, mental properties are not reducible to physical properties. Is it then property dualism? One could talk of one stuff having two aspects, rather than admit talk of properties, but this raises deep and complex ontological issues. Monism is the view that states there is one type of stuff. Yet, anomalous monism claims that there is one stuff, physical stuff, while insisting that the mental is anomalous (Davidson's elegant way of saying 'non-reducible') in respect of that physical stuff.

Type identity functionalism allows that the mental is to some extent reducible to the physical. Functionalism, then, is compatible with reductionism (type identity), non-reductionism (at least no type identity), dualism (neither type nor token identity), and anomalous monism (token without type identity). However, functionalism with type identity is essentially a physicalist view.

As 'anomalous monism', being a physicalism of sorts, can seem dualistic, so too can functionalism be argued to show physicalism, in one sense, to be false. Though compatible with differing opinions as to the identity relation, there is room for arguing that functionalism is for or against physicalism. This hinges on two factors. One, what is a mental state, and two, what gives a mental state its identity: what it is that is common among pain states that make them pain states.

A functionalism which accepts token identity without type identity claims that mental states *are* physical states. However, it does not provide a physical account of the identity of a mental state. What makes a type of mental state that type of mental state *is not* physical, it is something else: it is its functional role or its causal role.

There are ways of constraining a token without type identity functionalism. Martian pain and Robot pain may not be physically the same, and the type 'pain' may not be identifiable via the physical. Perhaps a human-specific type identity physicalism is possible. In general, type identity may not hold, but it may hold in a species specific way, and so a species specific reductionism may hold. Kim argues against non-reductivist physicalism in this way c̃itekim:nonredcausation. The argument is essentially that within a certain domain, specific reductions are possible, and hence type-identity is possible. However, the non-type identity views (for instance, anomalous monism), argue against type-identity in principle, and thus domain-specific reductions may not count against non-reductionism.

Type identity provides an account of the identity, the common element of types of mental states, in terms of the physical. Token identity without type identity answers the ontological question in physical terms, but not the 'metaphysical' question: it does not say what the identity criteria are for a mental state in physical terms. I have used the term 'metaphysical' loosely, for comparison with ontology. Type identity does answer this 'metaphysical' question in physical terms.

Functionalism that accepts type identity between mental states and physical states must allow reduction as a possibility. This type of functionalism is compatible with physicalism. Consider that all particular pains are physical. My pain, Robot pain, and Martian pain are all genuine pains. They are physical states. Nevertheless, what pains have in common in virtue of which they are pains need not be something physical. This is what functionalism essentially says, and this allows for multiple realisability. Martians are made out of Martian-stuff, robots out of silicon, and human brains from neurons, but all can have pains.

1.4.2 Functionalism and computationalism

The 'function' in functionalism can be described in different ways. There is a model of 'function' in computation theory. Abstract models of computation allow 'function' to be described in those terms. The functional details can be described in the manner of computation. Computation can be realised; some abstract models of computation, such as the Turing Machine, can be imagined in a physical way.

However, there are functionalist accounts that do not explicitly tie 'function' to computation theory, although all functionalist accounts rest on computation theory. There is a functionalist view that considers that the common sense view of mind is a functionalist account. It does not further tie this to computation systems, such as the Turing Machine. The functionalism of Lewis is an example (Lewis 1966).

Other functionalist accounts are more directly tied to computation systems, to a lesser or greater degree. Putnam's original functionalism (see (Putnam 1960)) tied mental items to states of a Turing machine, where the particular Turing machine was considered the appropriate functional characterisation of the mind. Later, he argued more explicitly that functional organisation, not physical makeup, mattered (Putnam 1967).

Other functionalist views, though ultimately a computational description, range between the common sense description and explicit computational description. In addition, there are functionalist accounts which are content to consider functionalism to be functional *description, or characterisation*, of the mind, while other functionalist views take this description of mind to be the sole defining characteristic of mind. The former views mental states to be describable as computational states, while the latter considers us computers made of flesh and bone. This latter view, if allowing for a very liberal multiple realisability, is radical functionalism.

1.4.3 Functionalism and Content

Functionalism can state the identity criteria for mental states in terms of functional role. It may allow that mental states are physical states. Physicalism would answer the identity criteria question in physical terms. But is this enough to explain mental states?

Functional states are in the head, yet meaning and content are in the world. At least, our accounts of meaning and content are in the world, and functional roles are considered in the head, though they can be construed widely. If Twin Earth is taken to succeed at what it was intended to demonstrate, accounts of content must be externalist. There is no commonly accepted account of narrow content. Functionalism is internalist and narrow, but some functionalists would argue against a narrow account of content (Jackson and Pettit 1988), while others would argue that it must be possible (Fodor 1990).

Yet, even the narrow/wide distinction is disputed for particular functionalist accounts. Marr's theory of vision is seen to be a wide, rather than narrow account by Burge (Burge 1986), while Segal disagrees (Segal 1989).

A narrow functionalist account of mental states does not solve the problem of a functionalist account of the meaning and content of those mental states. If ordinary propositional attitude contents do not supervene on the totality of the state of a persons head, then a functionalist account of mental states leaves rather a lot out, as functional roles are ostensibly narrow.

Consider a narrow functionalist account that deals with meaning. It must bring meaning into the head, and it does so by considering the meaning of a term as a functional state. Thus, the meaning of a word and its function are related, where this relation may be regarded as identity. In cognitive science, there is such an account called procedural semantics. In philosophy of mind, one such account is called conceptual role semantics.

One can consider 'function' widely also, specifying functional roles with respect to one's environment; this can even be extended into the temporal dimension by considering a historical context. The other way is to use something else other than functionalism in an account of content.

Functionalism can tell us what makes a mental state *a desire* for a weekend in Paris by specification of some functional role; mental states are defined or identified with functional roles. However, it fails to tell us what makes a mental state a desire *for a weekend in Paris*; that may require an externalist account of content. Functionalism can tell us what makes a mental state the type of state it is, but no more. In other words, functionalist accounts may have an account of desire, but they may not provide an account of content, especially if it seems an externalist account of content is necessary, as functional roles are considered narrowly. Thus, functionalism deals with the *desire* for a weekend in Paris, but without dealing with the content issues raised by the term ‘Paris’.

Allowing a functionalist or a type-identity physicalist view of mental states, while leaving out content, results in the problem of type identical functional states with different content. The externalist thought experiments point out this situation.

The content difficulty is the problem of representation in general. What, exactly, is it that makes something in one’s head representative of Paris? There is a serious issue with the coherence of claiming that something, considered alone, can be determined to represent something else. Fodor calls this issue, the issue of providing an account of what makes what is in our heads mean something, or be about something, which is not in our heads, ‘psychosemantics’. It is essentially the issue of intentionality.

Representation

Mental content is a problem related to that of phenomenal experience, if mental content has experiential aspects. The degree to which these issues are distinct is the degree to which meaningful mental states are distinct from the experience of meaningful mental states. Mental content can be considered as semantics, and can be considered independently of issues of phenomenal experience.

‘Thinking about Paris’ is a mental state. It is also an experiential state. An identity theorist for qualia would state that the experiential aspects of this state are identical to something within the head. An account of the meaning of that mental state, however, may need to refer to something outside the head.

Paris is not in the head (it would not fit). ‘Thinking about Paris’ is an intentional state: it is about Paris. An account of content bearing mental states may take into consideration things outside the persons head in giving an account of meaning. This is a ‘wide’ account of content, as opposed to a ‘narrow’ account, which deals only with what is in the head. A narrow account of content bearing mental states has the difficulty of giving meaning to a mental state about ‘Paris’ without being able to de-reference the term ‘Paris’. The qualitative experience of meaning for the person who is thinking about Paris, however, is considered a separate issue. It is unlikely that something mysterious reaches from our heads to Paris when we think of Paris. It is unlikely that a magical intentional lasso grasps the object of intention.

As mental content and qualitative experience are treated as separate issues, there need be no conflict between a wide account of mental content, and a narrow account of experience, though one can experience content bearing mental states.

Could an identity theorist look into the head and conclude, “Aha! A thought about Paris”? The ‘thought about Paris’ can be considered in an experiential or a semantic manner. An identity theorist with respect to qualitative experience would presumably allow that it is in principle possible to look into the head and conclude, “Aha, an experiential state of thoughts of Paris”. Yet, the experiential “thought about Paris” is not necessarily the semantic “thought about Paris”: the issues can be, and are, dealt with separately.

Perhaps a more accurate way to say, “this person is having a thought about Paris” is to say, “this person is having a thought about a city which they mistakenly believe is Paris”. However, the person looking at the other persons head plays a role here. They are providing the semantic meaning to the other persons mental states. For that reason, it is problematical to describe experiential states in semantic terms. Once we say, “the qualitative experience of having a thought about Paris” both issues are addressed.

There is the qualitative experience issue and the issue of providing accounts of mental content. Moreover, there is a third point to be considered, if content mental bearing mental states are considered that concern a belief.

Someone may mistakenly believe they have arthritis in the thigh, though this is impossible, as it is a disease of the joints (Burge 1991). However, their belief need not necessarily be inconsistent, though it is false. Another example (Kripke 1979): Pierre goes to London and finds it ugly; so he forms a belief that London is ugly. His friends go to a place called “Londres” (London), and say it is pretty, so Pierre forms a belief that Londres is pretty. Pierre’s beliefs are not inconsistent, as he believes that ‘London’ and ‘Londres’ refer to different cities.

The examples here deal with intentional content. However, intentional content, though it is ‘about’, can not be considered in the context of what the intentional terms seem to refer to or denote.

“The king of France is bald” is false. However, enumerating all the bald and not bald things reveals the king of France to be in neither list. If we take “the king of France is bald” to be a proposition, and expect that proposition to denote, there is a problem. It is a problem for the excluded middle: the king of France is either bald, or not bald.

George IV wished to know whether Scott was the author of Waverley. Scott was the author of Waverley. Substitute ‘Scott’ for ‘the author of Waverley’, as they are the same. But George IV did *not* wish to know whether Scott was Scott: a problem with identity.

These examples were used by Russell to indicate that propositions, generally, cannot be evaluated if it is required that they be taken to refer, or denote, things that exist (Russell 1905). What a statement is about may not be anything at all, and yet the statement can be evaluated. Russell's account still indirectly referred to, or asked questions of that which exists, but without the assumption that all must denote existing things. There is nothing that is not bald and the king of France. Neither is there a bald thing that is the king of France. So, is the king of France bald, or not bald? Alternatively, if we dispense with the law of the excluded middle, perhaps the king of France wears a wig[13]. Russell's answer to this dilemma was to ask, "is there an entity that is now the King of France and is bald?" That question is easily answered, yet 'king of France' does not denote anything (any existing thing); it is not about anything, though the mental state 'the king of France' is an intentional mental state.

Before Russell, evaluating such propositions as "the king of France is bald" resulted in ontological proliferation, with 'square circles' and 'the even prime other than 2'. Attempts to fix the difficulties of such a view by having a class of non-entities did not work. Before Russell, there were arguments that "the round square that is round" was a true proposition. Russell stated, "there is one and only one entity which is round and square, and that entity is round", and concluded that this was false.

The semantics of mental content can be provided with a wide content approach. Intentional content has 'aboutness', but it may be about nothing at all, and despite this, it can be meaningful. 'Aboutness' is not simply what is denoted, or what is referred to. In addition, an account of the qualitative experience of content bearing mental states need not require those items that a semantic account of content needs.

Intentionality is 'aboutness', but 'aboutness' need not be about anything in an existential sense. Intentional mental states are *represented* as sentences such as 'The king of France is bald'. That sentence is about an intentional state. That particular sentence fails a test of existential inference. The other example concerning Scott and Waverley fails tests of substitutionality. Those sentences are intensional with respect to these tests for existentionality. However, the truth conditions for these sentences do not require that the world be as represented by the original intentional states. Those sentences represent the content of those intentional states, and such content can be reported independently of the existence of objects referred to by the representation.

'The king of France is bald' is a representation of an intentional state that is representational. The truth of that sentence is dependent not on how things are represented by the intentional representation. It is dependent on how they are in the mental world of that intentional representation. Thus, 'the King of France is bald' is not subject to the laws of co-reference or substitutability. That sentence does not refer to what is represented in the intentional state; it expresses the content of that state.

[13] This was Russell's jibe at Hagelians, "who love synthesis"

This is all very fine, but it does not explain intentionality. It is akin to representation, and so long as representations of intentional representational content are not taken to be representative of what is represented, all is well. ‘There are angels’ does not represent angels, but represents intentional representational content.

There are arguments for and against ‘intrinsic’ intentionality. Searle is a vocal proponent of intrinsic intentionality. Derived intentionality nobody disagrees with, as it is self-evident: I do not understand Chinese, but the symbols do represent, for many people. Derived intentionality has a name: ‘meaning’, and we have to learn to give meaning to symbols. As to what, exactly, ‘intrinsic’ intentionality is, there is much confusion. Is there a magical lasso that reaches from my head to Paris when I think about Paris? If there is not, then what is intrinsic intentionality? Unfortunately, Searle does not give an account of what intrinsic intentionality is, as it is a basic self-evident premise of his view.

The difficulty with intrinsic intentionality is the difficulty that is evident in providing accounts of mental content. Content includes intentional content, and intentional content is representational. For the empiricist, it is the difficulty of determining what it is in neural firing that ‘represents’ something, as neural firing on its own can be said to have merely ‘derived’ intentionality: we assign it meaning. But that neural firing is in someone, contributing to supposed ‘intrinsic’ intentionality.

In considering intensional statements (which are usually about intentional content), we are taking those statements to represent something (the intentional content). This is itself a form of intentionality. However, the inkblots arranged as the text, ‘There are angels’ could represent anything. We bring representation to it; not so, it is sometimes supposed, for our own intentional states. Symbols in the world have ascribed, or derived, intentionality, while the mental realm has ‘intrinsic’ intentionality, Searle would claim.

Syntax and Semantics

‘Book’ has meaning for me. It is representational. However, it is not representational for someone who does not understand English. ‘Angels are beautiful’ is an intentional mental state, it is representational, but it does not necessarily represent an existing thing. The meaning of ‘Angels are beautiful’ is not dependent, specifically, on existing Angels. The meaning of ‘Paris is beautiful’ is somewhat dependent on (or at least an account of this mental state would include) the existing thing ‘Paris’. Both these states are intentional.

It is a problem of representation, and there is a related statement of the same problem. It is this: syntax is meaningless, until we assign it semantics. So there is a distinction. Then anything can be considered a symbol, and so a distinction between anything and its ‘meaning’ can be made. This includes one’s neural firing. What is it about one’s brain state that determines one’s mental content as the intentional state about Angels (that do not necessarily exist)?

Similarly so for functionalist accounts of mind: functional accounts are representational; the symbols used are of themselves, meaningless. A written functional account of mental states in terms of probabilistic finite state automata has no intrinsic meaning (being just arrangements of ink on a page), however, this is used as a functional account of meaningful intentional mental states.

What is it in a symbol manipulating functional device, such as a radical functionalist would claim us to be that gives these symbols meaning? Take this argument further, add that we have intrinsic intentionality, and conclude that we are not just symbol manipulation devices, as Searle does.

However, this all hinges on one point. Representations of intentional representational content do not have intrinsic intentionality, even if it is supposed that the intentional states they express do have this special intentionality. Accounts of mind are always going to have derived intentionality. The syntax/semantics distinction, if used against functionalism, makes the error of considering this avoidable when it is not.

What do syntax/semantics arguments actually say? If they say that our neural structures in functional states can be seen to represent anything, then it says very little. Pick any symbol and give it any meaning. Looking inside the head of any person, or any robot, or into the ethereal stuff of any Angel, is an act of assigning meaning.

In any case, if there is 'intrinsic' intentionality such intentionality cannot be 'found' by looking inside heads; the claim of intrinsic intentionality is going to be justified in terms of some first person epistemic argument. This is what Searle declares; he claims it is self-evident, from the first person point of view, that mental states have intrinsic intentionality. Thus, the argument that accounts of mind do not provide an account for it does not conclude for intrinsic intentionality. If it exists, it is not an empirical issue or an issue of third person accounts, if it does not exist, it is not an issue anyway.

Within the syntax/semantics debate is the debate on the ascription of functions or computations to objects. In a similar manner to differentiating the syntax and semantics of a term, the function or computation performed by an object is differentiated from the object. Some view the notion of 'function' and 'computation' to be purely ascriptive. Searle is one vocal proponent of such a view (Searle 1992). In this view, objects do not perform computations *per se.* There have been attempts to bolster this opinion by arguing that any object is open to an infinite number of (or too many) functional accounts. As for the infinite case, these arguments have been rejected. If two apples fall from a tree one after the other, followed by two which fall together, it is not appropriate to claim that the tree performed computational addition of $1 + 1$. That attribution required a lot of interpretation work on our part.

There is a distinction that can be made between assigning a computation to an object and considering that object as implementing that computation. In the former case, there are no explicit restrictions. In the latter, however, there may be restrictions, and different types of restrictions can be required. Chalmers gives an account of what it is for an object to implement a computation: the 'causal structure' of the object should mirror the 'formal structure' of the computation (Chalmers 1994). This would rule out Putnam's rock as implementing most of the infinite computations that Putnam originally assigned to it, so too with Searle's wall (see (Searle 1992)) running a word processor program (Chalmers 1996b).

Determining the computation an object implements, or choosing the computation to assign to an object, is dependent on what we know about that object. There is potential for slicing the object into a multitude of levels: a coarse-grained view of its external behaviour, or a fine-grained view of its internal workings.

The definition of computation allows this. This slicing is a feature of abstraction; it is this slicing upon which the notion of universal computation is built: universal computers can perform (potentially) any computation.

Putnam's argument, and Searle's rest on a similar fact: what aspects of the rock are being considered? Looking down at an extremely fine grain reveals a considerable amount of internal activity: each atom's interaction with the environment is a potential input/output. At another level, looked at in another way, an object, which performs the function we designed it to perform, can have many different functions. Computers compute, as we intended them to. They also function as heaters and headache inducers, with their flickering screens and whirring fans. The fans function to cool the computer, by actually generating more heat in the computer, but expelling hot air a lot faster. Refrigerators function as heaters as well as coolers.

Determining the computation an object implements is not easy, even if it is taken that there is more to an object implementing a function than a purely external assignment of function.

A related concern is the context in which functions occur. An action may be considered part of a function only as far as it contributes to some further end. A teleological account of function is unnecessary in a computational account of function, but is used in philosophy of mind, where it is seen as a possible restriction on the assignment of functional roles. Teleological characterisation of functional roles has two problems. Firstly, there is no acceptable account of teleology. The second problem relates to Swampman: teleology requires the right sort of history or environment. A chance creation of Swampman would not comply with these: do Swampman's mental states have content?

1.4.4 Liberalism and Chauvinism

A thing may seem to function as a thinker, but we may not believe it is a thinker. Just as we may view a chess machine as playing good chess, but not being a chess player, we may think that the fake thinker is not a thinker. Thinking, it is supposed, is anything but mundane, and so the functions and computations which occur (which are assigned to, or which are implemented) in a thinker must not, therefore, be mundane. Block conceived of 'Blockhead' (Block 1981), a thing which seems to think, but, because of its nature (it is a fancy lookup table) he argued that it does not think. Blockhead could pass the Turing test, if it acted appropriately; the test is not concerned with the structure of the candidate.

Searle's Chinese Room is an argument that points to internal constraints also. It is an argument that pumps the intuition that a certain type of internal construction, a symbol manipulation device, could not posses semantics (Searle 1980). Searle's argument concerns a person who does not understand Chinese, performing symbol manipulation tasks to translate and reply to Chinese statements. Searle points out that the person within the Chinese room need not understand Chinese. However, this was assumed in advance, so the intuition pushed is that if the person does not understand Chinese, then the room and its elements (paper, pens, rule books the person follows) does not understand Chinese. However, Chinese speakers understand Chinese, so there is a difference. Searle left the way open for others to argue that the complete system, room and person, could 'understand' Chinese, while accepting Searle's intuition that neither person nor room considered separately, understood Chinese (Cole 1991). Searle was trying to assign the important part of what the system did ('understand Chinese') to a single part. I understand English, but a bunch of neurons in my head, considered separately, do not (see (Searle 1990) for further comments from Searle on his Chinese room).

The Chinese room and Blockhead arguments have different specific points to make. Searle was arguing against functionalism while pushing a clean syntax/semantics distinction, while Block was demonstrating that input/output functionalism is false. The important issue, that of internal constraints, is raised in both arguments. What is it about an object that provides sufficient condition for us to claim, at a functionalist level, that it implements a computation, and at a higher level, that it fills some folk-psychological functional role?

Ultimately, functionalism equates thinking (and other various items in our mental medley) with functional role, or causal role. In addition, if a type-identity view is held, these functions/functional roles/computations are equated to types of physical state, and what is understood by 'physical state' will depend on empirical knowledge. If type identity is rejected, there may be a constraint on what can realise these functions: mental states, accounted for functionally, are token identical to physical states (Davidson 1980).

The constraints depend on the functionalist view; and will fall somewhere between liberalism and chauvinism with respect to that which supports the necessary functional roles or functional organisation.

1.4.5 Concerning 'loose' qualia

Phenomenal experience, or 'qualia', appears conceptually distinct from function and behaviour. This conceptual distinction may or may not indicate that this distinction is logical or empirical. However, if the distinction is acknowledged, and is taken as suggesting that qualia and function are logically distinct, then certain cases of logical possibility arise.

That qualia are conceptually distinct from function follows from intuitions that function under-specifies qualia. This allows logically possible cases in which qualia are absent or inverted.

The first case is the zombie scenario: a person without phenomenal experience, but functionally equivalent to those that have phenomenal experience. The second case is the inverted spectrum[14]. Inverted spectrum arguments are used to argue against functionalism and physicalism. For reasons of absent qualia, Block and Fodor reject functionalism (Block and Fodor 1972). As a criticism of functionalism, the zombie scenario raises the possibility that realizations of any given functional account of mental states may lack qualia.

The decision to accept physicalist functionalism over functionalism based on absent qualia possibilities is not, however, a straightforward way of solving the difficultly. Adding a physicalist constraint to functionalism does not necessarily avoid the missing and mixed up qualia situations, however. If qualia are conceptually distinct from function, then adding a physicalist constraint to functionalism does not help, because qualia are still conceptually distinct from function. A physicalist identity constraint, which fixes qualia by physical state, still leaves this distinction. It may be argued that, since there is physicalistic constraint, qualia are present. However, these qualia may be mixed up. The physical state may fix the qualia, but there may be inverted qualia with respect to function. Perhaps absent qualia are avoided, but there is still the inverted qualia possibility.

A solution could identify qualia with functional state. Thus, the distinction between qualia and functional state does not apply, and inverted or absent qualia are avoided. This is a functionalist constraint. Physical-state constraints do not fix qualia in a way that avoids the inverted qualia problem. For this reason, there are arguments that suggest, if phenomenal realism is held, then the only options are functionalism, or transcendental dualism (White 1996).

If we can 'introspect', in any sense of that term, our own 'qualia', then there is an argument to be made that they are functional. Absent qualia are impossible, if that type of introspection is possible. Shoemaker pointed out this case, summing it up as a conditional: if qualia are introspectable, they are functional, and absent qualia are impossible (Shoemaker 1975). Churchland has mentioned that apprehending qualia is the direct introspection of 'brain state', but pointed out that this is a problem for functionalism, as there is no way to explain what apprehending a 'brain state' is, in a functional

[14] First mentioned by Locke (Niddith 1975)

account (Churchland 1989b). Churchland is more of a functionalist than a physicalist (see (Churchland and Churchland 1981): this view is not easy to categorise).

This would be akin to a computer program having access to the actual physical state of the computer directly (as opposed to indirectly via data coming from sensors, for instance), and this is impossible. Functionalism alone does not allow for the introspection of physical state, so apprehending a non-functionalist state cannot be given an account within functionalism. Churchland thus tends to physicalist functionalism. Shoemaker replied to Churchland pointing out that the sense of 'introspection' can only reveal functional state, and so the equating (or reduction) of qualia should be to functional state, not physical state (Shoemaker 1984).

The possibility of absent qualia is not, however, reason to reject the possibility that qualia are 'introspectable' in some way. 'Introspectable' is not meant to infer any functional or causal account of introspection. More generally, absent qualia possibility does not provide justification for rejecting any sense of 'knowledge' or 'epistemology' of qualia. Many phenomenal realist views accept the logical possibility of such cases, yet also accept the epistemic justification for claiming phenomenal realism.

Both physicalism and functionalism have difficulties providing sufficient criteria for 'fixing' qualia. This difficultly results from distinctions that are made. This is not to say that the distinctions are at fault. However, distinctions between qualia and function, between qualia and behaviour, do lead to absent and inverted qualia arguments.

1.5 Inessentialism

1.5.1 Explanatory irrelevance

If a complete functional, behavioural, or objective world view explains everything, then there is either nothing it does not explain, or the supposition that it explains everything is false. If a functional or behavioural view is considered sufficiently complete, phenomenal properties could be reduced to functional or behavioural facts. Alternatively, they could be declared non existent (eliminated), or assigned to a separate ontological category.

Inessentialism is class of accounts in which an explanatory endeavour is seen to explain everything in certain terms, while leaving something else out. An example is a view that explains all behaviour in functional or causal terms, yet, phenomenal experience, (in the ontological sense), is accepted. As all behaviour can be explained without reference to phenomenal experience, phenomenal experience is irrelevant to the explanation of behaviour. It is explanatorily irrelevant in explaining behaviour, and thus inessential in that regard. Phenomenal experience is a difference that makes no difference, in behavioural terms.

These views are classified as inessentialist. There are also classifications of inessentialist epiphenomenalism and inessentialist emergentism. The commonality is the explanatory irrelevance of phenomenal experience to the explanation of behaviour and causation. The difference is in how phenomenal properties are considered. The distinction between inessentialist dualism and epiphenomenalism is not clean: phenomenal experience in inessentialist dualism can be considered an epiphenomenon. Emergentist views provide an account of how phenomenal experience arises from (emerges from) something else[15].

If mental (psychological, experiential) explanation is reducible to something else, to what degree are such explanations useful? In the context of inessentialism, such explanations are not necessary in principle. Such explanations cannot therefore be autonomous. Yet there are complex and subtle (and difficult to understand) arguments that push for some autonomy of psychological explanation within an inessentialist view. An example would be the argument that the relation between psychology (an aspect of the mental) and neuroscience is more complex than reduction (Fodor 1968). Fodor attempts to provide an account in which the relation between psychology and neuroscience is one of mutual constraint. 'Mutually constrain' cannot mean that they exert anything like a specific influence over each other, as this would entail causally efficacious emergent phenomena. A table does not 'mutually constrain' four legs and a flat board into being a table. If the mental can be reduced to the physical, then it cannot be the case that mental explanation is autonomous from physical explanation (Churchland 1982).

Similar difficulties in attempting to reconcile mental explanation and inessentialism are evident in the work of Searle. He states that consciousness is a feature of the brain in the way that liquidity is a feature of water. He also states that the brain causes consciousness. Liquidity is a feature of water, but water does not 'cause' liquidity. To say this is similar to saying those four wooden legs and a flat board 'cause' a table. He calls consciousness, solidity, transparency, and liquidity 'causally emergent system features'. Therefore, he has different meanings for the terms 'causal' 'emergent' and 'feature' than others.

1.5.2 The status of the 'irrelevant'

If mental explanation refers to the mental (experience, qualia, and so on) causing in virtue of itself, then inessentialism does not hold. If inessentialism holds, the mental 'causes' only in virtue of its being, or being related to, something else. The status of the mental is often enhanced by describing it in such a manner; it makes inessentialism more palatable.

Chalmers describes phenomenal experience as being causally efficacious in virtue of the causal capacities of that which they supervene upon. His is an inessentialist view: to causation and behaviour, phenomenal experience is irrelevant. Nevertheless, phenomenal

[15] An analysis of the term 'emergence', how it is used in the areas of dynamical systems theory, physics, and how philosophy attempts to use the term, is provided in (Silberstein and McGeever 1999). In this paper, we conclude that emergence is but a convenient description in all cases apart from one particular case in physics.

experience is related to that which can cause. Saying that phenomenal experience has a causal role in virtue of its relation to something else does not challenge the inessential and explanatorily irrelevant status of phenomenal experience.

In inessentialist views, the mental is, in principle, explanatorily irrelevant. Often, this extreme declaration of irrelevance is softened; and this is done by declaring the mental *qua* 'something else' is causally efficacious, or that the mental 'in virtue of its relation to something else' is causally efficacious.

The case where 'functional states' are given causal status is similar. A 'functional state' can be an abstract conception. However, there may be a physical realisation of some function or computation. Thus, it could be said that a particular object has a causally efficacious functional state in virtue of the fact that the object implements that functional state, and the object has causal capacities. The physical object does the causing. The 'functional state' cannot be considered to 'cause' in any way aside from the fact that a physical device which is seen to realise, implement, (or have that functional state attributed to it) can cause.

Returning to inessentialist mental phenomena 'causing' in virtue of their relation to something else: either the mental, of itself (*qua* itself), causes, or it does not. If it does not, inessentialism holds. If it does not, one can say the mental 'causes' *qua* something else, but this merely says that something else does the causing.

Whether it is said that the mental does no causing, or does 'qua' causing makes no difference. In both cases, inessentialism holds, and the mental is explanatorily irrelevant to causation.

Mental causation, of itself rather than 'qua' causation, is not compatible with the causal closure of physics, or the 'something else' to which the mental is related (Baker 1993). It is incompatible with any explanation that accounts for all causation or behaviour in terms that are independent of the mental.

Kim calls this mental 'qua' causation 'epiphenomenal supervenient causation' Kim (1984b). In order to claim that the mental has 'qua' causation status, in an inessentialist view, requires that the mental be related to the physical (or something else: function or behaviour) (Kim 1979).

Mental causation, of itself, in addition to the causal closure of the physical would entail downward causation reminiscent of 1930's emergentism (Kim 1992a), where emergent higher level properties influence the behaviour of a lower level. Non reductive physicalism can have this difficulty: non-reductive wholes emerge from parts and then constrain those parts (Kim 1992b).

It can be argued, based on interpretations of physical theory and empirical results, that there are non-reductive wholes that emerge from parts and have causal efficacy in their own right. But these wholes do not violate causal closure of the lower level in possible situations. This is so because these wholes do not constrain their parts in a downward causal manner. These empirically validated cases do allow for a coherent non-reductive view of emergent causation of a sort (Silberstein and McGeever 1999). However, the physical situations in which this occurs are not similar to the physical situations that occur in brains.

Conceivability, contingency, inessentialism and irrelevance

Contingency, and therefore modal concerns, can result from inessentialism. If something does not appear to do anything, then it is easy to imagine a case in which it is absent. This is the case with phenomenal experience: it is explanatorily irrelevant, and it is easy to conceive of its absence.

An epistemic explanatory gap between phenomenal experience and the physical seems to allow for the logical possibility that phenomenal experience may not obtain. Thus, there are arguments for the logical possibility of persons who do not have phenomenal experience.

The logical possibility that phenomenal experience does not obtain, entails that the facts of phenomenal experience are further world fixers. Contingency means additional modal constraints.

In the inessentialist case, the absence of phenomenal experience is *conceivable.* In philosophy, what is conceivable is generally taken to be logically possible. It is conceivable that there are non-conscious persons, so it is logically possible. It is logically possible as long as there is no contradiction arising from this notion.

It is also conceivable that things could travel faster than light. Travelling faster than light is just the conceiving of something going *very fast.* It is, however, *not* logically possible that this is so in the context of current physical theory. In the context of theory, it leads to contradictions, to the travelling object having ‘negative mass’, and of requiring infinite energy to reach the speed of light before it even gets a chance to have ‘negative mass’. In the context of some other physical theory, it may be logically possible. However, independent of specific physical theory, there seems no logical contradiction in something going faster than light.

There is a distinction between logically possible and metaphysically possible. Physicists would argue that it is not metaphysically possible for an object to travel faster than light. That is their opinion, based on what they know currently. A philosopher, considering it a matter of logic divorced from physical theory, may agree, while accepting that travelling faster than light is logically possible.

This distinction between metaphysical possibility and logical possibility, and conceivability arguments is apparent in inessentialist literature. This is especially so in the question of zombies. Zombies are the non-conscious persons mentioned previously. Inessentialism states that phenomenal experience is inessential and explanatorily irrelevant to behaviour. Thus, it is conceivable that there are persons, exactly as conscious persons in every respect except one: they are not conscious.

Zombies have been around for a long time. Descartes considered the possibility of zombies, but concluded against them. For Descartes, language use was limited to conscious persons, and so zombies were ruled out. Early users of the zombie notion include Campbell and Robinson, but perhaps the most ardent supporter is Searle, followed by Chalmers (see (Campbell 1970), (Robinson 1976), (Chalmers 1996a), and (Searle 1992)). What are we conceiving of when we conceive of a zombie? Are we conceiving of a person exactly like us, who is not conscious? How can we conceive of a person with no experiential life? It is akin to imagining what it would be like to be something for which 'what its like' is meaningless. We cannot imagine being a zombie, this being akin to imagining what it is like to be non-existent. Some people would not think that zombies are conceivable. Conceivability is complex, and so the issue is usually phrased in terms of logical possibility.

The logically possibility of zombies as stated by Chalmers is not a statement of metaphysical possibility (Chalmers 1996a). It is a statement that phenomenal experience is an additional modal constraint, over and above those provided by third person empiricism, on the space of possible worlds. However, it says nothing of this world specifically.

For Chalmers, the supervenience of phenomenal experience on the physical is a logically contingent matter. Nevertheless, some further statement needs to be made, concerning *this* world. Moreover, in this world, at least one person, Chalmers, is not a zombie. Thus, though the supervenience relation is contingent and not necessary, it happened also to be the case.

Chalmers had two options. He could have allowed zombies in this world, as well as conscious persons. This would have lead to a considerable other-minds problem for him. He did not choose this, and so, in this world, zombies are only a logical possibility. In this world, phenomenal experience always supervenes where it can; the contingency of the relation is not 'exercised'. Most inessentialists would not want zombies and conscious folk sharing a possible world, least of all our own world. Chalmers allows consciousness to fail to supervene, but only in worlds where it always fails to supervene.

In stating the supervenience relation in such a case, further constraints are necessary. It is not so that consciousness necessarily supervenes, and stating that it is logically possible that it fails to supervene, does not account for this zombie-less world. Chalmers needs some relation, such as one which states that consciousness 'necessarily supervenes in this world, but this necessity is not metaphysical necessity'. This kind of weak necessity could be called empirical necessity. Chalmers uses the term 'nomic necessity': consciousness nomically supervenes in this world (or in some group of worlds which have some 'extra ingredient' which 'force' necessity throughout those worlds).

Zombies, or absent qualia, are a statement of inessentialism. Inessentialism rests on the assumed irrelevance of phenomenal experience. This rests on the assumed 'completeness' of functional, behavioural, and causal explanation. Most physicalist and functionalist views are inessentialist, and inherit the difficulties inherent in absent and inverted qualia.

Chapter 2

Explain

2.1 Introduction

Functionalism, of whatever form, rests on mathematics; mathematics rests on logic; and logic has counter intuitive and controversial ontological foundations. So there is a difficulty with functionalism making ontological claims. However, functionalism for the most part, does not make ontological claims; it makes metaphysical claims about the identity criteria for mental states. An ontological claim may be provided by allowing for token identity to physical states for instance, but that ontological claim says nothing about the identity of mental states.

The token-identity claim, which is compatible with functionalism, is not an ontological claim that arises from functionalism *per se*; it is an additional statement. The claim that a functionalist view of mental states is cause for eliminativism with regard to aspects of the ontology of metal states, such as qualia, is not a claim that comes from functionalism *per se*.

Function, of whatever form, is considered a third person concept in the sense of not referring to first person phenomena. Functionalism is often considered to be, or to be an aspect of, a view that takes all behaviour to be explainable in functional terms. For example, physical theory explains behaviour in functional terms; it is unlikely that a physical theory would admit to behaviours that cannot be so characterised.

Functionalism is built on mathematics, which rests on logic, and logic concedes that there are certain propositions that are *a priori*, that are prior to that provided by experience. Logical positivism concedes this point. This is part of the foundation of functionalism, so the fact that functionalism does not explicitly refer to these is not a concern. A concern is that considering functionalism as indicating that everything can be treated in a third person absolutist manner ignores functionalisms foundations, as it ignores questions of the *a priori*.

Dennett's functionalism does not refer to the *a priori* (Dennett 1991). These things cannot be dealt with in a third person empirical manner. Yet, Dennett's views rule out anything and everything that cannot be dealt with in a third person manner. Where does the certainty of his extreme third person absolutism come from?

Functionalists of this persuasion could argue that the *a priori* is analytic, that it is not instructive, that it does not provide more than what empiricism provides. With such an argument, there would be more reason to accept ontological eliminativism. Such an argument would lessen the importance of non-third person items (such as the *a priori*), and so lend credence to a third-person absolutist view.

Empiricism does not tell us of necessity, as how can particular instances tell us of necessity, how can some possibilities tell us of all possibilities? The analytic does not tell us of necessity, yet functionalism rests on mathematics, which talks of necessity. If we know something of necessity, it is from the *a priori*, from those non-empirical, non-inferential, items which the third person absolutist presumably wants to be rid of. Arguing that the *a priori* is analytic, however, is not easy. One needs to show that these *a priori* propositions are merely propositions by virtue of the meanings of their terms, that they are true just because their negation is false, and that there is nothing 'new' in them.

Dennett does not accept yellowness beyond the report of yellow, or the belief of yellow, but does he accept that $2 + 2 = 4$? He presumably accepts that it is so, in this case, in this world. Does he accept that $2 + 2 = 4$ necessarily, and if so, what is the justification? If necessity goes so do modal logic and modal reasoning. If these are to be kept, then something prior to a third person view, prior to empiricism, must be accepted; and this will not be explained within the context of a third person or empirical view.

Getting rid of yellowness because it seems innate, non third person, and too liberal of ontology, is fine. However, there are other liberal items of ontology that seem innate and not third person, though they are not as immediate as supposed qualia. How can one say that believing in ontological experiences is naive, and yet implicitly accept an innate, non third person, non empirical *a priori* proposition that tells that $2 + 2 = 4$, and is so necessarily? It seems that $2 + 2 = 4$ necessarily, and it seems that there are ontological experiences. Eliminativism with regard to the latter may be appropriate, but what of the status of the former?

The Peano axioms, and ZF set theoretical foundations can be taken as is, in which case the ontological and epistemological aspects are not explicitly addressed. However, these foundations must be justified, even if only in a pragmatic way. Moreover, accepting mathematical necessity is as difficult to justify as accepting that there are ontological experiences. Similarly for the converse: getting rid of the first person *a priori* is as difficult to justify as getting rid of first person ontologies.

Functionalism cannot provide its own foundation. Concerns of the foundation are ultimately epistemic. The acceptance of the axioms, of the particular set theory axioms we accept is an epistemic matter. If these foundations are taken as-is, it is not considered an epistemic matter. If the foundations are taken as-is, then there is no direct reference to epistemic concepts. The foundations are ultimately epistemic, but this is not to say that there is an epistemic account.

The epistemic justification of the foundations can be purely pragmatic: if it works, it is justified. There need be no further epistemic discussion. Naturalised epistemology is the attempt to bring the justification of foundations into the system, which was built on such assumptions. Epistemology of this sort can seem circular, or it can seem to deny any coherence in epistemology aside from the pragmatic epistemology it advocates. The circularity was dismissed by Quine. Yet, in naturalized epistemology there is an admission that there is an epistemic problem.

Perhaps it is not relevant to point out these foundations and ask for justification. Perhaps it is compatible with third person absolutism to accept mathematical necessity, although third person absolutism cannot provide an account of necessity. Perhaps, but the point is that there are still implicit ontological assumptions being made, and statements taken at face value, for which there are no epistemological analyses forthcoming.

As Lewis says, “it’s too bad for epistemologists if mathematics in its present form baffles them, but it would be hubris to take that as any reason to reform mathematics”, and “our knowledge of mathematics is ever so much more secure than our knowledge of the epistemology that seeks to cast doubt on mathematics” (Lewis 1986, 109). This is the way it is taken, and for good reason. Nevertheless, the result is that in a single breath, phenomenal realist epistemology and phenomenal ontologies can be dismissed, while strong statements about modal realism and necessity are expressed. What can be concluded? Do not dismiss ontologies easily, nor dismiss statements of epistemic certainty easily. “Phenomenal realism is so” is dismissed by some, who then assume they know with certainty that $2+2=4$. But, as Lewis asks, “Can you *really* not know that $2+2=4$ . . . I doubt it.” (Lewis 1986, 133).

Lewis’s views in this matter lead him to a very strong modal realist view. One thing is sure, where any shade of mathematics is introduced, then strong statements follow, and these may be statements regarding ontology and epistemology, but they will not have an epistemic justification. Functionalism, then, inherits this liberty.

2.2 Effectively computable and computable

2.2.1 Computable

As to what mathematical reasoning is, there is no consensus. There are many views on this matter. Each one addresses (or does not address), either explicitly or implicitly, the ontological and epistemological issues underlying it. A Platonistic view of abstract items is strongly ontological; it is strongly epistemological too, if we are concerned with the justification for the abstract ontology. Intuitionism explicitly rejects an abstract ontology. This is reflected in the fact that intuitionism accepts only constructive proofs (for what do the other proofs refer to, if there is no abstract ontology?), or in the abandonment of the law of the excluded middle. Intuitionism fell out of favour. Most of us hold, implicitly, a strong Platonic concept of abstract entities in this regard. As mentioned previously, most of these concepts are considered prior to even epistemology. Our intuitive notions in this regard are, however, fallible.

The wrapping of abstract entities in sets has caused difficulty in the past. A decision as to which set theoretical axioms to choose had to be made; and the restriction on sets that are 'too big' indicates to us that our intuitive notions of abstract mathematical entities are not perfect. Concerns over the justification of axioms can be considered an epistemological matter, or a pragmatic matter; yet, matters of pragmatism are implicitly epistemological. In addition, there is the status of the *a priori*, over the justification of necessity; although this and all other foundations can be taken as-is, and the epistemological justification or the implied ontological commitments can be left aside.

We are part of the world, and mathematical reasoning occurs in ourselves, so it is *at least* related in some way to the world. And this is reflected in the fact that we can use mathematical reasoning to build functionalist accounts of aspects of the world.

In the past, whilst mathematicians agreed on most points, subtle mathematical proofs became cause for disagreement. Without specific knowledge of the nature of mathematical reasoning, there was no means to justify a subtle proof. Without that, there is no clear way in which to resolve a disagreement between someone who claimed to have a proof, and the opponents who contended that it was not a proof.

The question as to what constitutes proof in mathematics and how such a proof is to be recognised needed answering. That this needed answering was made explicit by Hilbert. His question as to whether it was possible became known as Hilberts 10^{th} problem, being one of a list of problems (or questions) he posed at the 1900 Paris International Conference of Mathematics. The tenth problem asked whether there was a way to decide whether an algebraic equation has a solution in whole numbers. Essentially, he asked whether it was possible to determine whether a given mathematical statement was true or false.

Hilbert attempted to build a formal axiomatic system that would allow the construction of well formed mathematical statements that could then be checked for theoremhood. Hilbert wanted such a system to be complete, that all well formed statements would either be theorems, or be theorems if they were negated. If the statement is a theorem, then this was to be final, and if it were not a theorem, then the contrary assertion would be a theorem. And the system was to be consistent; if something is not a theorem, then the statement that it is not a theorem is a theorem. More formally: a certain statement is not a theorem, but its negation is a theorem. Formal axiomatic systems such as this do not have procedural rules of proof: there is no rule that tells us what to do next when we are attempting to construct a proof. However, theorems can be constructed indirectly, as existing well-formed statements could be checked for theoremhood, as Hilbert envisaged a decision procedure for this purpose. With a decision procedure, disagreement between mathematicians over a purported proof could be resolved in a manner acceptable to all. The checking of theoremhood is procedural, and the construction of well-formed statements is procedural; thus all well-formed statements can be listed and checked for theoremhood. The construction of theorems is not procedural, but the decision problem—showing that a given well formed mathematical statement holds—is. The British Museum method will find them from a list of well-formed statements.

The extent of Hilbert's system was to be such that it encapsulated all of what was understood to be mathematics. Hilbert's system was to be a product of mathematical reasoning that encapsulated mathematical reasoning. Hilbert equivocated on the truth of mathematical statements and their theoremhood. Truth was to be proof, and what was provable was to be true. Hilbert wanted truth (proof), the whole truth (completeness) and nothing but the truth (consistency). If a statement was provable, then its negation was not to be provable; this consistency he got, but not completeness.

Hilbert's assumption that mathematical truth and proof can be equated did not work[1]. That mathematical truth and proof are different requires a demonstration that there are well-formed statements that need not necessarily have both the characteristics of truth and proof.

In order to construct such a statement, a high degree of expressive power is required within the axioms and foundations upon which such a statement is based. Without a sufficient degree of expressive power, it is not possible to separate truth and proof with regard to well-formed statements. Propositional calculus does not have the flexibility to construct well-formed statements that could express such a division. Propositional calculus is complete and consistent.

For any countable set, there is a one-to-one function from that set to the set of natural numbers; thus, every element of that set can be encoded by a unique integer. Gödel numbering is such a one-to-one assignment of a subset of the natural numbers to elements of countable sets, with some conditions (Godel 1931). There must be an algorithm to calculate the function; an algorithm to test whether a natural number is the Gödel number of some element in the set (the range of the function is not necessarily the whole of the natural numbers); and finally, a method to determine what the corresponding element of a Gödel number is (the inverse Gödel function).

Gödel numbering allows the encoding of various discrete structures (such as graphs,

[1] If Hilbert had been correct, a system such as he envisaged would have encompassed *all* of mathematical truth and proof, and all mathematics would be, as Wittgenstein said, tautology: if the system is complete and consistent, then the decision problem is solvable, and the whole enterprise becomes trivial, because there is a procedural way to settle any question that can be formulated within the system

or tuples) in integer form. A Gödel number can be part of a well-formed statement, and thus the statement can refer to the element of the set the Gödel number represents. And so Gödel achieved the task of constructing a well- formed statement which referred to the statement of its own proof. The statement actually referred to the statement of its unprovability. The statement does not refer to itself directly. It says that if you perform a certain procedure to calculate a number, this is the Gödel number of a statement which cannot be proved. The number that is to be calculated is the Gödel number of the entire statement. This form of self-reference is possible because of the expressive power of mathematics. A system would need to be expressive enough to allow Gödel numbering, and this essentially means any system in which it is possible to deal with the positive integers.

A well-formed statement, which refers to its unprovability, is not a simple matter. It is easy to say 'we can see that the statement, "this statement is unprovable", is true'. This is a somewhat Platonic conception of truth. It is problematic to make such a vague claim. The statement is one that seems correct to us if indeed it is unprovable. The statement "this statement is unprovable" is hardly 'true' if we can verify that the statement a theorem. It is necessary to accept that the system is consistent; then we can consider the statement true.

The Gödel statement is not some link to abstract mathematical truth; it simply refers to the fact that, if the statement is unprovable, then the statement that the statement is unprovable seems 'true' to us. The statement must be considered in the context of the system in which it was constructed. The statement is true because it refers to its unprovability in the consistent system of mathematics. The Gödel statement is meaningless if indeed we can prove it (which we cannot do in the system of mathematics, as it happens). It is 'true' if it is meaningful, and it is meaningful in the context of the system in which it is constructed.

It was the vague interpretation of self-referential mathematical statements of the sort Gödel created that lead to sustained debate questioning whether our minds are not limited in the way that mathematical systems are. These debates hinge on an informal version of Gödel's result: "we can see that the Gödel statement is true, but the system cannot prove it". The 'truth' which Gödel showed is simply the meaningful correctness we give to a statement which is well formed within a system which is consistent. This Gödel 'truth' must, and can only, be considered in these terms. Indeed, Gödel's work indicated that our intuitive notions of 'truth' never quite turn out as expected.

As the Gödel statement is unprovable, yet expresses the 'truth' about its unprovability, there is a disjunction between mathematical truth and proof; and with true statements which are unprovable, the complete system of Hilbert is not possible. Of course, in an inconsistent system anything can be proven, so there is a choice: consistency or completeness, and it is an easy choice. Gödel considered whether or not theorems were provable.

From Gödel's result alone, we know that consistent systems cannot be complete. Turing showed that there was no decision procedure to show the truth of any randomly chosen mathematical statement (Turing 1936).

In the paper where Turing introduces the halting problem (though he did not use this term) for his demonstration that there is no general decision procedure, he was talking about computable *real* numbers. His argument is essentially Cantor's diagonal method applied to the computable reals. The computable reals are a denumerable subset of the reals.

Turing considered a list of all functions from the real numbers to the real numbers. He specified an abstract Turing machine for this purpose. Next to each program is the real number it generates. Following Cantor, he constructed a new number using the diagonal method. He took the first digit after the decimal point of the first number in the list, and changed it. This number becomes the first digit of a new number. Then he took the second digit of the second number, the third digit of the third number, and so on. The new number so constructed, with a decimal point in front, will be a member of the list of numbers generated by programs. It does not matter if the N^{th} program does not produce the n^{th} digit in the number it generates, as any number can be chosen as the n^{th} digit of the new number. The resulting number will still be different from all numbers in the list.

The new number, not being a member of the list of computable reals, is an uncomputable real number, so there must be a reason why it cannot be computed. The construction of this number is essentially the step, "take the N^{th} program, and take the N^{th} number it generates, change it, and print it out as the N^{th} digit of a new number". This is an algorithm for producing an uncomputable number, so there must be a problem; and the problem is not with simply changing a number. The problem therefore lies with the apparently simple task, "take the N^{th} program, and take the N^{th} number it generates". So there is a difficulty with the general task of getting the N^{th} program to generate the N^{th} digit.

A possible difficulty would arise if it does not produce an N^{th} digit. This is possible, as a program may not produce an N^{th} digit, it may simply produce only $N-1$ digits. If it does not produce an N^{th} digit, there is no difficulty; as it does not matter to the diagonal construction if an N^{th} program does not produce an N^{th} digit. However, we need to be definite that the N^{th} program does not produce an N^{th} digit, before moving on to the $N+1^{th}$ program, in our program which carries out the diagonal procedure.

Thus, if we could answer the question, "Does the N^{th} program produce an N^{th} digit", we could have a program for constructing an uncomputable number. The program would operate in this way. It would take the N^{th} program and wait until it produces an N^{th} digit, or verify that it does not produce an N^{th} digit. Then it would change that digit if it is produced, and use it as the N^{th} digit of a new number. The program would repeat that process for every N. However, we cannot answer the general question "does the N^{th}

program produce an N^{th} digit". Obtaining the N^{th} digit of the N^{th} program is a specific case of the halting problem: one cannot, for all N, get the N^{th} digit of the N^{th} program, as that program may not halt.

If Hilbert turned out to be right, and there were a consistent and complete mathematical system with a decision procedure, the halting problem would be solvable; we would have a general mechanical procedure for determining whether a given program halts. This is because we would be able to run through all possible proofs until we found one that the program halts or does not halt. So the decision problem, incompleteness, and the halting problem are intimately related.

Algorithmic action is defined in terms of Turing computability, or other systems, such as that of Post, which are equivalent to Turing computability. Yet at the time, when the equivalence between these systems was not known, there was a question as to whether Turing computability captured all that is 'effectively' computable. The term 'effectively computable' was used to refer to those computations that could be carried out mechanically.

The conjecture as to whether all that could be regarded as effectively computable is encapsulated by Turing computability is known as the Church-Turing Thesis. This thesis was introduced by Church in 1936 (Church 1936), and arose from the work of Gödel (Godel 1931) and Kleene (Kleene 1936). It is the thesis that equates 'effectively computable' with 'Turing computable'. The Church-Turing thesis as formulated, is not one that can be answered, because there is no rigorous formal definition of what 'effectively computable' means. Evidence, however, supports the thesis. Other conceptions of computable systems, such as Post machines (Post 1936), or the lambda calculus of Church (Church 1941) compute the Turing computable functions. No one has arrived at an intuitively computable function that is not Turing computable. So 'effectively computable' is now taken to mean 'Turing computable'.

2.2.2 Effectively computable

Do we want algorithmic action going on in the heads of mathematicians for which we can find no physical analog? Do we want 'effectively computable' to refer to persons only, and not physical devices? That would entail some startling facts about persons. If this possibility is ruled out, then what is effectively computable must have a physical analog; it must be related in some way to the physical world. If this were not the case, then there would be something going on in the heads of persons for which there is no physical analog; there would be effective computations restricted to persons.

The abstract reasoning in our heads is related to the actions of physical processes. Certainly, the effective computations we carry out are related to the actions of physical process. Thus, it is to be expected that abstract reasoning may indeed be able to uncover what is physically effectively computable. No one has yet uncovered an intuitively effectively computable function for which there is no physical analog. As far as effective computability is concerned, what is effectively computable in our heads is also physically computable in the world. If it can be computed by us, it can be computed in the abstract conception of a Turing machine, and this has a physical analog; the abstract mechanical process can become a physically mechanical process.

The effectively computable is related to what is physically possible. Our acceptance of Church's Thesis and the fact that abstract conceptions of computable can be physically realised indicates that we accept this. Consider the computable function addition. For addition to be effectively computable in this world, the world must allow physical processes that would allow for the realisation of addition. The world would need some physical process to which we could attribute the abstract computable function addition

Conceive of a world in which there are no discrete things. In such a world, there is no physical analog of the natural numbers. Moreover, there is no physical analog of addition in such a world. Addition would not be effectively computable in such a world. Could there be mathematicians in that world who considered addition computable? Perhaps, but could there be mathematicians in that world who could effectively compute addition? No, not unless they were somehow apart from that world. The mathematicians in that world would not consider addition effectively computable, and thus would not consider it a computable function.

Regardless of computability, regardless of abstract ontology, what is effectively computable is related to what is physically possible. If we can effectively compute a function in our heads then we must, in principle, be able to realise this function in some physical device. Let us say that we find an effectively computable function for which there is no physical analog. This would suggest that we are somehow more than, or not completely restricted by physical possibility in this world.

Deutsch (Deutsch 1997) advocates an extremely physically grounded view of computation. He takes it that computation is a physical process, and as such, what is computable is determined by what is physically possible. Given that he accepts this view of computation, it is odd that he also claims that Turing could not determine, through his abstract pondering, what is computable. He makes this claim because of his position that the physically possible determines what is computable. However, in the claim he makes a distinction between abstract and physical which, since he argues that what is abstractly computable is physically computable, is invalid. His argument is that what is physically possible can only be determined empirically, and since what is physically possible determines what is computable, then what is computable can only be determined empirically. Yet, in his view, abstract reasoning is physical process, so abstract reasoning could possibly determine what is physically possible. His view that Turing was working "in the wrong direction" is therefore meaningless. The work of Deutsch suggests that he is a modern day intuitionist, in that he explicitly rejects a Platonistic abstract ontology, however, he does not take on the restrictions of the old intuitionist school. To Deutsch, everything is what is physically possible; there is no purely abstract reasoning and no purely abstract entities.

Deutsch claims that the physically possible determines what is abstractly algorithmically possible; that what is physically computable is what is abstractly effectively computable. The more common view is that what is abstractly effectively computable is related to what is physically possible because abstract mechanical computing systems can be physically realised, or form the basis for physical realisations.

Deutsch, however, does accept Church's Thesis; he accepts that Turing computability encapsulates all of effective computability. Since to Deutsch, effective computability is physical possibility, he makes the claim that all that is physically computable is encapsulated by the system of Turing.

The ontological world according to Deutsch does not contain uncomputable numbers, though it certainly may contain persons who do mathematics involved with what they think of as uncomputable numbers. However, the symbols for these numbers, and their thoughts of these numbers, though they are considered to refer to such numbers, do not so refer. The computable function which doubles the number of discrete things in existence is a computable function, but it does not refer to a number, as there are no abstract entities, and the world lacks that many discrete things. Computable functions, which are not effectively computable, must be regarded as potentially effectively computable functions, and not considered in an abstract Platonic manner, in the view of Deutsch.

To Deutsch, well-formed mathematical questions may not have an answer. This is akin to what the intuitionists believed. They did not subscribe to any transcendent notion of numbers. To an intuitionist, the well-formed mathematical question "What is twice the number of discrete things in existence" is not a question to which there is an answer. If we assume there are only 12 discrete things, then in this conception there is no answer to 12+ 12. In what sense, then, is this a computable function? It cannot be effectively computed, and abstract ontologies have been explicitly rejected. We are conditioned to believe that there is an answer to this question as the question refers to a number. But that number cannot exist in a physically grounded way, it cannot refer to anything, because there is not a collection of things big enough to which it can refer, and abstract numbers are rejected. To distinguish effectively computable and computable means implicitly advocating some form of abstract ontology.

Functionalism comes with epistemological and ontological assumptions. However, these are not made explicit: mathematics does not need epistemology in that it is taken as is, without an agreed epistemic analysis. Functionalism is concerned with some form of computation or functional role. Some forms of functionalism view this functional role at a high level, as in the case of folk psychology functionalism. Other forms are nearer to a computational conception of functionalism. In both cases, the function or computations that matter are those that can be carried out. In computational terms, effective computations are required (though effectively computable and computable are now considered the same). There is a degree of multi- realisability if only in the sense that functionalists do not want functions which can occur only in the heads of persons. This would be a distinction between computable and effectively computable, and that would make strong claims about our nature, as opposed to the nature of the non-human physical world. The implied ontological commitment is to something that supports functional organisation.

Functionalism answers the question of the identity criteria of mental states in functional terms. But this does not necessarily say what mental states are specifically. That 'something supports this functional organisation' is the indication that what mental states are, is a question to be answered.

2.3 Assigning functional role

2.3.1 Functional descriptions

Pragmatic empiricism is concerned neither with the attendant notions of perception and observation, nor with any notions of realism. Empiricism neither transcends nor denies such issues. It is independent of explicit ontological commitments, and is not explicitly in conflict with particular ontological commitments. The complex ontology underlying logic, the questions of the synthetic/analytic *a priori*, and modal issues, will not be mentioned here.

I will consider functional explanation in a formal and abstract sense. I will consider it a tool to form explanations from empirical data. These explanations will be formal and abstract. I will start with the explanation of empirical data in an abstract manner. It is common that mathematical explanation of this type results in ontological commitments. The form of the explanation, where it is interpretable as 'functions' or 'rules' may gain ontological status. Such is the way with the concept of 'laws of physics', whereas in actuality, what a 'law of physics' means, aside from a formal abstract mathematical concept, is vague. Such 'laws', if considered in this way, take on a Platonic ontological flavour, and there are modal questions regarding their applicability in other logically possible worlds. But I shall not be making further comments along these lines.

Data is obtained by observation, but the nature of observation does not matter. If it did matter, it would be an ontological and epistemological concern, and I am taking a pragmatic stance. Observations are to be represented formally and abstractly. This is to say that they are removed from their particular ontological basis. The representations of observations are labels. What is of the observation beyond the label that it is given is not a concern in this context. The set of abstract formal labels needs to have some order, so elements of this set can correspond to a set of empirical observations. I will consider the case where labels are numerical, as natural numbers are conveniently ordered.

Once the labels are assigned to observation, the labels themselves become important, not what they stand for, refer to or designate. It is not the labels themselves that matter but the relationships between them. Addition works for us, who believe we have understanding of '2' in a physically grounded way, or perhaps a set theoretical way, or indeed a Platonic way. However, calculators work also, and they do not have any such concepts. The task is one of explaining a series of observations represented by a string of natural numbers, where each digit of that string corresponds to a particular observation. Mathematical functional explanation is applied to this string. The explanation is therefore concerned with the number string, not the observations, and not 'the world'. Where ontological commitments can arise is when an explanation of data is used to infer an ontological story about the world.

Consider three observations that occurred in time series to which were given the labels '1', '2', and '3'. This is empirical data to be explained. Explanation entails finding an abstract formal functional explanation that fits the data. Functionalist explanations are mathematical. A functionalist explanation of '123' fits that string in that it generates that string. An example of an explanation of some empirical observations represented by the string '123' is a functionalist explanation that states "print '1', followed by '2', followed by '3' ". Another explanation is "calculate 124 minus 1 and print the result". Yet another explanation could be "calculate the number of angels on a head of a pin, and print 123". In all likelihood, '123' is a string that could be extended with additional observations. Explanations that could predict future observations are more useful. Considering '123' as part of a larger, or infinite, string, would mean an explanation of the form "each result is an increase of one over the last one" is more appropriate. Yet, any bizarre explanation, which results in '123', explains that string. The form of such explanations may seen to invoke 'functions', 'rules', or 'laws', which may lead to an ontological story being told. However, without further commitments of an epistemological and ontological sort, or without looking at the form the abstract functionalist explanations take, there is no ontological story. Functionalist explanations, which have the same results, are equal, if the task is to simply produce results.

Explanations are themselves abstract and formal, just as the data they explain are represented in an abstract and formal way. As ontological concerns are ignored, the formalism in explanations has no reference to 'the world', just as the natural numbers, as we are considering them, are nothing but abstract entities. The explanation, being abstract and distinct from any ontological grounding, can be represented in many different abstract ways. One way is in the form of natural numbers. In this way, an explanation of a string of numbers can itself be represented as a string of numbers. The explanation of the string beginning '123' can be represented numerically. The numerical representation of this explanation can be considered in the same manner as the original data string. Thus, the explanation can itself be explained. There is a string, an explanation of that string (a generator of that string), and a further explanation of the original explanation. If this succeeds, the second explanation can replace the first explanation. The second explanation explained the first explanation, and the first explanation explains the data string. The second explanation explains what the first explanation explains; it explains the original data string. Further explaining of explanations is one way to get neater explanations.

A way to compare explanations, apart from their correctness in producing required results, is to compare their sizes. When represented numerically, comparing two explanations is easy. Either they have the same number of digits, or one has less than the other does. Therefore, this is a criterion with which to judge explanations. I will say that the numerical form of any explanation will be shorter than the data it explains. I will say that if this were not so, it would merely a larger way of representing the data. Both the longer and shorter cases are representations of the data, but I choose to call shorter representations explanations. In practice, shorter explanations will be the norm because data is usually part of a potentially infinite string, and explanations need to be finite. Considering a potentially infinite data string, the finite explanation that can generate this infinite string can be considered a compression of the string. If compression—explanation—is possible one can say that the information content[2] of an infinite string can be contained in a finite string. Quantifying information content is possible, but it is dependent on the particulars of the abstract mathematical formalism used in explanation[3]. For relative information content, we can compare the lengths of two strings. Length can be considered the length of the string in digits (the examples here are in base ten), or the length of the string when represented in binary. Consider the case in which a finite segment of a potentially infinite string has an explanation which can be represented by a numerical string of length n. This explanation generates the infinite string. I am assuming that it could potentially generate the entire string and not just the finite segment we used for explanation. Thus, the infinite string has an explanation that is a finite string. It can be said that the information content of the infinite data string is not more than n. In non-quantified terms, we can say that the information content of string A is equal to the information content of the shorter string B, if the string B generates the string A. Many relations are possible. One can consider the mutual information content of two strings. This will usually be less than the sum of their information content, as they will usually have something in common. That is to say, having an explanation of one string will make explaining the other string easier. One can usually combine parts of the explanations of

[2] This discussion concerns Algorithmic Information Theory (Chatin 1987)

[3] Chaitin has changed the particular formalisms during his research in this area, improving on his earlier work, which had some difficulties (Chaitin 1995) in the area of program concatenation, and thus difficulties with relative and mutual measures of information.

both strings, and in so doing make the combined explanation shorter. Thus, there can be measures of how much 'cheaper' it is to explain strings together rather than separately. There are other relations, such as relative information content. Given an explanation of a certain string, we can use this as an aid to explaining another string. We can ask what the information content of a certain string is, given some other string. Consider the infinite string that begins '234876234987786'. This finite segment of the infinite string is used to explain the infinite string. The explanation will necessarily be based on a finite segment of an infinite string, and so it may not generate accurately the entire infinite string, but I will assume that the explanation predicts each additional digit with accuracy. Let us say that the explanation that generates the infinite string is represented in the finite string '3467'. This finite string can generate the infinite string that begins '234876234987786'. That infinite string has been compressed into a finite short string. Now, the finite string '3467' could potentially be explained; it could potentially be generated by an explanation represented by a shorter string. But a shorter explanation of '3467' may not be found. There are two possible reasons for this. We may not be able to find it, or it may not exist. If we find a shorter string, we can in turn attempt to explain that string. If the process of taking strings and explaining them, and taking the explanations (represented as strings), and explaining them in turn is repeated, eventually there will be a string that we will not be able to explain. The explanation of such a string would be longer than the string itself. That is to say, we cannot shorten certain strings. It is not possible that the infinite string which starts '234876234987786' is generated by the finite string '3467' which is generated by the shorter string '4' which is generated by ''. It is self evident that shortest strings exist. If it were not so, something could be explained by nothing. There are finite shortest strings and there are infinite 'shortest' strings and here, the term 'shortest' lets me down, and I should use 'incompressible' instead. Replacing strings with shorter ones is a process of compression, and this works if the information content of the original string is less than or equal to the information content of the shorter string that replaces it. For any given string, which is not a shortest string, there is at least one shortest string that generates it. This is a process of compression, and compression must eventually stop.

There is information in a string, and the string may contain redundancies. If it contains redundancies it can be explained and can be compressed (it is the redundancies which would allow us to see 'order' in the string). There is a 'shortest' string, a string that is incompressible, which will contain the same amount of information as the original string, but in a form in which there are no redundancies.

A string in which there are no redundancies is one in which there is no order. It is not merely that there is no discernible order, but that there is no order. Incompressible strings have no order, which is why they cannot be explained and made shorter. They can be considered unexplainable. In information terms, they are informationally maximal. They contain the most information in the smallest space. Here is where one factor of our common sense notions of 'information' fails. We tend to equate 'information' with order. However, there is no order in an informationally maximal string. Moreover, as there is no order, it will seem chaotic to us. It will seem random. We may wonder what information could there be in a maximal string (a string which is 'random' and will appear so to us). And the answer is, so much we cannot comprehend it. So much that we cannot explain and compress it. There are no redundancies upon which to base a 'fit' to a maximal string. If a string were ordered, it would not be random. If there were order in the string, then a shorter explanation would be possible. Without an explanation of a string it is not possible to predict or generate future members of a string, and so, an informationally maximal string will appear random[4]. Each additional digit of the string adds information that was not there already. If it did not, an explanation would generate this digit, and thus show that it does not add additional information. Informationally maximal strings are beyond functionalist explanation. Whether there are processes in the world which generate maximal strings is unknown. Types of functionalism which rest on the view that there are functionalist accounts for all behaviour would rule out processes that generate maximal strings. There is an instance, however, which is considered as being a candidate for a process that generates maximal strings. Radioactive decay is considered a statistical process. We do not have a deterministic account of it. That is not to say that such an account is not possible. If, however, it were taken that radioactive decay is genuinely a non-deterministic process, then it is a process that would generate an informationally maximal string. I say that it would generate an informationally maximal string, because I am considering the infinite string that it would eventually generate. In computational terms, maximal strings cannot be computationally generated by an algorithm shorter than the maximal string. The set of infinite maximal strings, represented in base ten (in the natural numbers) is the set of uncomputable natural numbers.

If processes that generate maximal strings exist, there are some difficulties. We would only ever observe a finite segment of the string it generates. A functionalist account will always explain a finite string, even a maximal string, though in such cases, the functionalist account will contain more information—it will be longer—than that of the finite segment. However, as this none too short explanation would be seen as a finite explanation to a potentially infinite string, it may be accepted.

We can never know that what we are observing is part of a potentially infinite maximal string. Thus, using the example of radioactive decay it is not possible to say for sure

[4] An accessible explanation of Algorithmic Information Theoretic randomness is (Chaitin 1988). The implications this understanding of randomness has been captured by (Stewart 1988) and (Gardner 1979), when they consider a particular random number that Chaitin defined in terms of the answers to mathematical questions. The number is thus a representation of the answers to all mathematical problems in the shortest space. Unfortunately, being random, it can be defined, but not computed, and if found by accident, we could not verify that it was that number.

that it is a non-deterministic process. Despite the fact that we have tried and failed to find a deterministic account, we cannot use this to make a definitive claim that it is non-deterministic. We cannot prove, cannot verify, that potentially infinite strings are maximal. To do so would be to state, "on the basis of what I have seen, further parts of this string will not add information that isn't already there". It is clear that this is impossible. The string is maximal, and so there is no pattern, correlation, or structure. There is no order to see in such a string. There is nothing on which to base such a judgement. Further parts of a maximal string are not related to previous parts. We can only verify that our predictions, so far, are accurate; or contrariwise, we can only know that we have failed to provide a deterministic account. We cannot know that we are observing a part of a potentially infinite maximal string, but we can know if finite strings (and this can include a segment of a larger string that we are considering in isolation), are maximal. Finite maximal strings can be part of larger strings that are not maximal. For instance, a finite maximal string could be part of a larger non-maximal string if that larger non-maximal string is just a repeated series of the shorter maximal string. If a string is finite, and we consider it in isolation, we can find out if it is maximal. The way to decide if the string is maximal is to generate all possible strings that are shorter than the string we are testing, and verify that none of these strings generate this string. An example of a finite informationally maximal string is '1'. An example of a non-maximal string is the infinite string '11111111. . . ', though it can be considered a string of maximal strings. I mentioned mutual information and other relative measures above, and this is an example. It can be cheaper to explain non-maximal strings together, rather than separately, as explaining one will form part of the explanation for the other. In explaining things in the world, there are no definite finite strings as there are always further observations that could be made. If the world can generate maximal strings, for instance, if radioactive decay is a genuinely non-deterministic process, we will not know this through empiricism and functional explanation alone. We may be able to know this in some other way, but it will not be through attempting to verify such strings directly. Mathematical explanation, and hence all functionalist explanation, cannot verify that finite strings are part of infinite maximal strings. Thus, functionalist accounts will be seen everywhere, if that is what one looks for. All things will be seen to be implementing functions, or to have functional accounts attributed to them, if that is what one searches for. This is so even in the case of processes that could in principle, generate an infinite maximal string. Functional explanation will therefore never cease. There will always be a continuous stream of new 'observables' which are deemed important, which will be used as data for functionalist explanation. New and previously overlooked aspects of what is to be explained will continue to emerge as new data for new functional explanations. Functional explanation, considered this way, is just description. However, there is a way to bring an ontological aspect to this mathematical, functional description. I shall consider a path to the ontological aspect; a path from a functional explanation of some aspect of the physical world.

2.3.2 Explanation and prediction

Functionalism is about abstract descriptions of a described. Functionalist accounts need make no explicit ontological commitments. Ontological commitments can be made regarding that to which functionalist accounts are applied. But, alone, functionalist descriptions are abstract. Functionalist accounts predict what can be observed from the described. Functional description tells us about the described by telling us about the abstracted representation of the observables of the described. The functional description just describes a representation of the described. It does not describe the described directly, it describes a data string. However, in functionalist explanation, there is always more than just the data to be considered. There is a background context, and that can include ontological commitments; it may also include restrictions as to how functional explanation is attributed to particular things.

The dripping of a tap, a double-jointed pendulum, and the noise on a phone line are three examples in which there is no functional explanation that allows for accurate prediction. It is practically impossible for functionalist accounts of a double-jointed pendulum to predict where the end of the pendulum will be one hour after being released from a known position. Yet, we accept that we have a functionalist explanation of the actions of such a pendulum. The failure of functionalist explanation to predict does not matter to the acceptance of certain functionalist explanations. If data alone were considered in the absence of a large background context there would be little justification for accepting any functionalist account that does not predict with sufficient accuracy. In a background context, prediction may become a secondary criterion of the success of functionalism. The form a functionalist explanation takes, is as, or more, important than its predictive ability. Predictive failure of functionalist explanation may be traced away from the functional explanation itself and onto the context in which the explanation resides. In the case of the pendulum or the dripping tap, the difficulty is one of the accuracy of measurement. The functionalist explanation, though it may fail to predict, can be accepted. In addition, the form of the functionalist explanation would indicate that there is considerable divergence between predicted positions and initial positions of the pendulum. The failure of practical prediction need not mean that the explanation itself is useless. As long as there are other reasons for keeping an explanation of that form, it may be retained.

In the discussion about data strings, the criterion for the success of abstract functionalist explanation was accurate generation of the string. Considering the data alone, there is no sense in which a functionalist explanation can be judged if it does not predict. Without a large background context, and considering data alone, functionalism is abstract description, and there can be no basis for making claims based on the form of the functionalist explanation. Yet in the pendulum case, the form of explanation is considered important.

The correct view of functionalism, if abstract mathematical thought is independent of particular ontological and epistemological commitments, is as abstract description only. The issue is one of how to relate abstract functional explanation with that which it explains. The latter will include ontological commitments of various sorts.

The view of functionalist explanation as prediction only is useless because it seeks to become an oracle. Prediction is what an oracle does, but oracles do not explain. Given only the observational string of the position of the pendulum, an oracle would tell us, precisely, the position at any time. However, if we had the data string, on its own, unaware of what generated it, the accurate predictions of the oracle would not help us understand what generates this data. Oracles can predict the future, but without telling how this future will arise. Explanations tell the future by finding out how the past and present arose, then extending this into the future. Thus, the 'form' of the explanation must be important, because there is nothing else but its 'form' which explains. Thus, the ontological and epistemological concerns of mathematical reasoning are important, for there is where functionalism can gain stature, and create a bridge to the world it describes.

Functional explanation takes on a slight ontological colour. It is not merely descriptive. There is a need for a relation between functional description and that which it describes. This is difficult terrain. We may talk of a certain thing 'having' a particular function. We may say that a certain thing implements a certain abstract function, or that it realises the function. The difference between ascribing function to something and describing it functionally is important. The constraints on how functional description is ascribed to particular objects are the essential element in functionalist accounts of mind. Each differing functionalist account of mind has different restrictions and that determine when a particular object can be said to 'realise', 'implement', 'have', or 'perform' aspects of the functionalist account

2.3.3 The ontology of function

A computer is a device, which is designed with the specific intent of implementing computations in the physical world. As such, we can say that it carries out the computation we designed it to perform. Nevertheless, computation is independent of particular physical processes, and so there is nothing about the object or its actions, which allow us to determine that it is carrying out a particular computation. Computation is attributed to physical processes. A physical realisation of a computation system is a physical object (collection of objects) that can be constrained to work in a way that easily allows the attribution of computations to it.

Claims of function, as attributed to an object, refer to particular aspects of the object.

A computer that is set up to add numbers is an object that is adding numbers only in the sense that certain aspects of that object are deemed important. That we may have built this object with the intent of creating an adding machine does not change this. The object, if conceived of in a broader way, could have many different computations and functions attributed to it, depending on what aspects of the object are considered important.

Computations are not dependent on the particulars of physical instantiations. Effective computations are dependent on their being the possibility for a physical instantiation, but not on particular instantiations. The aspects of the instantiation upon which computations are dependent to not refer to particular facts of an ontological nature. This is what universality entails. Thus, no particular ontological claims can be made from functionalism. Functionalist accounts of mind may place restrictions on the nature of the object that is the target of functional explanation. Statements, which contain both a term with implicit ontological weight and a term regarding function, are problematic. For example, the statement ‘this machine performs this function’ does not give any information regarding the status of the ‘machine’ to which it refers. The function referred to could be either an effective computation, which is not limited to that particular machine, or an abstract computable function. The machine performs the function in so far as the effective computation which ‘this function’ refers to can be attributed to it. That ‘this machine’ can be seen to perform ‘this function’ indicates that ‘this function’ is an effective computation that can be realised in other objects. As ‘this machine’ is seen to perform ‘this function’, another object could simulate the necessary aspects of ‘this machine’, and the simulation of ‘this machine’ would be seen to perform ‘this function’.

The important aspects of any object that can be seen to perform a computation can be simulated in a suitably powerful realisation of a universal computation device. By the important aspects of the object, I mean the aspects of the object that allow it to realise a computation.

If there were behaviours of particular physical objects to which a computational or functional attribution could not be made, then important behavioural aspects of that object could not be simulated, accurately and completely, within universal computational systems. Certainly, an acceptable simulation could be made of a process to which no functionalist account is attributable in certain situations, but this would be accurate only to a certain degree, in certain situations. An example of simulating a process for which there is no functionalist explanation would be generating a pseudo random sequence in place of an apparently random sequence the actual process generates. If we can attribute computations to the behaviour of objects, then the behavioural aspects of those objects can be simulated.

It is impossible to verify that the behaviours of any particular thing cannot, in principle, have an effective computation attributed them. A case in which we have failed to attribute a functionalist explanation is in the case of apparently random processes, and these can be replaced by good enough pseudo random generators. Thus, as far as we know, all types of behaviour at all levels in the physical world can be simulated in universal computation systems, contingent only on the storage capacity of those systems. This is a means of abstracting behaviour from a specific ontological base. The behaviour can be simulated in another object, a universal computation device. Moreover, the behaviours of that universal computation device can be simulated in yet another device and so on.

Functionalist explanation allows behaviour, which can have computations attributed to it, to be lifted from any specific ontological base. Thus, in a functionalist or behaviourist account, separate ontological commitments would need to be made if any ontological story is to be told. In the absence of such constraints, all behaviour can be viewed functionally, and all computations can be realised in any universal computation object. The only constraint, then, is the basic constraint of the possibility of effective computation: that computations can be realised, that it is possible to build universal computation systems. There are functionalist accounts, both of the mind and of the world generally, which do not have further ontological constraints. In these accounts, it is the function or the computation that matters, whereas in the constrained accounts, there are restrictions on that which realises the computations or functions.

In these unconstrained functionalist views, no ontological story is told. In such views, it is the computations that matter, and nothing else. All behaviour, in these views, can be functionally described, and so all behaviour can be simulated. There are no behaviours inherently tied to particular ontological commitments. In such a functionalist view, the process that carries out a computation does not matter, because the important aspect of that process is simply that it could implement a computation. So, the important aspects of that process can be simulated in another physical process (a universal computation device), and that process itself could be simulated in another process, and so on. What matters is that there is an implementing base; however, that base could itself be implemented in some other base. The 'physical process' which realises a computation could be a 'physical process' only in that it is part of a larger simulated context.

The difficulty with such a functionalist view is that the difference in meaning between the physical process which realises a computation and the computation it realises is blurred. The physical process itself could be another computation within a simulated context. It is not surprising that this blurring occurs. This functionalist view does not tell an ontological story, and so it is not expected that the 'physical process' which realise computations have any ontological weight; the only ontological weight they have is in that they can realise computations.

2.3.4 Notes on radical functionalism

The functionalist view of mind is, as with all views of mind, limited. This allows functionalist views of mind to have additional constraints, a larger ontological or epistemic context in which the functionalist view is held.

The unlimited functionalist views, which are the province of cognitive science and artificial intelligence, but not so much philosophy of mind, necessarily have difficulties. These views equate all that is important with function, and so do not have the explicit ontological grounding or other constraints which philosophical functionalism has. These views are 'complete' in that they seek to explain all behaviour in functional terms, and do not refer to items outwith the functionalist view: there are no constrains on a realising base, for instance.

Whereas philosophical functionalism would draw a distinction between explaining mental states and the brain, this radical functionalism does not. It is a view that explains the brain, rather than a view that explains mental states in a functional manner as realised by the brain. In simple terms, radical functionalism is a view that views the brain as a computer. We are computers made out of flesh and blood; and a robot with the same program as a human being would *ipso facto* be all the things that a human is: conscious, capable of feeling pain, and so on.

These functionliast views do not explicitly address any concerns regarding the foundations of mathematics upon which it is built. Nevertheless, in being a complete view, it does, implicitly, explain how persons come to understand such foundations, and what types of opinion they will come to regarding these foundations. This is so if mental states are explained completely and solely by functional role: that functional role, therefore, determines the epistemic justification for mental states with abstract mathematical content, for instance.

This radical functionalism has no answer for the question of necessity; necessity is most likely implicitly assumed in such a view, although the question of necessity need not be addressed at all. But it explains us, and so explains our thoughts on that question, as well as defining the limits of our thoughts on that question. This functionalism does not answer the question of necessity, or the ontological and epistemological aspects of the foundations of functionalism. However, it explains how we think (and thus what we can think) regarding these questions.

This radical functionalism, if not addressing the ontological issues of the mind initially, does address them eventually. It does this through its explanation of the brain, which defines its epistemic bounds. Therefore, it implicitly answers the issues that it did not need to address initially. It answers them because it defines our cognitive capacities and epistemic bounds, and so answers these questions by answering in what way we could claim to know the answers to these questions. The difficulty with such a radical functionalist or computationalist view is simply that it is all encompassing. It is a utopian project.

Functionalism itself cannot justify its foundations, so it cannot address the question of $2+2=4$ in general. Nevertheless, it uses these foundations, whatever they are. However, radical functionalism does address these issues because it addresses the manner in which we can think them, and the manner in which we can claim to know and form answers to them.

Radical functionalism is the task of providing a complete psychological description of humans. It equates us with computers, and computer states can be individuated; necessary and sufficient conditions can be made for any one particular functional state. We are isomorphic to Turing machines (with limited storage and other restrictions), and a machine table distinguishes each state from all other states. Indirectly, radical functionalism provides a complete psychological and mental description of us. Yet, it is to be noted that in such an all- encompassing context, the meaning of the "complete psychological description of humans" is problematic. It is problematic because in this context, if radical functionalism is taken to be so, we allow ourselves to be able to posses this "complete psychological description" of ourselves.

Arguments invoking Gödel sentences are used to make certain extraordinary claims concerning minds. However, they can be used in another way. Radical functionalism implicitly states that we, as probabilistic automata, can find the complete functional description of ourselves. This also implies that we would know such a description when we came across it. We would have to claim to know this description if we were to claim that radical functionalism is correct.

Putnam, who for a time was the important figure in the functionalist school, finally rejected functionalism. Originally, Putnam's functionalism was a kind of radical functionalism, where mental states we considered akin to Turing machine logical structure (Putnam 1960). Later, this direct equivalence was weakened (Putnam 1975b), while maintaining that functional organisation was still all that mattered. Radical functionalism argues that no further ontological or other constrains are required over and above functional organisation, and would thus rule out physicalist functionalism, which Putnam has argued against (Putnam 1967) .

The reason Putnam finally rejected functionalism was that he considered it unreasonable to believe that we could be justified in claiming to know our own complete psychological description. That it is unreasonable to accept that we, as computers, would find and know our own computational description.

The Gödel results can be used to make a case for this, but not completely. If we assume that what it is justifiable to for us to accept is determined by a recursive procedure which is specified in our ideal functionalist description D, then it is never justifiable for us to accept that D is our ideal functionalist description.

This is an epistemic argument. It concerns whether or not we can know that radical functionalist claim is justified. It does not concern whether or not radical functionalism is correct. It is an argument that concludes that, if we are computers, then we can never know what our ideal computational description is.

The more common arguments that invoke Gödel are very different. The argument described briefly above concludes, "if we are computers, we are not going to know our formal computational description, and could not verify it if we were given it". The more extreme Gödel based arguments conclude that we are not limited in the same way as formal systems (or computers) are, in regards to what we can justifiably claim about formal systems. There are several vagaries in the first epistemic argument. Firstly, that argument only applies if we assume we are consistent functionalist devices: a justifiable claim has to follow from our recursive description. If we are equated with probabilistic automata, then the situation is different.

The argument holds in cases where it is assumed that our justified beliefs are the result of innate 'rules'. Putnam argued against Fodor's language of thought hypothesis using such an argument (Putnam 1985), an argument which formed the basis for his rejection of functionalism.

2.4 Conclusion

Empirically, functional organisation depends on what one looks at; it is not, from the empirical point of view, an intrinsic fact about an object. Empirically, it meaningless to 'find' the function an object performs; this must be assigned to the object. Empirically, it is not possible to verify that behaviours are *not* computable. That means it is not possible to determine that behaviours are random, or purely statistical. Although this is preliminarily accepted in the case of radioactive decay, it is not verifiable. Thus, empirically, the world is full of functions, full of physical behaviours acting like computers. Even if the world contains processes that could generate infinite maximal sequences, we cannot verify this. Thus, regardless of the fact of the matter, everything will empirically be seen to play a computable, functional role. This implies there needs to be something non-empirical, and non third person, about choosing functional roles in functionalist accounts of mind. For this reason, folk psychology, or common sense functionalism, is acceptable: the functional roles must come from somewhere, and there is a difficulty with relying purely on empirical knowledge, so common sense functionalism takes them from our experience. This also means that strong eliminativism with functionalism is somewhat incoherent. Third person absolutism, with functional overtones, such as the view of Dennett, is incoherent. Third person absolutism is incompatible with functionalism.

The foundations of functionalism are prior to epistemology, and thus a functional account of mental states will not be a complete account of mental states with content regarding these foundations. Radical functionalism, however, tries to be a 'complete' view in this way.

Functionalism considers that, as well as the abstract functional role, there is something which has this abstract functional role. Functionalism provides the identity criteria of mental states in terms of abstract functional role, but acknowledges that there is a separate issue of what mental states actually are. Radical functionalism, or functionalism without any constraints on realising base, tries to answer both the questions of what mental states are, and what makes a mental state the mental state it is, in terms of abstract functional role. Thus, radical functionalism is a 'complete' view. Non radical functionalism can allow that there is something about mental states which is left out by providing the identity criteria of mental states, as identity criteria is identity criteria, only. There is thus no conflict with functionalism and the need for externalist accounts of content in functionalist views. Radical functionalism, however, does have a conflict, as functional roles in the head must account for everything: it is not merely an identity criterion.

Chapter 3

Functional

3.1 Introduction

"Whenever a system has the relevant functional organisation, it has the qualitative experience in question" (Cole 1994, 297). This is a statement of pure functionalism with regard to phenomenal states specifically. Indeed, Cole's functionalism qualifies for what I have termed radical functionalism. Another statement of functionalism is that "every phenomenological distinction is caused by/supported by/projected from a corresponding computational distinction" (Jackendoff 1997, 24).

Physicalist functionalism is different in that it uses terms that are implicitly or explicitly ontological. The physicalist functionalism of Churchland is described as the view which takes it that "the essence of our psychological states resides in the abstract causal roles they play" (Churchland 1989a, 23). 'Causal role' here is implicitly ontological, whereas 'functional role' is not. Physicalist functionalism has a constraint on what can have functional organisation, which radical functionalism does not.

There are various reasons why physicalist functionalism may be chosen over pure functionalism. Physicalist functionalist accounts consider functional role as the identity criteria for mental states, but this does not say what mental states are specifically: it says what the defining characteristics of mental states are. What makes a car is that it functions as a transport device. But a car is a lump of shaped metal and plastic. Physicalist functionalists answer the latter question of what mental states are specifically, in explicitly ontological, physicalist, terms.

Functionalism, alone, allows for some looseness as to the ontological nature of mental states. One aspect of the ontological nature of mental states is the phenomenal aspect. Arguments for physicalist views attempt to point out that functionalism alone allows for missing or mixed up qualia, and that a physicalist grounding solves this problem of 'loose' qualia (Levine 1988). Some functionalists, however, reject physicalist functionalism, and merely declare that absent or mixed up qualia do not occur. The statement of Cole's functionalism above, is such a declaration; but it is not an argument.

The well-known spectrum inversion puzzle is one such case of 'loose' qualia. It argues that functional role does not determine qualia, by describing an instance in which two people are functionally alike, but have different qualitative experience. One person's colour spectrum is inverted. That person experiences 'red' when looking at 'green' grass. The experience of 'red' they call 'green'. They behave exactly like people with the more normal colour experience.

One argument against this case would claim that spectrum inversion is not possible. It may argue that if we looked close enough, we would find that persons with inverted spectra would not be functionally like the rest of us. For instance, some colours seem 'warm', others 'cold', and so on, and this would lead to functional differences. Thus, someone with inverted spectra may call their experience of some other colour 'red', but it will not affect them in the way that 'red' affects us, or so an argument could go. If arguments such as this are correct, then phenomenal experience does supervene on functional role, and the 'loose' qualia objection does not hold.

If functional role under determines phenomenal experience, then this 'looseness' in qualia requires tightening, and that requires a further constraint, over and above functional role. The 'loose qualia' argument is not an argument that phenomenal properties fall outside a physicalistic picture, it is an argument that functional role does not provide the entire story as regards these properties. Phenomenal properties are grounded in neurophysiology, perhaps, or some other physicalistic constraint, in physicalist views.

In the introduction, I described the case of the colourblind neuroscientist called Mary. Mary, in her black and white environment, supposedly knows everything there is to know about colour, yet has not experienced colour. In the introduction, I mentioned this thought experiment in the context of the modal force of facts about experience. I mentioned that this hinges on whether or not experienced redness is considered a 'fact' over and above the facts she could potentially know before experiencing colour.

In a physicalist picture, 'everything there is to know about colour' can be taken to mean 'all the physical facts about colour'. In that case, her lack of experienced colour causes some difficulty, if experienced colour is taken to be a 'fact'. If it is, then this 'fact' is not a physicalist fact, and thus physicalism is false.

Counters to 'the knowledge argument'—a name for the Mary's sitation—involve showing that the 'fact' of experienced colour is not a further fact over and above physical facts, that it is not a further modal constraint on possible worlds.

Mary's case is one that can be used as an argument that the physical picture of the world under specifies with respect to phenomenal experience, in the same way that inverted spectra argues that functionalism similarly under specifies.

This chapter shows the difficulties that arise from assuming that functional role can fully determine phenomenal experience. Hence, the conclusion will be that there is a need for further constraints. In this context, physicalist functionalism is preferable to functionalism.

3.2 Setting the scene

The basis for computational modeling rests in the assumption that the world can be simulated in universal computational systems. It is the assumption that a computational functional description can be assigned to physical phenomena. That implicitly assumes that the world can be seen to perform effective computations, in that effective computations can be assigned to the world. Physical processes can then be represented as computable processes, with universal computers simulating the behaviours of these physical processes.

3.2.1 Simulating physical processes

If the behaviour of a physical process can be considered in a computational way, then in so far as function is concerned, physical processes can be simulated. If the behaviour of physical process can be captured computationally, and physical theory is computational, then the behaviour of physical process can be simulated, to any degree of precision.

There is one instance in which this may not be possible, and that is in the case of apparently stochastic processes, such as radioactive decay. In such cases, good enough random number generators would simulate this process effectively. The other question concerns whether the empirical variables are ultimately discrete or continuous. This, however, is not an important issue, as for any case, the physical process could be simulated, discretely, to a sufficient degree (see (Feynman 1982) for a discussion on simulating physical processes, particularly for the question of dealing with apparently continuous empirical, physical-world variables). In any case, there is difficulty in proving, via empirical results only, that empirical variables are continuous rather than discrete.

3.2.2 Simulated environments

John Conway's life[1] is an example of a simple virtual environment. The Life system is represented as a two dimensional grid of occupied or vacant cells, with simple update rules governing the state of each cell. The rules relate to the number of occupied cells around a particular cell[2]

The Life system is computationally universal: it can implement Turing computations. Any computable system can thus be implemented in the Life system. An analog of a

[1] popularised by Martin Gardner ((Gardner 1971) and (Gardner 1970)) in his mathematical games column in Scientific American. A 'game' of a similar sort was considered by (von Neumann 1966).

[2] Death: too few (0 or 1) or too many (4 to 8) neighbours (loneliness or overcrowding); Survival to next iteration: two or three neighbours (supportive friends); and Birth, a new cell becomes occupied: 3 neighbours (threesome required).

Turing machine could be implemented within the Life system. Because the system is universal, a simulation of another class of universal system can be implemented within the Life system, with implemented programs running within that implemented system. Just as a PC can simulate a Mac, and that simulated Mac can run its own macintosh programs, so too can universal systems be implemented within the Life system. A PC can implement the Life system, which runs an implementation of a universal system, which mimics a Turing machine, which is performing some calculation.

There are two levels to the Life system: the hardware, and the abstract logical Life system itself. The hardware is computationally universal, and implements the Life system, which itself is computationally universal.

Thus, in the abstract world of the two dimensional grid, machines can be built which are computationally universal. The existence of such machines is dependent only on the Life system, and not directly on the implementation of the Life system. Such machines are dependent on the logical structure of the Life system only, though they are indirectly dependent on that which implements this logical structure. This is just as it is with a Mac program that is dependent on the logical Mac, and not a physical Mac, since a PC can simulate a Mac.

In as much as the behaviour of physical processes can be captured computationally, so can the Life system simulate these physical processes to an arbitrary degree of accuracy.

3.2.3 The status of simulated environments

When the Life grid is displayed on a computer screen, it is easy to give it status as an existing thing in its own right. It appears on the screen as an actual concrete grid, with actual concrete elements in the occupied cells of that grid. The grid display is an interpretation of selective aspects of the physical implementation. The grid display is not part of the Life system, which could be implemented and run with it. The addition of display hardware allows for what appears to us as an actual concrete grid. Only if we have additional physical hardware to create a display of the grid does it seem ‘concrete’ to us.

There are two levels to the Life system: the hardware that implements it, and the logical Life system itself. Computational universality allows for clean distinctions between levels. Within the Life system there could be an implementation of a universal system. Thus, there would be three levels: the physical implementation of the Life system, the Life system itself, and the universal system within the Life system. The universal system within the Life system could itself implement a Life system.

Consider another system, one that is built purely as a realistic virtual environment system. Physical processes are simulated within this system in such a way that when we immerse ourselves in this virtual environment we are suitably pleased with its accuracy. The way in which we are amazed with its accuracy, however, is by rendering an image of the simulated environment.

It is the image, and not the logical environment, which we think of as ‘concrete’. If the environment was never rendered, or only rendered in black and white, would we think it was as real? In what sense, if never rendered, could the mimicking of the physical process of reflection of blue light be thought of as real? Virtual environments seem real enough to us if they are displayed to us and we can interact with them (even if this interaction is limited to merely looking at a rendering on a screen).

The redness of the virtual environment we consider ‘real’ if we see it in the form of a suitably accurate colour rendition. Without rendering, the most complex virtual environment system we know seems somehow unreal to us. As far as virtual environments go, it is not the logical structure of the environment that we care about, but our ability to interact with it.

Virtual environments, however, are distinct from the renderings we make of them, just as the existence of the Life system is independent of our choice to create a display of an updating grid on a screen.

3.2.4 Simulated persons

The assumption in this chapter is that functionalism is sufficient for capturing qualia. The assumption is that the cases of inverted spectra do not apply. The assumption is that there is no need for further physicalist constraints to determine qualia.

This being assumed, persons can be created out of systems that support the relevant functional organisation, with the additional criteria of suitable sensory motor capabilities.

If there is a suitably complex virtual environment with which we can interact, there opens a possibility. A persons mental states, including phenomenal properties of those states, is dependent only on their functional organisation. In us, it is the organisation of neural structures. In robots, it may be the organisation of silicon. A robot could interact with the virtual environment in the way that we do.

The virtual environment system (the system which implements the virtual environment) is a computationally universal system. As such, it can support, in principle, any computational functional organisation, subject to storage limitations. Thus, that system could support the functional organisation necessary for simulated persons with genuine mental states. If functionalism holds, then this is possible in principle. To argue that it is not possible is to argue that something other than functional organisation is necessary to determine genuine mental states.

The virtual environment system is one that simulates the behaviours of physical processes. Thus, the virtual environment system can simulate the behaviours of neurons and of collections of neurons. The virtual environment system can simulate the physical processes of a complete person, and this simulation of the physical processes of a complete person supports the functional organisation necessary for genuine mental states. To claim that this is not so is to add a further physicalist constraint, which I am assuming is not necessary.

There are three levels to the virtual environment system. First, there is the universal computational hardware that implements it. Second, there are the simulated physical processes that are part of the virtual environment. Some of the physical processes that are simulated are in the form of complete persons. Third, these simulated persons support the necessary functional organisation for genuine mental states.

Rather than conceive of a robot which supports the necessary functional organisation, and which interacts via sensory motor equipment with ourselves, we conceive of a virtual world, within which are virtual persons which interact with this virtual world.

3.2.5 The status of simulated environments to simulated persons

The 'reality' of the simulated world to the simulated persons is not dependent upon our having a rendering of that simulated world, or our being able to interact with that world. It need never be rendered in any form, yet it ought to be 'real' to the simulated persons within it. Functionalism says that our mental states are determined by functional organisation, and the virtual world supports the relevant functional organisation for those people within it to have mental states.

The ability to render the simulated world does not mean that we can claim that the simulated persons within that world will experience sounds and colours in the manner that we do when interacting with the rendered simulated world. The rendering of the simulated world is irrelevant to the claim that simulated persons experience their world in a manner similar to how we experience a rendered view of their world.

There is no inference from the simulated world as rendered by us, to how the simulated world would appear to the persons within that world. The redness of the renderings of the simulated world that we see are not part of the simulated world: these renderings are part of our world. The physical characteristics of the implementation and the rendering equipment we use do not play a role in the virtual world. Inferences cannot be made from these.

Our experience of 'redness' of the rendered simulated world therefore, has no bearing on, and no part of the ontological claims we can make on behalf of the simulated persons in that world. The simulated persons in that world are not able to see our renderings of that world. Our only claims regarding the simulated world are those that are implications of the functionalist view that allowed us to create that world, complete with simulated persons, in the first place.

3.2.6 An equivalence assumption

The reality of simulated environments is bound inexorably with reality as we know it. Reality as we know it is how it appears to us. The apparent realness of the simulated world, as experienced by simulated persons, is determined solely by the functional organisation of the implementation of that world. Because these persons are part of the simulated world, the appearance of the world, to them, is dependent on the functional organisation of the simulated world.

The simulated persons are trapped within their own world. They are trapped within the appearances of their own world as we are we trapped in our own world. We can only assume it has the same status as our world, whatever the ontological status of our world is. We assumed a functionalist view, which allows for simulated worlds, and so we consider their world is to them to be as ours is to us. All we need to know to construct their world are matters of functional organisation, so in the explanation of their world, ontological phenomenal properties are irrelevant.

Because of the appearance of our world to us, we allow that their world has equivalent status to them. This is not explicitly ontological. We rely on equivalence. If we have phenomenal experiences, then they have. If we make particular claims about what phenomenal experiences are, then their claims of that nature are as valid as our claims.

We can form a hypothesis about the state of simulated world as it is to the simulated persons. The apparent situation of the simulated persons is equivalent to our situation. We accept that their epistemic situation is the same as ours.

3.2.7 Claims of simulated persons

There is a simulated world and there are people in it. There is our world, and we are in it. It is a functionalist view that allowed us to conceive of this situation. Our mental states are fixed by functional organisation, and this includes phenomenal properties. However, we have not stated what ‘phenomenal properties’ are in a manner aside from being determined by functional organisation.

The simulated persons, being functionally equivalent to us, share the same general mental states, including phenomenal states. Thus, they see the simulated world as it appears to them. Moreover, we do not deny the status of this appearing of the world to them. It happens, however, that there is no specific ontological story is told.

Functionalism rests on mathematics, and that has an unspoken collection of epistemic postulates supporting it, but the epistemological grounds upon which it rests are not stable, and are a matter of debate. Therefore, there is no fixed manner in which claims by simulated persons (or indeed us) can be epistemically evaluated within the functionalist framework. This would require an account of mental content and epistemic justification in the context of functionalism, and there is no such (agreed) account.

Thus, there cannot be epistemically justified stable ontological commitments. This is not a problem of functionalism, as functionalism can ignore ontology, while dealing with the metaphysical question of the identity criteria for mental states. Many of the difficult epistemological and ontological questions regarding virtual persons and their virtual world, and by extension us, remain unanswered. However, we have assumed that functionalism determines mental states, and that includes states we call 'phenomenal states'.

The functionalist view upon which this possibility of the simulated world is based says nothing of phenomenal experience directly. However, we can simply state that, since our brains embody the necessary conditions for phenomenal experience (functional organisation, without further constraint) a simulated persons brain within the simulated world must similarly enjoy phenomenal experience of, for example, 'redness'. We need not have concern for whata that may ontologically entail.

An explicit tackling of phenomenal experience is not necessary. We are making an assumption, and with that, how phenomenal experience seems to us is how phenomenal experience seems to the simulated persons within the simulated world. There is no mention of what phenomenal experience is. Whatever it is, if it is, the simulated persons must enjoy it also, by assumption. Thus, if we feel we have genuine experiential consciousness, we are justified in that claim in as much as the virtual persons would be justified in that claim.

3.2.8 The situation regarding simulation

The situations of the simulated persons and us are equivalent, because we consider the simulated persons as 'real' as we are. In addition, we consider that they consider their world as 'real' as we consider our world. Thus, what they can justifiably claim, we can justifiably claim, and vice versa.

An argument can use this equivalence. The claims that the simulated persons make are justified to the extent those similar claims we can make are justified. Secure and grounded epistemological evaluation of claims is precluded, but relative epistemic evaluation is not.

The following argument is formed around the various claims that can be made. There are different claims that can be made regarding the situation of the simulated world. Firstly, there are the claims that we can make about the simulated world. Secondly, there are the claims that the simulated persons can make about the simulated world. Thirdly, there are the claims that we can make about our own world. When comparing similar claims by the virtual persons and us, we use the criteria of equivalence.

The argument will hinge on a particular case in which we know that certain claims the simulated persons make are, from our point of view, incorrect. However, these claims can be shown to be justified claims that the simulated persons can make. Since our situations are equivalent, if we think that such a claim is justified in their case, then the similar claim we can make must be justified. If we do not consider that similar claim to be a claim we would like to make, then we must deem it unjustified for the virtual persons to make that claim.

The tensions in these possibilities lead to unavoidable contradictions. Thus, functional organisation under-specifies with respect to phenomenal experience. A physicalist functionalism is required. The contradiction in the reductio argument points to the necessary criteria a physicalist functionalism must have.

3.3 Claims about worlds and simulated worlds

3.3.1 Problematic claims

There are claims that can be made by simulated persons within a simulated world that we must deem unjustified. This is not a question of the truth of the claim but of the justification of the claim, and the justification of the claim is dependent on the claimant.

There are claims that can be made by simulated persons that cannot be the result of their sound judgement. *That these claims can be deemed unjustified is independent of the functional organisation of the virtual person.* If a simulated person makes such a claim, that claim is unjustified. The simulated person may feel justified in making such claims, but we deem such claims unjustified.

The claim is one we deem unjustified not because we deem it incorrect. We deem it unjustified because our perspective on the simulated persons and the simulated world shows us that these are not justified claims that simulated persons can make.

It is in the context of what we know about the capacities of the virtual persons that we can deem such claims unjustified. It is not that such claims are 'true' or 'false', nor is it that we consider these claims correct or incorrect. The claim could only be 'true' as far as we believe that we would be justified in making it.

Certain claims by virtual persons are unjustified. But we do not know enough, given the lack of a formulation of epistemology within the functionalist context, to say whether or not a claim a virtual person makes is justified. The best we can do is to invoke the equivalence of our apparent situations. If we made the claim that they made, and we deem it justified, then by equivalence, we could say it is justified for them.

This equivalence is already showing signs of strain. We can justifiably claim something that they cannot justifiably claim. We know about their unjustified claims. This is something that will show difficulties in the assumption that simulated worlds, and simulated persons, are possible. Our judgement that some of their claims are unjustified is a claim which itself must be justified.

We are justified in that we assume a functionalist view, and that entails that there are claims that simulated persons can make that we are justified in judging unjustified. The justification of our claims of their unjustified claims rests on our assumption of the possibility of the simulated world.

There are unjustified claims from simulated persons that are unjustified purely in virtue of the claim. Of course, the claim is dependent on, as it arises from, the functional organisation of the simulated person and the simulated world.

3.3.2 Ontological claims

We are assuming that the simulated world is as ‘real’ to the simulated persons as our world is to us. We assume the simulated world and our world have equal status to the people in these respective worlds. There are no definite opinions on the ontological status of our world, however, so referring to the worlds as ‘real’ is not explicitly ontological. It can be said that our world appears to us as their world appears to them. This statement does not make particular ontological commitments. The only manner in which ontological claims of simulated persons can be evaluated is by equivalence.

Considering the ontology of the simulated world means considering it in terms of the ontology of that world according to the simulated persons. The primary apparent ontological fact of the simulated persons is that of appearances. These may be categorised as secondary properties described in terms of the senses or in terms of the immediate impact on phenomenal experience generally. The ontological status of secondary properties and phenomenal experience is not the issue. Regardless of the status of such supposed ontologies, there is a world as it appears to us, and—as we are assuming computational completeness—a world as it appears to the simulated persons. That there is an apparent world is not questioned, even if its ontological status is. To say that there is a world as it appears makes no specific ontological commitments.

Ontological claims of the simulated persons are either justified or unjustified. For cases in which we do not know how to evaluate the claim, equivalence can be used: if we claim it, we acknowledge their similar claim. There is the possibility that a claim by a virtual person is one that we deem they are justified in making. We can examine claims that they make, and consider whether we would be justified in making such a claim about our situation. If we would be justified in such a claim, then by equivalence, so are they.

However, a claim we feel they are justified in making—a claim that we would feel justified in making about our own situation—may also be one that we deem incorrect, if made by the virtual persons. The incorrectness of the claim in this instance stems from the fact that we would not feel justified in making that claim about the simulated world. Thus, there are claims we feel the simulated persons are justified in making about their situation, but these claims are not ones that we would feel justified in making about their world.

The instance in which a simulated person’s claim is justified, but incorrect, is one in which we are using our viewpoint to make a claim about the simulated world. Our judgement of such claims as incorrect, however, does not mean that this claim by a simulated person is unjustified.

We have a unique perspective on the virtual world, but this does not mean we have privilege to overturn their claims about their own world. Claims which simulated persons make regarding the simulated world may be justified, even if they are not claims we would feel justified in making about their world.

It is the justification of their claim, in the context of what they can justifiably claim, and not our judgement of its correctness in terms of whether it accurately reflects what their world is to us, that is the issue.

3.3.3 Our claims about simulated environments

From our point of view, there is more to the simulated world than at first the simulated persons may realise; there is more to the simulated world, from our point of view, than how it appears to the simulated persons. Our epistemic boundaries are equivalent, but the contexts are different. The situations are equivalent as regards our claims about our world and the virtual person's claims about their world, but since we built their world we have a perspective on their world which they lack. We can justifiably make claims about the virtual world that the virtual persons cannot.

These claims concern the implementation of the simulated world. This may seem to indicate that our epistemic boundaries are greater than theirs. This is not so, as the equivalence of our situations ensures that our epistemic boundaries are balanced. It is the case, however, that the virtual persons can make claims about their world that we cannot make about their world.

3.3.4 Ontological views within simulated environments

The simulated persons may ponder what their world is, outside the context of how it appears to them. They may strive for a more objective view of the world. Everything they can say about their world will reflect the world as it appears, even if indirectly via their empirical instruments. They may argue that this is 'true' objectivity, or 'good enough' objectivity. Alternatively, they may consider the world independently of how it appears to them; they may consider the notion of noumena.

Noumena, for us, refer beyond appearances and derived ontologies to the world as it is in itself. It may or may not be a coherent notion. Similarly, for the simulated persons, noumena refer beyond the world, as it appears to them. From our point of view, their world, beyond its appearances to them, is a large chunk of implementation hardware.

We see this as the simulated persons attempt to ponder their world beyond how it appears to them. To us, this can only refer to the implementation. However, we know that the virtual person's concept of noumena cannot be of the implementation level as they are cognitively closed to it. Their concept of noumena does not refer to the implementation; their intentional mental states regarding the implementation are not 'about' the implementation.

3.3.5 Epistemic limits for simulated persons

Nothing of the specifics of the implementation of the virtual world is important, either to ourselves, or the virtual world and its inhabitants. What is important is that it is a universal computational system, and this is a fact which is independent specifically (but of course dependent generally) on the fact of the matter regarding the existence of the system. It seems that it must matter, in that an implementation must exist, but exist it does solely to ground functional organisation. We can say that the implementation matters only in that it exists, and that it is flexible enough to be a universal system. The functional organisation matters. If anything beyond its functional organisation mattered, functionalism would be making crucial ontological claims, but the functionalism under consideration here makes no such claims.

There are limits to the epistemic situation of virtual persons. Their world is not dependent on anything but the functional organisation of the implementation. They have no epistemic access to the details of the implementation. They are precluded from 'knowledge' of the details of the implementation.

That the virtual persons are cognitively closed to the implementation does not preclude them from making noumena claims. If a virtual person makes a claim about that upon which their world rests, that claim is unjustified. It is unjustified in the context of what they are, and what they can know. The claim may be one that we feel we would be justified in making. However, we have no alternative but to judge their claim to be unjustified. In judging so, we have made an explicit epistemological claim.

Virtual persons are cognitively closed to the implementation level, but it is necessary to know precisely what this means. Its meaning can be expressed in terms of the 'reality' of the virtual world to the virtual persons, which is the virtual world as it appears to them. The virtual persons will never have any knowledge of the specifies of the implementation level. Nothing of the specifics of the implementation has any bearing on how the world appears to them. It is the functional structure of the implementation level that contributes to the world as it is to them.

Knowledge of the implementation level may not be derived from appearances; it may be innate. Functional organisation determines the world as it is to them includes both the world as it appears, and the world as it seems, to them. This includes such things as innate knowledge that the virtual persons may claim to have. They may make a distinction between empirical and innate knowledge, but both these are determined solely by the functional organisation of their world. Thus, the simulated persons are cognitively closed to the specifics of the implementation level of their world. Their world is one of appearances and derived ontology, but not transcendent ontology

Virtual persons could be programmed to make accurate claims regarding the nature of the implementation level, but this knowledge is still unjustified, even if correct from our point of view.

3.4 Worlds in worlds

The simulated persons may eventually come to believe that their mental states are specifiable in terms of function. To put their lofty ideas into practice, the virtual persons go about building a simulated world, complete with simulated persons. These simulated persons, they say, have an apparent situation equivalent in all necessary respects to their own.

The virtual persons have adopted functionalism even to the point of creating simulated persons in simulated worlds. They have therefore discovered their actual situation, as we see it. The simulated persons are making a claim about themselves and their world that we feel justified in claiming about their world. Are their claims about their world justified? Are they justified in accepting functionalism?

If we accept functionalism for ourselves, we cannot deem their acceptance of it unjustified. By assumption, functionalism allows for simulated worlds with an equivalent epistemic situation to our world. The simulated persons claims of functionalism are correct, from our point of view. Thus, if we deny their claims, it reflects on our own claims of functionalism. By equivalence in any case, we must take it that their claims of functionalism are justified.

3.4.1 An instance of overstepping cognitive closure

We know that the simulated persons are cognitively closed to the specifics of the implementation of their world. The virtual persons accept this also. Their acceptance of functionalism and created simulated worlds entails that they have no access to an implementation level of their world, if they considered their world in that way.

Any claim, which is a valid inference from the functionalist claim, is a justified claim, if the functionalist claim is justified. That their world is dependent only on functional organisation is an acceptable view for them to hold. This view is correct, from our point of view. It is a valid inference from their acceptance of functionalism.

However, their acceptance of the functionalist view implicitly refers to the concept of ‘functional organisation’ upon which their world rests. Though they are cognitively closed to the implementation level, the acceptance of functionalism itself implicitly refers to an implementation level. This ‘functional organisation’ does not refer to the organisation of the appearances of their world, and it cannot refer to the world in any way as it is to them. It cannot refer to anything of which they have knowledge. It therefore implicitly refers to something outwith their epistemic boundaries.

There is a difficulty with referring to something that is admitted as beyond epistemic bounds. Simulated persons are precluded from justified claims about specifics of the implementation of their world, but they are not necessarily precluded from justified claims that there is an implementation, of which, by their own admission, they know nothing.

This is akin to our concept of noumena. We may accept a noumenon as an unknowable thing, but yet we refer to it, have concepts for it, and have a name for this unknowable thing. Being an unknowable thing does not mean that its existence is also unknowable. If the existence of an unknowable thing is unknowable, then even pondering the concept of such a thing is unjustified. Nevertheless, in our case, we may accept that we can refer to the existence of an unknowable thing: we know of it, but not about it. However, there is a difference in the claim of noumena being justified, and the implementation level claim of the simulated persons.

The simulated persons concept of noumena is independent of their acceptance of the functionalist hypothesis. The concept of the implementation level, of which they know nothing, cannot be described as a noumenon concept. A noumenon is fully unknowable, but the implementation level is not.

If the simulated persons considered the implementation level at all, they consider it as within the class of things that can implement the class of Turing computable functions. The implementation must be computational universal, and this fact is implicit in the concept of an implementation level.

The notion of an implementation level gives the world of the simulated persons ontological status beyond the ontological status they may give to the world as it appears to them. They rule out specific claims about an implementation level as being epistemically unsound. There can be no noumena or other deeper ontological commitments made by the simulated persons.

Are the simulated persons justified in considering their world a simulated world? Are they justified in saying that the world as it appears to them is dependent only on functional organisation? The answer is, it depends on what ‘functional organisation’ is taken to mean, it depends on whether this implies there is something which supports this functional organisation.

3.4.2 Implications of cognitive closure

Consider the simulated persons options with regard to the acceptance of the notion of ‘functional organisation’ upon which their world depends. That which has this ‘functional organisation’ does not refer to anything in the simulated world that the simulated persons can know.

The options are either to accept that ‘functional organisation’ refers to something which has this functional organisation, or treat it as having no such referent. One of these options must be valid, if functionalism is a justified concept that simulated persons within a simulated world can hold.

Accepting that ‘functional organisation’ refers to something amounts to accepting an inference from the acceptance of functionalism to the necessity of a substrate that supports this functional organisation. It views ‘functional organisation’ as requiring a ‘that’ which ‘has’ functional organisation.

This referent is the implementation substrate. The details of the implementation are unknowable to the simulated persons, yet they know it exists and that its important actions are encapsulated within the computationalist hypothesis.

The simulated persons have knowledge of what ‘functional organisation’ can refer to when applied to objects within their world. They know that the actions of the implementation substrate can be simulated in their world, as their world has computationally universal objects.

The implementation substrate, which has functional organisation, does not matter beyond the functional organisation that it can support. Its ability to implement these actions is not unique or fundamental. The important actions of the implementation substrate can be defined terms of objects within their world. Their understanding of the implementation substrate does not rely on referring to something outwith their epistemic bounds.

3.4.3 Ontological privilege

Because the important features of the implementation level can be defined within the world as it appears to the simulated persons, there is no sense in which they can claim that the implementation level is fundamental or privileged.

The important actions of the implementation substrate could be simulated within their world. The implementation substrate is not privileged, as any computationally universal system could support the relevant functional organisation.

The simulated persons refer to an implementation level which is outwith their epistemic bounds, and then find that there is nothing privileged about it. The implementation level itself—its characteristics that allow for functional organisation—could be simulated in another computationally universal substrate.

The implementation level of their world could have its own implementation level. The simulated persons have no reason to discount this possibility. Thus, they have no reason to claim that the implementation level to their world is privileged.

The difficulty is that the simulated persons are referring to an implementation level which they are cognitively closed to. They cannot treat it as a noumenon, because they do know something about it: its important feature is the ability to support functional organisation, and this is a feature of any computationally universal system. Their implementation level may be a simulated system within some other computationally universal system.

This leads to an infinite regress, as they must consider that the implementation level may itself be implemented in another context. In order to avoid the difficulty of an infinite regress, they could abandon the notion of an implementation level to their world. But for the time being, they are taking it that their world is dependent only on functional organisation, and so this approach is not an option.

The simulated persons could declare that the implementation level of their world is privileged or fundamental: that it supports the relevant functional organisation, but has ontological aspects that precludes it from being simulated: it has aspects which cannot be encapsulated by functional description.

In order to claim this, the simulated persons need to make a strong ontological claim regarding the nature of the implementation substrate: that it is an ontological ground that supports the functional organisation relevant for their world. However, all that they can know is dependent only on the functional organisation of the substrate, and not any ontological fact about the substrate. Thus, this claim cannot be the result of innate or derived knowledge of any kind. In invoking the privilege of the implementation level they are invoking a supposed aspect of that implementation level which has no bearing on their world or themselves.

From our point of view, the claims the simulated persons make regarding functionalism and their world are correct.

3.4.4 Ignoring an unanswerable ontological question

The simulated persons may realise the problems arising from considering an implementation substrate to their world. They may then consider that in even pondering an implementation level, they are going beyond the bounds of closure set up by their functionalist views.

This is not a resolvable situation. An entailment from a view that they accept breaks the epistemic boundaries that the view creates. However, there is one possibility, and it is to take it that the functional organisation upon which their world rests does not have an external referent. The simulated persons deny any pondering over ‘that’ which has this functional organisation, yet they keep the notion of functional organisation as being the determining factor of their world.

The simulated persons attempt to ignore the question of an implementation level by not acknowledging it as a valid concept. Nevertheless, they still have the notion of ‘functional organisation’ defining their world. However, they do not consider the external ‘that’ to which this would refer, as this is outside their epistemic bounds. They hope in this way to avoid epistemically unsound transcendent referents.

The concept of their world relying on ‘functional organisation’ now becomes nebulous. They do not acknowledge anything to which this functional organisation could refer. Functional organisation, taken alone, is an abstract mapping, a Platonic concept. They can consider the notion from within the epistemic confines of the virtual world only.

From our point of view, this is meaningless. The virtual persons cannot have strongly ontological views regarding abstract entities. They are dependent solely on the functional organisation of a device that is not abstract. There is no justification for a claim of the purely abstract in the situation of the virtual persons, from our point of view.

The virtual persons, attempting to resolve the dilemma, consider that their world is defined functionally, but without further ontological commitments. It is a functional mapping, and nothing more is said. There is no external referent implicit in ‘functional organisation’, so it is not functional organisation of anything.

Nothing privileges the functional organisation of the simulated persons world. They have no ontological grounding to this ‘functional organisation’. If they claim that their world is dependent on functional organisation, but are not allowed an ontological grounding to this, then all functional organisations which could support worlds have equal status to their own world.

All that can be said about their world is that their world is that which is this functional organisation. Functional organisation is not functional organisation ‘of’ something, it merely is functional organisation. Their world is not privileged. To say that their world is the only world is to imply that there is a further ontological commitment to that which supports the functional organisation of their world.

The number of other worlds with equal status to their own is the number of functional organisations that can support worlds. The many worlds are the many logical possibilities of functional organisations. It is thus modal realism of a sort.

The statement of modal realism in this context is one in which there are several independent and non-interacting worlds of appearances, akin to the world of the simulated persons. They are non-interacting and independent in an epistemic sense. They are independent of the world as it appears to them. (This is similar to the strong modal realism of David Lewis, who argues for independent and separate worlds with equal status to our own world (Lewis 1986)). The ontological status of these worlds is described in terms of appearances alone, as the simulated persons have difficulty with transcendent ontologies. We too have this difficulty, as the way in which we describe the reality of the virtual world is either in terms of how it appears to us or in terms of how it appears to the simulated persons.

The simulated persons are forced to accept that there are other worlds which appear to the persons within them—for those worlds that have ‘persons’ to which a world ‘appears’—

as their world does to them. They must give some worlds the same status as their world, in the same manner in which we give the simulated world the same status to the persons within it as our world has to us.

From our point of view, the ignoring of the implementation substrate difficulty by the simulated persons is incorrect. In addition, it is also unjustified from the point of view of the simulated persons. As they ruled out an infinite regress because of a lack of epistemic justification, so too they must they rule out the alternative of a sort of modal realism.

Infinite regress was ruled out as it oversteps the bounds of epistemic closure. The statement of many worlds is one in which other worlds are given equal status to the virtual world, and that oversteps their epistemic bounds. If they acknowledge an implementation level, they are overstepping their bounds. They cannot ignore it completely, as it is part of the functionalist hypothesis they accept. Their attempt to ignore it voids their world of privileged status. Thus, this leads to notions of many worlds, and that implicitly oversteps epistemic bounds. To avoid this means returning to the requirement of privileging their world in some way, and the only way to do that would entail overstepping epistemic bounds.

From our point of view, their world is necessarily dependent on the concept of an implementation. There is nothing privileged in this implementation. Their acceptance of an implementation level is correct, from our point of view, but it is unjustified. It would lead them to an infinite regress, unless they make a claim that from our point of view is both unjustified and incorrect. Their ignoring of the implementation level is incorrect. Ignoring it completely is an attempt to stay within epistemic bounds, but within this claim, there is the implied claim of many worlds, and so the claim is not justified.

3.4.5 Our supposed world

We gave the apparent situation of the simulated persons the same status as our own apparent situation. We postulated that our epistemic situation, and that of the virtual persons, is equivalent. Their epistemic concerns are our epistemic concerns. Their concerns about their world are our concerns about our world. We know that their situation is an implemented world, and so we must accept the possibility that our situation is an implemented world. Our world could be a simulation on some other hardware that we know might look to the builders of that hardware as computer hardware looks to us.

We could simply accept that implementation levels are beyond our epistemic boundaries. Of what we cannot know, we could remain silent. As far as the existence of the world is concerned, all that matters to its existence is the abstract functional organisation that could support it. Without further elucidation on that point, what is it that makes one abstract functional organisation privileged? That question cannot be answered, and remaining silent on this point belies the sense that there is no privilege, for privilege would need to be an extra postulate which, in itself, would be outside our epistemic justification. With the world considered as abstract functional organisation, all abstract functional organisations have the same status. All worlds exist, and have equal ontological status to our world, if our world has any status at all. The modal realism of the simulated persons has caught up with us. This is expected, because of equivalence. As the simulated persons have no valid routes through the epistemic difficulties of this issue, neither do we.

We could try to ignore these difficulties. We can ignore all pondering over implementation contexts, and ignore all wondering about extreme modal realism. We could deny that the ungrounded notion of functional organisation leads to many worlds. Thus we make no commitments, and maintain a stance of silence. All we have left is the world as it appears to us. Any further ontological claim we cannot make, no transcendent commitments, even implicitly, are possible. We have become idealists. Our idealism is more idealist than the idealism of Kant or Berkeley; they had notions of justified transcendent inferences which are precluded in this case.

The radical functionalist view is too effective. Functional role, alone, is not enough to fix phenomenal experience, neither are internal architecture constraints enough: an ontological constraint is required, such as a physicalist functionalism can provide. We considered functionalism independent of specific ontological commitments, and in accepting it, end up independent of ontological commitments. Yet, we still have the world as it appears. We still have this because being independent of ontological commitments is not being eliminativist of ontological commitments. This functionalist view merely ignored, but did not eliminate them. Functionalism was successful at eliminating explicit ontologies of the simulated persons, but it left the minimal ontological residue: the appearing of the world.

Some functionalists are eliminativist. However, there is only one thing left to eliminate, and that is the appearing of the world. Eliminating this results in nothing at all. What is left when there is nothing at all left? Presumably, the acceptance of functionalism is left. This states that functional organisation is all the matters. So perhaps eliminativists have an abstract ontological commitment of some sort.

However, our world—the appearing of the world—is not abstract. It is composed of instances, of particulars. There is more than abstract functional organisation in our world. To say that these appearances are dependent upon an abstract, ungrounded concept of 'functional organisation' is meaningless.

Functionalist eliminativism is the result of not realising how good functionalism is at eliminating ontologies. It leads to idealist ontology, and if the idealist ontology is eliminated, we are left with nothing, except perhaps abstract functional organisation in a Platonic sense.

There is nothing but the seeming of phenomenal experiences, for there are no deeper ontologies. Yet, the sense in which elimination is invoked is to get rid of old fashioned and embarrassing subjective ontologies. These are thrown away in favour of more transcendent, or fundamental, or privileged, or 'objective' ontologies (more down to earth, more 'physical', or 'material'). Yet, if functionalist eliminativists eliminate phenomenal ontologies, they eliminate everything.

3.5 Conclusion

These contradictions are explicitly embraced by Tipler and Toffoli (see for instance (Tipler 1989) and (Toffoli 1982)), who combines the concept of abstract functional organisation with phenomenology and extreme modal realism. To them, the fact that the world is a "huge ongoing computation" (Toffoli 1982, 165) is the essence of their view. Because of the epistemic limitations that follow from this view, no talk or reference to what implements, realises or instantiates this 'ongoing computation' is admitted. There can be no ontological commitments either. As Barrow says, in the context of this extreme functionalist view, "such a physically real universe would be equivalent to a Kantian thing in itself. As empiricists, we are forced to dispense with such an inherently unknowable object". The unknowable object is that which supports this 'ongoing computation" (Barrow and Tipler 1996, 155), which Barrow realises is something which is incompatible with the epistemic boundaries of the views of Barrow, Tipler, and the radical functionalist view. Barrow considers our situation to be exactly the situation of simulated persons in the simulated world, who preclude themselves from any talk of an implementation. Radical functionalism has these implications, but Tipler, Barrow, and Toffoli accept them explicitly.

In the case of simulated worlds, there is nothing more than things as they appear to a simulated person: no further ontological commitments can be made. Thus, the person ought to be an idealist. However, to do so would be to give ontological status to the appearing of the world. Thus the appearing of the world is not 'nothing but' what is ultimately functional and behavioural.

What can be concluded is that secondary properties, phenomenal experience, or qualia cannot be eliminated. They cannot be 'nothing but' functional states. There is a requirement for a specific, explicit ontological commitment. Physicalist functionalism is one view that attempts to do this by saying that functional states are physically grounded. The term 'physical state' involves ontological commitments.

The concept of simulated persons within simulated worlds causes difficulties, and these difficulties provide reason enough to take it that simulated persons within simulated worlds are not possible. This extends also to simulated persons within our own world: robots, for instance. The epistemic situation of the robot is also dependent only on its functional organisation. How the world 'appears' to the robot is dependent only on is functional organisation. Robots in our world cannot tell that our world is not a simulated world. There may be non-biological or artificial persons, but their epistemic situation will not be dependent purely on the functional organisation of their internals.

Functional organisation is not enough. Our apparent situation—how the world appears to us—is dependent on more than functional organisation. Specific ontological commitments are required. Physicalist functionalism must be chosen over pure functionalism, if functionalism is to be kept.

Chapter 4

Phenomenal

4.1 Introduction

The question of phenomenal experience is both ontological and epistemological. The distinction between the ontological and epistemological aspects of phenomenal experience may not be 'clean'. If phenomenal experience is ontological, it is reasonable to argue that the justification for the phenomenal realist position is the phenomenal experiences themselves. It is reasonable to suggest that phenomenal experience would be its own justification, if phenomenal realism (in this sense) were correct. But there is still need of an epistemic premise. The epistemic premise, in phenomenal realist cases, is that there is core-epistemology, a type of direct knowledge of experiences, and this reveals them to be ontological.

4.1.1 An epistemic premise

Stating that phenomenal experience is an ontological kind is valid if it is an ontological kind, but is it? The premise is that there are phenomenal ontological kinds, and this is known because they exist, and that is known because of their epistemic aspect. The epistemic aspect, the direct knowledge of experiences, must be secure; we must be able to say, justifiably, that there are phenomenal kinds because we experience them

The premise is essentially one that states, "I have phenomenal experiences and about this I cannot be mistaken". In any phenomenal realist premise there is a degree of first person authority. It need not be as strong as the certainty of Descartes, but it needs to be strong enough to override other accounts (empirical or functional, for instance) that are used to suggest eliminativism. The phenomenal realist premise is based, to a large degree, on its argued immunity to the skeptical premise.

However, the immunity to skeptical arguments is inherent in the sense of, and meaning of, 'phenomenal experience', and this calls into question the reliance on this immunity. Skeptical arguments rely on something seeming to be a certain way, when in actuality being quite different. This description of skeptical argument, however, rules it out as being applicable to the case of phenomenal experience. The 'seeming' is the *fact* of phenomenal experience. Thus, 'seeming' to have 'phenomenal experience' when one is not, is impossible in virtue of what we understand as seeming to have 'phenomenal experience'.

Skeptical argument, then, is applicable in cases of brains in vats, false memories, being trapped in virtual environments, and, according to Descartes, to the existence of our own physical bodies. However, it is not applicable in the same way to many concepts of phenomenal experience, as there is too close a connection between the meanings of 'seeming to be one way' and 'phenomenal experience'. So we return to the issue of immunity to skeptical argument.

A link

The phenomenal realist claim can be considered as having two aspects. There is the fact of phenomenal ontological kinds; there is the fact of this being claimed; and there is the direct knowledge we supposedly have of phenomenal experiences. It can be considered an epistemic claim about ontology. There are, however, implications arising from making a distinction between ontology and epistemology in this instance.

If a distinction is made, and the phenomenal realist premise is to be justified, then there must not be a contingency between the epistemic knowledge of phenomenal experience and the ontological fact of phenomenal experience. For those that argue immunity to skepticism, this must be so; the ontology of phenomenal experience and the epistemology of phenomenal experience go hand in hand.

If the ontological and epistemological aspects are considered logically contingent then logically possible cases in which someone has the epistemology of phenomenal experience without the ontology arise. An example would be ersatz pain. The question then arises as to the epistemic difference between real and ersatz pain. The answer is, there is no difference, since what we know in both situations is the same. Thus, the ontological aspect makes no important difference, and so the epistemic aspect of phenomenal experience *is* phenomenal experience. This may lead to plausible eliminativist accounts, where it is not denied that people epistemically believe there are ontological phenomenal kinds. Shoemaker rules against ersatz pain on the basis of considering qualia as functional, and therefore introspectively accessible (Shoemaker 1975). This is Shoemakers account of the necessity for having the ontological and epistemological aspects linked. If qualia are not functional, they are not introspectable, Shoemaker claims. And if absent qualia are possible, then in what manner are qualia introspectable? (Davis 1982). Thus, absent qualia are possible, and furthermore, there is no difference between real and ersatz pain. There is room for eliminativism of the ontological view of qualia. Averill, in this context, opts for eliminativism (Averill 1990). These arguments tend to conflate two issues: the direct knowledge of phenomenal experiences, and issues of introspection. There are two senses of knowledge. Shoemaker combined the senses of direct knowledge, and introspection, in order to avoid absent qualia.

To ensure a difference between real and ersatz pain requires that what we know of our experiences reveals something of the experiences themselves. Hence, a need for some direct knowledge of phenomenal experiences. The epistemic and ontological aspects of phenomenal experience then, would not be distinct, or would not have a contingent connection between them. For phenomenal realism, then, there is either no clean distinction between the epistemic and ontological aspects, and if a clean distinction is made, the connection between them holds necessarily.

This rules out certain accounts of the epistemology of phenomenal experience. It rules out all accounts that have an unreliable link between the ontological and epistemological aspect of phenomenal experience. Thus, this epistemic knowledge of phenomenal experience is not open to functionalist or causal accounts as with causal chains there may be over determination along the line, and that which supports a functionalist account may break. A functional or causal account of the connection between the ontology and epistemology of phenomenal experience does not provide a necessary connection. Hence, the need for some direct knowledge of phenomenal experiences.

Since phenomenal realism is a premise—in the sense of not generally being derived through argument but defended as a postulate—there need be no particular difficultly with merely accepting it, and building functionalist, behaviourist, or computationalist views around it. The resulting view, however, must not cast doubts upon the requirement of epistemic certainty. Views built around the phenomenal realist premise do not need to supply an account of epistemic certainty; but they must provide a space into which epistemic certainty is placed. This is to say that, as the epistemic and ontological aspect are linked necessarily, an account built around this premise must not rule against that necessary link.

4.1.2 Inessentialism

The premise of phenomenal realism need not have an impact on explanatory endeavours. Indeed, because of the success of explanatory endeavours, it is often not seen as having an impact. In such cases, the phenomenal ontology and the epistemic knowledge associated with phenomenal realism are not seen to be essential to particular types of explanation. The phenomenal is considered explanatorily irrelevant with respect to an explanatory endeavour.

If the explanatory endeavour is taken to explain the entirety of a particular domain, then in that regard, the phenomenal is inessential to that domain. For instance, consider that behaviour can be explained in a way that does not refer to phenomenal experience.

Thus, phenomenal experience is irrelevant to the explanation of behaviour.

That something is irrelevant and inessential in this manner is not to say that useful and accurate explanations that refer to that thing are precluded. Such explanations can be both useful, and accurate. Explanations, which invoke inessential phenomenal experience, are accurate explanations of the world. Inessentialism merely states that it is possible, in principle, to have a full and accurate explanation of a certain domain (behaviour, or function, for example) without referring to the inessential item.

These cases concern explanatory endeavours that do not have explicit ontological referents. Behaviourism and functionalism do not have specific ontological commitments. Physicalism, however, has explicit ontological commitments. In such cases, the explanatory irrelevance or inessential nature of phenomenal experience may be a stronger claim. The claim that phenomenal experience is inessential to the ontological story of the world, is a very strong claim, leading either to dualism or eliminativism. Alternatively, such a claim may lead back to reconsidering the statement of inessentialism. A further option of a non-inessentialist monism that encompasses phenomenal experience is open.

The phenomenal may be explanatorily irrelevant. That is not to say that it cannot be described as causally efficacious. This is causation of the phenomenal in virtue of that to which it is related via identity, reduction, or supervenience. However, it is explanatorily irrelevant, and so need not be invoked in a causal story of the world. Thus, phenomenal experience, of itself, is causally inefficacious in these cases.

The degree of irrelevance and inessentialism depends on the particular view. It may be that, with regard to behaviour and function, phenomenal experience is irrelevant, but it is not seen to be so with regard to the ontological story of the world.

Views in which phenomenal experience is seen to be inessential to the scientific, or ‘objective’ ontological story of the world, yet essential with regard to the behaviour and functioning of that ‘objective’ world are, at present, quite rare. Cartesian dualism is an instance of this class of view: the mental is in a separate ontological sphere, yet influences the functioning and behaviour of the material ontological sphere.

Where the phenomenal is assigned ontological status, contemporary views lean mostly towards declaring it explanatorily irrelevant and inessential with regard to the behaviour and functioning of the world. There are two advantages to this. Firstly, interaction between distinct ontological kinds is avoided. Secondly, ontological phenomenal experiences are maintained, but are fixed and determined by the physical to which it is related by identity, supervenience, or reduction. This allows ultimate authority to be given to objective empirical explanation, while at the same time allowing authority to be given to first person phenomena. This type of inessentialism is considered in this chapter.

Specifically, the inessentialist view considered here has the following elements. Phenomenal experience is taken to be something in addition to function and behaviour; it is not open to eliminativist accounts, and it is not explained in purely functional or behavioural terms. Phenomenal experience is inessential to the functioning and behaving of the empirically knowable world, and thus is explanatorily irrelevant in these domains. It is not necessary to state explicitly that phenomenal experience is ontological, in this view. That it is not encompassed by empirical third person explanation is enough. However, given this, phenomenal experience must have an ontological aspect: it is something that is not reducible to behaviour, function, and empirical knowledge.

4.1.3 Authority

Inessentialism can maintain the authority and in-principle completeness of explanations of function and behaviour, while accepting the authority of phenomenal experience. The phenomenal realist premise is the acceptance of such authority. Inessentialism ensures that the acceptance of phenomenal realism is not in opposition to the in-principle completeness of explanations of function and behaviour.

First person authority is thus not in opposition to third person authority, as the first person is inessential with respect to the third person. The authority of phenomenal experience does not compromise third person explanation of function and behaviour because of the inessentialist premise. In the phenomenal realist premise is the sense that this premise is justified. There is the sense in which there is epistemological justification for this premise, even if such justification is not given, and the premise merely asserted. There is no justification for the premise to be found with functional and behavioural explanatory endeavours. These endeavours are not dependent on the premise.

With regard to the phenomenal realist premise, then, there are the following basic points of authority. Firstly, our basic epistemic understanding of phenomenal experience is accepted. It is taken to be something irreducible to accounts of behaviour, function, and empirically derived knowledge generally. This may be supported by our intuitions that empirical methods to not account for it. Nevertheless, the concern here is not with supporting the premise, but with accepting it, and seeing what the implications are. Secondly, it is taken to be inessential with respect to behaviour and function generally, at all levels of the empirically knowable world. Thus, there is no conflict between it and functionalism, behaviourism, or empiricism generally.

In addition, there are also the following implicit points of authority. Firstly, the phenomenal realist claim is epistemically justified, though no justification need be given if the premise is merely asserted. Secondly, any other explanatory endeavour must not rule against their being an epistemic justification for the premise. The second point states that explanatory endeavours used alongside the phenomenal realist premise must be consistent with this premise. This only requires that they do not imply that the pheomenal realist premise is false, and do not imply that an epistemic justification for the premise is precluded.

4.1.4 Argument outline

The argument put forth is one that attempts to show the incompatibility of the following two premises. One, phenomenal experience is an existent irreducible to empirical explanation. Two, phenomenal experience is explanatorily irrelevant to empirical explanation. In essence, the argument attempts to show that phenomenal realism and inessentialism form an inconsistent pair.

The options that follow from this are twofold. The phenomenal realist premise can be dropped, or the irrelevance of phenomenal experience to empirical explanation can be dropped. This latter option sounds like interactionist dualism, but this is just an extreme view which may follow, not the only one.

The argument develops in the following way. Firstly, it is pointed out that the phenomenal realist premise requires a strong degree of epistemic certainty. Secondly, it is pointed out that this certainty cannot be contradicted by empirical explanatory endeavours. It is shown that this entails that a particular scenario is impossible. A case, in which this scenario is not just logically possible, but empirically possible (contingent on the correctness of the assumptions), is described. This empirically possible scenario is a case that embodies the fact that empirical explanatory endeavours can override the required degree of epistemic certainty necessary for the phenomenal realist premise. Thus, inessentialist empirical explanation is seen to disallow any justification for the phenomenal realist premise. The argument undermines either the phenomenal realist premise, or the inessentialist premise. Thus, this argument alone can be used in an eliminativist manner.

4.2 A possible scenario

4.2.1 Making a claim

The phenomenal realist premise does not refer to causal, functional, behavioural, or other third person empirical account. The phenomenal realist premise is accepted because we believe that we are justified in its acceptance. From the first person point of view, I choose to accept the phenomenal realist premise for myself, for it seems epistemically evident that it is so. This aspect, the epistemic justification for the premise, will be called core-epistemology: I accept the premise of phenomenal realism because I have core-epistemology of that fact. Core-epistemology is inherently first person, it applies to that person who makes the phenomenal realist claim.

The claim is accepted for all persons. We accept the claim that others make as being similarly justified as our own. We accept that others make the claim for the same reason of core-epistemology. From the first person, I accept the claim for reasons of core-epistemology, but I accept it for others because I accept that they are making the claim for their reason of core-epistemology.

The phenomenal realist claim can be seen from both the first and third person points of view. The claim refers to a supposed fact: epistemic surety of phenomenal experience. This claim can also be seen from the third person point of view, and it can be considered in a third person manner.

Looking at someone making a claim from the third person manner, we can consider it a behavioural act, and we can provide a functionalist account of that act. We need not provide a behavioural or functionalist account of the content of the mental state, 'I have core epistemology of phenomenal realism'. What is important here is that inessentialism allows us an in-principal complete and accurate functional and behavioural explanation for the *act of claiming*. This is not, however, an explanation of *what is claimed*.

Thus, a persons claiming of the phenomenal realist premise can be explained without reference to phenomenal experience. This is so, because inessentialism is accepted. The third person claiming can be explained. The explanation of the claiming is independent of what is claimed. The difference phenomenal experience makes is to the first person, and then the difference is core-epistemic. It makes no difference to the third person.

The important fact here is this: There are phenomenal kinds, and our claims to that effect are epistemically sound, yet our claiming can be explained without reference to phenomenal kinds.

4.2.2 Empty claiming

Since the explanation of claiming is independent of what is claimed, an explanation of claiming cannot pass judgement on what is claimed. In third person terms, the claiming could occur in the absence of what is claimed. The claiming is third person, and the claim first person, but this matters not to the explanation of the claiming. We are thus precluded from making first person judgements from a purely third person point of view.

From the third person point of view, we can see someone making the phenomenal realist claim. We do not know, however, whether the claim is just empty claiming, or whether it is a genuine claim.

That claiming is independent of what is claimed allows for a logical possibility. That claiming is distinct from what is claimed is a feature of inessentialist views, and it is those views that allow for the logical possibility of zombies. Zombies are persons who can make the claim, but the claim is an empty claim: they are just claiming, and nothing more.

Zombies are conceived as having no experiences, no phenomenal ontology, and no phenomenal epistemology. There are no first person 'feels' associated with their existence. Of course, it is impossible to directly conceive of zombies because we cannot imagine what it would be like to be a zombie, as there is nothing it is like to be a zombie. We cannot conceive of what it would be like to be dead (assuming an atheistic view), or imagine what its like to not exist. However, our intuitions as to the distinction between behaviour and experience open a way to conceive of zombies.

Zombies are functionally and behaviourally identical to us. This is the extent of the zombie notion used here. Others may accept that zombies can also be physically identical to us.

In the absence of a demonstration of the contradiction of the zombie notion, we are free to accept it as a logical possibility. Empirical possibility is not the issue. It may be impossible, in this world, to have zombies, yet they are a logical possibility. Arguments against the zombie notion need to show a contradiction in the logical, not empirical, possibility of zombies.

This type of zombie is unrecognisable (and so distinguished from folkloric zombies) to us through empirical third person behavioural and functional analysis. They are identical to us in these ways. Being functionally and behaviourally identical does not rule out empirical detection, however. Zombies may have chalk for brains, while being functionally and behaviourally identical. The argument presented further on does not require zombies to be physically identical, just behaviourally and functionally identical, as the conclusions do not rest on this type of empirical detection. Nevertheless, it can be taken that they are physically identical also.

Just like us, zombies may accept the logical possibility of zombies. They may say that zombies are the same as them, but without consciousness, which, unbeknownst to them, they do not have. Zombies believe in zombies, but zombies do not believe that they are zombies, for the most part. There may be zombie eliminativists, just as there are non-zombie eliminativists.

Zombies show us that, whatever the epistemology of the phenomenalist premise, claims of 'phenomenal experience' are allowed, even in the absence of phenomenal experience. It also points to the other minds problem: how we evaluate the claims of others.

If I were a zombie, I would act no differently, and I may claim the phenomenal realist premise. Therefore, from your point of view, you have no reason either way to judge me as zombie or non-zombie. Nevertheless, I know that I am not a zombie, as I have core-epistemology of that fact. From the first person view, all non-zombies have core-epistemology that they are not zombies.

The fact that I would deny zombiehood even if I were not a zombie has no bearing on the fact that I claim it to be non-zombie. If I were a zombie I would deny it, and this does not affect my denial of zombiehood.

Phenomenal experience entails core-epistemology of that fact, this being the implicit epistemic justification for the phenomenal realist premise. However, nothing follows from both the facts that phenomenal experience entails core-epistemology, and the possibility of zombies. Where phenomenal experience obtains, people have core-epistemology of that fact; when phenomenal experience does not obtain, people may claim to have 'core-epistemology' of that 'fact'. If I were a zombie, I would deny it, and I am not a zombie.

Now I wish to consider a separate case. I will use the zombie notion to introduce it.

4.2.3 Mechanical claiming

Claims of phenomenal realism will include terms that we consider having a referent. Our claims and zombie claims are different in that our claims refer while theirs do not. We claim because there is such a referent. Zombies do not claim for the same reasons that we do.

There exists a complete functional and behavioural explanation for the act of claiming. The act of claiming is the claim considered from the third person point of view. The act of claiming is the same in the zombies and us. This explanation would explain our claiming of supposed inherent first person ontologies. It would explain our claimimg of core epistemology, and about how we are clearly not zombies. This explanation would be an explanation of our behavioural acts of claiming.

That such an explanation exists indicates that part of the explanation of our functioning and behaviour will include explaining the third person reasons of our first person claims of phenomenal realism. The explanation will state that, given this functional organisation, it is possible that there will be acts of claiming "phenomenal experience is irreducible". It will be able to point to the third person causes of such a claim. The third person aspects of the claim "phenomenal experience is irreducible" can be explained in third person terms. Our third person authority can then state that our behavioural claims of 'phenomenal realism' are the result of various dispositions to behave, functional organisations, and third person empirical causes.

The third person explanations of the third person aspects of claiming refer to behavioural dispositions, functional organisation, or other items of third person accounts. The entirety of this explanation of our claiming of phenomenal realism, including our concepts of first person specific phenomena, refers to a part of us. Part of the functional account of ourselves will account for our claiming. It will state that there is part of ourselves, explainable in third person terms, which provides us with the ability to make claims in accordance with our supposed core-epistemology.

This part of ourselves, this 'mechanism', which is referred to in these explanations, is that which allows us to express our core-epistemic situation in third person behavioural terms. Thus, the explanation of our claiming phenomenal realism can be considered the explanation of a mechanism within us, the purpose of which is to generate claims of phenomenal realism. This mechanism is independent of phenomenal experience; it is a purely third person mechanical functional part of ourselves. This mechanism exists in both zombies and us and is the same in both cases. In zombies it allows for claims of phenomenal realism also, it is just that in such a case, it is empty claiming. In our case, however, that claim is in alignment with our core-epistemic situation.

"Phenomenal realism is mysterious in that it does not succumb to third person explanatory attempts. It has aspects of subjectivity, of an indexical subject, that are not adequately addressed in third person terms. In short, third person absolutism is incoherent. It is just immediately apparent that this is the case. It cannot be denied, as it is immune to skeptical argument. It just is that way, and I have core-epistemic knowledge of that fact." That is what I say, because I have a mechanism that allows me to make such claims. Phenomenal realists write long and complex books attempting to show that their phenomenal realist assumption is true for the reason that there are important mysterious aspects of phenomenal experience that are left behind with third person explanation. This is to be expected, because they have a mechanism, independent of the truth or falsity of the phenomenalist realist premise, which allows them to produce such arguments.

The mechanism creates the behaviour of claiming and justifying the phenomenal realist premise, and this applies to behaviour generally. It does not refer to outward moving behaviour of persons or their vocal behaviour alone. It refers to any behaviour that is discernible in third person terms, any empirically discernible change, at whatever level, counts as behaviour.

My claim of phenomenal realism is correct but my claiming that this is so has nothing to do with its being so. My claim, from my point of view, may indeed refer to ontological kinds, but my claiming can be explained without phenomenal experience. Phenomenal realists, who argue for phenomenal realism, are correct in their claims. However, that they make such claims can be explained for reasons apart from phenomenal realism.

Phenomenal experience plays no empirically discernable role in our claims of phenomenal experience and our secondary claims that this is justified. Phenomenal experience does not interfere with the third person domain. The 'reason' for our claiming is core-epistemology. But this 'reason' plays no causal, behavioural or functional role.

It is because our core-epistemology plays no such role, that we require the mechanism. It would be awkward if we did not have such a mechanism, independent of phenomenal experience, which allows for claims of phenomenal experience. However, we have such a mechanism, as do zombies. Both zombies and ourselves make claims, and our claims are true. This mechanism is vital to us, allowing us to make claims about our situation.

Without this mechanism, we would never speak of phenomenal experience as an ontological kind. We would never behave in a way that would indicate that it is an ontological kind. Without this mechanism there would be no behaviour in us that would indicate, or be involved in indicating that the phenomenal realist premise is correct. We would still, however, have core-epistemology of experiences.

4.2.4 An absence of mechanism for claiming

The mechanism, being independent of phenomenal experience (it is present in zombies), is not essential to phenomenal experience. There is no reason to suggest that we would not be conscious without it.

The mechanism is that which allows us to make claims both to others and ourselves. However, core epistemology is not a self-claim; it is just a blunt core epistemic fact. We do not 'claim' to ourselves that we have core-epistemology. We have core-epistemology, and we can make this claim to others.

Without the mechanism, we would not be able to make a claim of core-epistemology, but we would still have core-epistemology. The mechanism is that which allows for claims of first person items, and without it, from the point of view of others, we would be phenomenal eliminativists. However, we would still have core-epistemology that eliminativism is not so.

The mechanism allows us to make behavioural claims to ourselves regarding our core-epistemic situation. Thus, it allows us the third person aspect of thinking to ourselves, "phenomenal realism is so, and I shall defend this basic epistemic fact against eliminativism".

Now, if 'thinking' involves third person discernible behaviour, then we would not be able to make claims about phenomenal realism to ourselves. We would not be able to stand in front of the mirror and say, "indeed I do have non-eliminativist phenomenal experiences". Nor would we be able to say this silently to ourselves. All these things have third person behavioural aspects, but the third person mechanism that allows for such aspects is missing.

We can say that, if thinking has behavioural correlates, the mechanism is necessary even for us to tell ourselves about our core-epistemic situation. It is not necessary for our core epistemology of course. It is necessary for our judgements and 'second order thoughts', of our core-epistemic situation regarding phenomenal experience.

Without the mechanism we would have certain difficulties reminding ourselves of that fact of our core-epistemic situation, or thinking about that fact, or discussing that fact, or making certain beliefs about that fact, or creating theories about that fact. We would be incapable of 'second order' thoughts about, and resting on, our core epistemology.

In order for us to claim to ourselves that phenomenal realism is correct, we require the mechanism. Moreover, if we do not have this mechanism, then in order for us to think about phenomenal realism, this thinking process must not be behavioural. If thinking is necessarily correlated with behaviour, we cannot think about phenomenal realism without this mechanism.

4.3 Considering the scenario

What is the experience of a person without such a mechanism? They are left with core-epistemology, but no second order thoughts about this epistemic fact. They still have experiences and core-epistemology as they are not zombies. However, the mechanism, which ensures that they can make claims in alignment with their core-epistemic situation, is missing. Such persons will not claim to see any mysteriousness in phenomenal experience, or any ontological fact of experience.

These mechanism-less persons, from the third person point of view do not express any difficultly in the question of phenomenal experience. The person without an intact mechanism (or with a broken mechanism), will not see any mysteriousness in consciousness at all. They will deny that there is an essential indexical or that there is anything special in the first-person point of view. They will claim that third person explanation is all that is necessary. When asked why third person explanation does not seem to entail any facts about phenomenal experience, they will appear baffled; they will reply, "entails what? There is nothing it could entail".

They act in such a way because they do not have second order thoughts about core epistemology, and that core epistemology has implicit 'facts' about ineffability, of phenomenal realism, of the incoherence of eliminativism and so on. Such people still have experiences. They complain about headaches. However, they will say, "my judgement of a headache that I have is the headache that I have".

Phenomenal experience is accepted. We claim phenomenal realism and this claim is true. Nevertheless, the reason we make the claim is that we have a mechanism. The third person explanation of this mechanism explains why we make claims that third person explanation is insufficient. Third person explanation holds the key to our behavioural aspects of claiming third person explanation as inadequate. Without this mechanism, we would find the third person explanation of why we make phenomenal realist claims to be complete and adequate. We would not see any role for our own phenomenal experiences in the phenomenal realist claim.

4.3.1 Core-epistemology distinct from judgement

Those without the mechanism have genuine phenomenal experiences. They also have core-epistemology of this fact. This is part of the condition of an inessentialist view: those that have phenomenal experience have core-epistemology of that fact.

The mechanism-less do not make claims of phenomenal realism or of core-epistemology. Thus, they have core-epistemology of something that they will never claim to know. Indeed, they will claim not to know about this core-epistemology, and may claim 'core-epistemology' to be meaningless. Does the fact that they have core-epistemology of something that they will deny, cause difficulties?

If thinking and other mental states that are judgements about core-epistemology have third person behavioural correlates then the mechanism-less will never even judge that they have this core-epistemology even to themselves. They will appear, from the third person point of view, to never desire to correct the claims they make. They have core-epistemology, but deny this fact, and they will not appear to see a problem in such claims.

The mechanism-less will not think to themselves, "well now why did I say that? Of course, phenomenal realism is so, but why do I keep claiming I otherwise? I have this core-epistemology, of which I cannot be mistaken, but I consistently deny it". They have core epistemic facts, but they do not have the mechanism that ensures the capacity to make judgements, claims, and second order thoughts in alignment with these facts.

The mechanism-less have no internal conflict. They do not suffer: "This is terrible! I keep saying things which I know to be false, but I cannot seem to help it!" In order for such mental states to be possible, these states must have no functional, behavioural, or causal role.

Searle seems to allow for such a possibility (Searle 1992). At least, he mentions a case where a persons ability to make claims regarding their epistemic situation is compromised, and they are aware of this fact. He imagines such a case as arising via brain replacement by silicon devices. In such a case, our actions would seem not to be under our control. In Searle's brain replacement thought experiment, their can be internal conflict.

Physicalism, rather than functionalism, may allow for internal conflict scenarios, if they considered a static physical state as fixing mental states of conflict. If such a static physical state was shown to have no bearing on the otherwise normal functioning of the person, then perhaps internal-conflict scenarios are plausible. However, mental states of conflict are complex and can be extended over time. In short, allowing non-behavioural physical states to determine such mental content is problematic. It allows for mental states and behavioural states to be utterly independent; it allows for trapped homunculi. It allows for mechanism-less persons who are aware of the fact that they are denying what they know (in the core-epistemic sense), and desire fruitlessly not to deny what they know.

The case of internal conflict requires complex mental thought processes over time, which have no behavioural correlates. Thoughts such as "now why did I claim to be an eliminativist along the lines of Dennett just then, when I clearly side with Searle?" must have no behavioural aspects. In addition, the case of internal conflict resides on the premise that, if we did not have the mechanism, we would realise that we cannot make claims about our core-epistemic situation. This is not coherent, as our core-epistemic situation is not the reason we make claims in the first place. We have core-epistemology, but it is the mechanism that allows for claims; phenomenal experience plays no direct causal role, and so cannot be the instigator of behavioural claiming. These claims are in alignment with the facts in our case, but not in the case of zombies. Since our behavioural claiming is independent of our core-epistemology, we would not notice a difference in what follows from our core-epistemology if we were mechanism-less.

Nevertheless, say that internal-conflict situations are accepted. Then this is what we have: genuine Chalmers-like dualists walking around trapped in the bodies of Dennett-like eliminativists, having complex content mental states (perhaps preparing complex arguments) which they will never express, all the time experiencing that they have no control over their behaviour and speech acts in this regard.

In any case, it does not matter to this argument whether one rejects the internal conflict situation or not. The possibility of the mechanism-less is the issue. If the mechanism-less are considered a problem, then that problem reflects back to inessentialist phenomenal realism.

4.3.2 Attempting to refute the scenario

The mechanism-less are a problem. We want people who say they are conscious to be conscious. The possibility of the mechanism-less has more problematic implications than the possibility of zombie possibility. Zombies are merely mistaken in their claims (it matters not, as they are zombies with no internal lives), but the mechanism-less seem to be, from the third person point of view, in denial (and they do have genuine internal lives). And they never even think about what is true for themselves, and they never realise that what they say is false.

Inessentialist phenomenal realism must not allow for the mechanism-less. This means it must provide an account of why the mechanism-less is not even a logical possibility. However, part of inessentialism is the fact of the mechanism. There is a distinction between core-epistemology, and claiming, so there is a mechanism. And this mechanism is not part of having phenomenal experiences, as Zombies have this mechanism.

Thus, inessentialism must claim that having a mechanism is an essential part of having phenomenal experiences. It must claim a dependency between core-epsitemology and the mechanism. The mechanism operates independently of phenomenal experience, but in order for phenomenal experience to obtain, there must be a mechanism. Thus, if the mechanism were to be damaged in a person, they would turn into an eliminativist zombie.

Denying those without a functioning mechanism phenomenal experience is problematic on a number of counts. Firstly, it suggests that strict eliminativists may indeed be zombies. It openly allows for the possibility that Dennett is a zombie on the basis that he acts like his mechanism is damaged. The behavioural differences between those in which the mechanism is broken and those in which it is fully functional are not sufficiently different to claim that those in the former group are zombies. However, the mechanism is not related to phenomenal experience. It is reasonable to assume it can be removed, or rendered inactive, without effecting that to which it bears no relation.

Treating an eliminativist philosopher (such as Dennett, for instance) as a zombie would require knowing that his mechanism is damaged, and this would be a third person empirical endeavour. But perhaps Dennett is aware that he is a dualist, and merely argues eliminativism for the intellectual challenge. Unless we knew the specifics of the mechanism, we cannot judge it absent based on claims a person makes.

The attempt to rule out the mechanism-less can be stated as the condition that those who have phenomenal experiences must be able to make claims in alignment with this fact (without the condition that they will make such claims; that they could make claims is enough). This does not rule out zombies. It does, however, state that purely first-person experiences and epistemology only obtain in cases where there is potential for empirically discernable claims which are in agreement with this fact.

This condition, however, cannot be checked, just as we cannot judge Dennett as mechanism-less based on his claims. The condition merely states that it must be possible to make claims. Perhaps someone does not make these claims. On the other hand, perhaps they do make the claims, but they are lying. Perhaps Chalmers is a strict eliminativist, but finds arguing for dualism an intellectual challenge.

4.3.3 Concerning the mechanism

We know of the mechanism from the first person, but cannot know of it from the third person. The mechanism, being third person mechanism, ought to be as understandable as any mechanism. Its job is to provide us with claims of third person inadequacy, the mechanism is the embodiment of a third person explanation for the epistemological explanatory gap. Moreover, the epistemological explanatory gap is the reason of why we make the claims that we do.

However, we cannot know the mechanism from the third person. We know that we have such a mechanism. If we knew and understood the mechanism in ourselves, then the mechanism would not be working. Our understanding, therefore, would not qualify as understanding of the mechanism. The mechanisms job is to provide claims and judgements about phenomenal realist notions. Understanding the mechanism in ourselves would be the same as understanding why, in purely third person behavioural and empirical terms, we claim that phenomenal realism is not third person.

Imagine we could empirically find the mechanism and understand it. Then we know that we make phenomenal realist claims for reasons independent of phenomenal realism. We still have core-epistemology, however, but we would understand our claiming as being nothing more than the actions of a third person mechanism. We would say, “so that is what the mechanism is in myself. That makes it very clear to me why, in third person terms, I keep on claiming that phenomenal realism is not third person”. If we understood that our claims had nothing to do with phenomenal realism, would we still bother to claim phenomenal realism? Yet, if we would stop claiming phenomenal realism, then the mechanism would no longer be fulfilling its purpose.

If we claim something like, "I understand the mechanism, but it does not account for the core-epistemic fact", this is a claim for phenomenal realism. In addition, our understanding of the mechanism would give an account for our claiming that the mechanism does not account for the core-epistemic fact. Our understanding of the mechanism would give us an understanding of why we think we need to qualify our understanding of the mechanism: "but it does not account for the core-epistemic fact". Given this, our understanding of our statements of phenomenal realism, and that the mechanism does not account for it, would be altered. Thus, if we truly understood the mechanism, it would, by definition, not be fulfilling its purpose in ourselves. Understanding it to that degree is contradictory. Thus, if inessentialism (and thus the mechanism) is accepted, there are cognitive limitations that prevent our knowing and understanding it completely. There are epistemic limitations, as our makeup defines out epistemic abilities, so cognitive closure of this sort is acceptable.

An account which suggests something along the lines of the mechanism described is given by Elitzur: "Consciousness must be the reason people are bothered by problems of consciousness. If someone says that he cannot understand his experiences by what he knows about himself, this expression of bewilderment cannot be explained by any physical process, unless one resorts to the farfetched claim that the person expressing this bewilderment is lying" (Elitzur 1989, p. 9). Elitzur in this statement discounts the case that we may not be able to know our workings to such a degree. He assumes we could know our workings completely, in an empirical, *a posteriori* way. Thus, Eliztur argues, with complete knowledge, we could become eliminativists, or if we did not, experience must then have direct causal capacities. A thing cannot know everything about itself; this is a self-referential impossibility. It is akin to standing behind oneself, or assuming one can store a box within itself, simply because boxes can store things. However, Elitzur in the same paper considers, and discounts (because of the assumption of complete knowledge) the existence of the mechanism. There is the possibility for argument that we would not need to know everything about ourselves, to understand ourselves completely: there may be redundency, for instance. Howevever, there is no contradiction in accepting inessentialism, and therefore accepting the mechanism, which entails that we could not know this mechanism to a sufficient degree. The last condition merely requires that this epistemic limitation is accepted. The reason the inessentialist situation has difficulties is not for the reason Eliztur suggests.

4.3.4 A conflict with a premise

Inessentialism relies on judgments about core-epistemology being in alignment with core-epistemology. This allows Chalmers to accept his own judgments about his core-epistemology.

Zombies make such judgements too. But this is not a concern as long as those that have experiences can make judgements about experience. Inessentialism separates core-epistemology and judgement. However, it requires judgement in the presence of core-epistemology. It requires that, in addition to core-epistemology, we have judgement states about this core-epistemology.

The presence of judgement in alignment with the facts of core-epistemology allows Chalmers to communicate about core-epistemology. He has core-epistemology in any case, even if he does not talk about it. But his judgements about core-epistemology have nothing to do with core-epistemology. Chalmers has core-epistemology that refutes the idea that he is a zombie. In addition, because there are possibilities for judgement in alignment with this fact, he can claim that he is not a zombie.

Where does the justification for inessentialist phenomenal realism lie? The justification, ultimately, is in the fact of core-epistemology, and this is immune to skeptical argument: where there is phenomenal experience, there is core-epistemology. However, there is an implicit requirement that judgement about core-epistemology is possible.

The mechanism-less scenario is a possibility of someone who is entirely without judgement regarding his or her phenomenal experience. If the term 'conscious of' is used to indicate the presence of a judgement on a core-epistemic experiential state, then this person is not 'conscious of' their experiences. Because of this, they will not be a phenomenal realist. This person will not have any mental states of judgement of core-epistemology. This person will deny core-epistemology.

Where does the justification for inessentialist phenomenal realism lie now? We return to one premise of phenomenal realism: we cannot be mistaken about the fact that we have experiences. There is a degree of certainty. Yet, it seems that core-epistemic certainty is not enough in the face of the mechanism-less, because it does not ensure that we could tell ourselves, and make claims to others, about core-epistemology.

Whether or not the mechanism-less are considered a problem for inessentialism rests on whether it is acceptable that a mechanism- less person 'knows' that they have experiences. Here is where there is some vagueness. They have core-epistemology, but they will never claim it, never have judgements about it, and never have 'second-order' or 'higher-order' thoughts about it. In what sense, then, do they 'know' they have experiences?

4.3.5 On the requirement of alignment

The mechanism-less are a logical possibility. They are more possible than zombies are. The mechanism-less may even be a possibility in this world, if inessentialist phenomenal realism is so. The only condition for the mechanism-less is a degree of brain damage, and that is an actual possibility, not merely a logical one. The possibility of the mechanism-less is merely that some piece of machinery is missing.

It is not the mechanism-less that are the problem, so denying the possibility of the mechanism-less is not a solution. The difficultly is the separation between core-epistemology and claiming. Claiming includes our claiming to ourselves as well as others. It encompasses all judgements that can be made about core-epistemology. The mechanism allows for such judgements. There needs to be a mechanism because there is no necessary link between core-epistemology and judgement.

The mechanism-less are the result of this clean distinction between core-epistemology and judgement. Inessentialism does not provide an account of the alignment between core-epistemology and judgement. Arguing that the mechanism-less are not possible in this world is not good enough, just as arguing against zombies in this world is not good enough to refute claims that invoke zombies; it is logical possibility that counts.

Both zombies and the mechanism-less are cases in which there is miss-alignment between core-epistemic facts and judgement. In one case, there are no core-epistemic facts, and in the other case, there are no judgements. But zombies are not a problem, because zombies do not contradict the strong sense of epistemic certainty required for inessentialist phenomenal realism.

An account of alignment is required if inessentialism is to hold. This account of alignment must describe how it is *necessarily so* that those who have phenomenal experience have the ability to have judgements in alignment with that fact. This does not rule against zombies.

Denying the possibility in this world of the mechanism-less is one way of providing alignment in this world. But it is not necessary alignment. Necessary alignment would show that the mechanism-less are not logically possible. But, following from the inessentialist phenomenal realist premise, they are logically, and empirically, possible. Thus, an account of necessary alignment is not possible within an inessentialist context. An account of necessary alignment would provide a necessary link between core-epistemology and certain behaviours (encompassed by the mechanism) which have, by hypothesis, no necessary link to core- epistemology. This is impossible. Therefore, an account of the supposed alignment between core-epistemology and judgement would actually be an account which did not see these as being distinct in the same way as inessentialist phenomenal realism does.

The phenomenal realist claim is that core epistemology entails judgement knowledge in this world. It is not a necessary entailment. However, given the scenario of the mechanism-less, a necessary alignment is required. However, this is impossible within the constraints of the inessentialist phenomenal realist premise.

Perhaps the most recent and cited work on inessentialist phenomenal realism is the book by Chalmers. He devotes considerable time to this question of alignment (Chalmers 1996a). Yet his arguments for alignment are all arguments based on possibility *in this world.* Chalmers does not want mechanism-less persons, or zombies, in this world. He uses a 'fading/dancing qualia' argument to show that there is, in this world, alignment between core-epistemology and judgement claims. This avoids him having to add an additional postulate to the inessentialist phenomenal realist premise. This is *not* an account of alignment, it is an argument that there *happens* to be alignment in this world, though this alignment is logically contingent.

Chalmers ponders the need for an extra ingredient which provides alignment, but argues (with dancing/fading qualia) that this is not necessary *in this world*. He does not consider the need for this alignment to be necessary. His fading/dancing qualia arguments are about possibility in this world.

There are four possible cases regarding core-epistemology and judgement. First, there is the phenomenal realist, a person with both core-epistemology and judgement. Second, there is someone without experiences who makes 'empty' judgements; this is the zombie. Thirdly, there is someone with neither experiences nor judgement; this is an eliminativist zombie. Fourthly, there is someone with experiences and no judgement; this is the mechanism-less case.

Inessentialism is coherent if the fourth case is precluded, and precluded necessarily. In that case, if experience obtains, so does judgement. Thus, all know about their experiences, if they have experiences. Zombies are not a threat to this; the alignment is one way: core-epistemology to judgement.

4.4 Conclusion: what this scenario tells us

The distinction between core-epistemology and judgement is the problem. This distinction, however, seems to be intuitively correct. Such a distinction is argued for explicitly by Block, Chalmers, and Searle. The distinction is one between 'access consciousness' and 'phenomenal consciousness', also called 'core-epistemology' and 'judgement'. 'Judgement' encompasses the 'higher-order thoughts' of higher-order-thought theories.

One option is to question the validity of 'phenomenal consciousness' as distinct from 'judgement' or 'access consciousness'. This option is eliminativist. The other option is to maintain a phenomenal realist stance, but recognise the difficultly with keeping phenomenal experience distinct from judgement.

4.4.1 Eliminativism

To fix this situation, the easiest path to take is to jettison the concept of phenomenal realism in the view. This is essentially what Dennett does. He argues that the entire notion of core-epistemology distinct from judgement and behaviour is incoherent. In his view, knowing something means being able to make a judgement. Yellow, for Dennett, is the judgement of occurrent yellow.

The important point about Dennett's view is not that it is eliminativist, but that it does not contain a notion of core- epistemology divorced from judgement. The criterion for an experiential state, in his view, is that there is judgement about this experiential state.

Eliminativism is, however, a very clear and open option. The mechanism-less showed that there are people, who have experiences in accordance with the inessentialist phenomenal realist concept, but are never aware of them, never think about them, and consistently deny that they have them. They do not have any judgements about experiences. To an eliminativist, this situation is incoherent. Moreover, to the eliminativist, all that has to be done is to provide them with judgements. What is it that turns a mechanism-less person from being an eliminativist with regard to phenomenal realism, and a phenomenal realist? It is not experience, but the judgement. So to Dennett, the judgement does 'all the work', and nothing else is required.

4.4.2 A criterion of judgement

The criterion for judgement entails having phenomenal experiential facts supervene on behaviour and function explicitly. This may rule out the logical possibility of zombies. The difficulty is to maintain phenomenal realism in the presence of this criterion for judgement. The criterion for judgement is a necessary one; it is not a contingent alignment between judgement and core-epistemology in this world. Thus, the phenomenal realism concept will not have core-epistemology distinct from judgements about core-epistemology; it will not have experience, and knowledge of experience as distinct. They will be necessarily related.

This would be a difficult to formulate while maintaining inessentialist phenomenal realism, because zombies are seen as logically possible. What is required is to take phenomenal experience out of a behaviourless realm in recognition of the fact that the concept of experience utterly divorced from the behaviour of any aspect of empirically discernable world is problematic. Our intuitions about this case are, however, strong. They influence what we think of as logically possible, and there are arguments that experience is logically distinct from behaviour.

4.4.3 Concluding remarks: Is eliminativism the only option?

Neither functionalism nor physicalist functionalism is without this difficulty. Physicalist functionalism does not help, because the solution is not the provision of extra constraints in addition to function. The issue is not one of determining or fixing phenomenal experience, it is one of the distinction between core-epistemology and judgements about phenomenal experience.

The one aspect upon which the zombie scenario rests is the concept of our ability to know, in principle, the entire behavioural, causal, and functional process of the actual world. Further, there is the view that this complete knowledge would not refer to or entail, phenomenal experience. If this assumption of complete understanding is dropped, then our concept of experience as logically distinct from behaviour needs careful consideration. It may be that function and empirical physics under specifies, and if this is accepted, then the 'completeness' of empirical and functionalist views does not count for the irrelevance or inessential nature of phenomenal experience. And neither does an interactionist view result, as this requires 'completeness' of functional and empirical views which is violated by phenomenal experience.

Views, which explicitly consider our epistemic limitations, and so are more apt to disallow concepts of 'completeness' of empirical knowledge, have more room for phenomenal experience having a functional role, while at the same time avoiding interactionism. Thus, the core-epistemology/judgement distinction is not a problem in such a case. McGinn's view is one such view that acknowledges the problems presented here, and allows for a solution. McGinn takes it that function always underdetermines intrinsic nature, so absent/inverted qualia cases are not incompatible with consciousness having a function, as they are cases which arise in the context of incomplete knowledge (McGinn 1981). Inessentialism, therefore, is not something that arises in the context of his view.

It is easy to image zombies. However, with the judgement criterion, zombies are not logically possible. Does this entail eliminativism? It must entail eliminativism if we take it that we can, in principle, know the entire causal structure of the world without reference to what we consider terms of phenomenal experience. If this is possible, then zombies are possible. However, with the judgement criterion, they are not possible, so something must go. What must go is (a), the phenomenal realist premise, or (b), the idea that the empirical world tells us completely what the world contains, and thus the world is in principle knowable completely without reference to phenomenal.

Thus, if phenomenal realism is kept, then the 'inessentialist' concept is void. Also, there is need to inquire as to the epistemic notions underlying the belief in a third-person complete understanding of the behaviour of the world. For if this is maintained and phenomenal realism kept, then interactionist dualism is the result.

Chapter 5

Implications

5.1 Introduction

Functionalism rests on the notion that the functioning of an object can be captured completely by abstract formalism. Thus, it is independent of specific ontological commitments. If functional explanation of something is seen to encompass the explanation of the behaviour of that thing, then behaviour is explainable independently of specific ontology also.

Universal computation allows that the functional account of an object, being independent of specific ontological commitments with regard to that object, could be the same as the functional account of some other object which has different ontological aspects. By different ontological aspects is meant general differences of a physical, material, or structural nature. That a functionalist account is the same in both objects refers to the functional organisation of those objects and nothing more. It is also possible that a functional account could be found for some object, and that functional account used to constrain the actions of another object such that it then conforms to that functional account. The only criterion is that an object be capable of supporting functional organisation; in computational terms, it must be computationally universal.

Functionalism, then, separates what something does, from what something is. The latter is a question of ontology, and there are no specific ontological commitments beyond the implicit commitment that the object can support functional organisation. Because of this, there is no room for what-something-is, to be part of the explanation of what-it-does. This is just to say that ontological issues are not seen as necessary for functional and thus behavioural explanation.

Functionalism 'skims' functioning and behaviour (if not all behaviour, then important behaviour), from the ontology of an object. Thus, there is the assumption that decoupling behaviour and ontology in this way is justified. It is this point that caused the difficulties for inessentialist phenomenal realism, and leads most easily to eliminativism as an answer to this difficulty.

Avoiding eliminativism requires that the notion that functionalism, and empirical explanation, can completely specify behaviour be reconsidered. For if it is reconsidered, then there are no grounds for declaring either inessentialism or interactionism on the basis that there are 'complete' functional or empirical accounts of the behaviour of particular objects.

In essence, avoiding eliminativism requires that we consider the possibility that behaviour and ontology cannot be decoupled, that what-something-is is essential to what-it-does. The purpose of this chapter is to consider the implications of not decoupling behaviour and ontology.

In the argument concerning functionalism a difficultly arose concerning the epistemic contingency between phenomenal experience and the underlying ontology of the world. The problem being that there is no discernible epistemic relation between the appearing of the world and any underlying ontology. Functionalism, however, does have an implicit ontological commitment to 'something that can support functional organisation', though this phrase has no explicit ontological commitments associated with it.

The simulated persons epistemic situation, and so their phenomenal experiences, are in no way determined directly by the ontology of the implementation; they are dependent only on its functional organisation. These persons had three options, either a phenomenalist ontology, which drove them to a many-worlds view, an infinite regress, or a functionalist refuting ontological commitment. None of the alternatives are valid, since they all break the epistemic bounds of the simulated persons.

The lack of epistemic justification for a transcendent inference is the difficulty. It is a difficultly only because it is not possible to ignore or deny, to the extent necessary, the apparent situation of the simulated persons. With the concept of 'their apparent situation' are concepts related to phenomenal experience. These cannot be denied. The difficultly can be summed up thus: if phenomenal ontologies are, in and of themselves, important, and that there is an epistemic contingency between phenomenal experience and the underlying ontology of the world, a difficultly arises.

5.2 Preliminary concerns

5.2.1 A note on empiricism

Pragmatic approaches to explanation do quite well without concern for ontological issues. Every approach to explanation rests on certain foundations that are assumed, and positivistic explanation is no exception. 'The world' has to be observed, in some manner, in order to provide empirical data. There is direct and indirect 'observation' of 'the world'. Categories and terms are built around the results of this 'observation' of 'the world'.

There is usually a sense of privilege. For instance, redness and other phenomenal or secondary properties are deemed lesser than primary properties, or are denied ontological status, whereas ontological commitments to objective, empirical items are made based on observations.

’Observation’ of ‘the world’ comes from someone, or something. It is a view from somewhere, if not necessarily a view by someone. Agreement as to lots of ‘observations’ by someone or something allows the particular someone or something to matter less. A consistent view from many different somewhere’s means that a particular somewhere is not that important. In time, with stronger agreement, a view from any particular somewhere is acceptable: a view from anywhere.

The view from anywhere can be conflated with concepts of the world itself. The view from anywhere may be considered to provide a complete view of the world. Separating a view from somewhere, and the world itself, is a separation of phenomena and noumena, and that may not be accepted. Certainty, the line between so-called direct observation and inference is blurred, if observation and inference are kept distinct. This is even so in our direct visual experiences. The completeness of an empirical view suggests that noumenon concepts are somewhat flawed, or else it suggests that the empirical view details all the important aspects of the noumenal world.

Conflating the view from anywhere with the view from nowhere, and tending towards an objective view can arise from accepting that there is no clean distinction between observation and inference. In the absence of such a distinction, it is not appropriate to divide the world into phenomena (a view from somewhere/anywhere), and noumena (a ‘view’ from nowhere).

The world as it is in itself, is not a view from somewhere, or anywhere. It is a ‘view’, if that can be said, from nowhere at all. The ‘view’ from nowhere is not a view at all, and so no empirical information is forthcoming from it. A complete empirical view, in the context of noumena, seeks to combine the view from anywhere and the ‘view’ from nowhere, in that it would not admit to transcendent and empirically unknowable things (noumena) in the world. Where neither the epistemological nor ontological aspects of ‘observation’ in ourselves is understood, nor the line between observation and inference understood, then the combining of empirical accounts with noumena concepts is to be expected.

With a noumena concept as regards ontology, an empirical account describes the world, yet the empirical world does not necessarily contain just what an empirical account says it contains. The more common view currently is that it is in principle possible that an empirical account can tell exactly what the empirical world contains. In other words, ‘complete’ accounts are considered possible, in which all functioning and behaviour, and all causal antecedents to any event, are knowable empirically. Where there are apparent limitations to empirical knowledge, these limitations are known, and are not seen to be relevant at the level of cognitive science or of philosophy of mind.

The view from anywhere has implied concepts of observation, as do all empirical accounts. The degree to which such accounts can then be called ‘objective’ depends on what meaning is given to this term. Empirical accounts, if purported as potentially complete in principle, suggest a degree of independence from us, from observation, and from the epistemic context in which they were derived. Such independence is impossible.

This is the difficulty with strong ontological eliminativism with regard to phenomenal experience. An empirical or functional view, one that seems independent of ‘observation’ and the attendant notions of experience suggests that there is something in the concept of ‘observation’, which the view does not address. Thus, the reaction to this can be to eliminate that item which it seems not to address. This is done on the basis that the account addresses enough already.

5.2.2 A note on determinism and causal closure.

Computability and determinism are related. If a system is computable, it is deterministic. Computing something allows it to be determined. Whether or not a particular particle will ever collide with another particle is a question, which can be asked of a deterministic universe, which is semi-computable (semi-decidable). In general, it can be answered in the affirmative, but not the negative. In specific instances, the question can be fully decidable, but in general, it is semi-decidable. The answer to the negative case is, if it never collides, the answer is no. Implementing the ‘if it never collides’ bit is tricky. This is just the halting problem. Such a billiard ball system is a fully deterministic and computable system, yet the asking of a question of this computable system need not be an effectively computable task. Determinism and computability are distinct in that way. No conclusions about determinism from our inability to compute can be drawn; neither does a limitation on empirical accounts, which may lead to those accounts having apparently stochastic processes, entail anything regarding determinism.

The view that, in general, behaviour is not captured completely in an abstract functional way does not have any implications regarding determinism. It only has implications for determinism, as we can understand it. It implies that our abstract functional accounts will not determine behaviour completely, in principle, though in practice, this difficultly may be slight. This is just to say that something may appear non-deterministic to us. Currently, radioactive decay appears non-deterministic to us. This says nothing of whether radioactive decay is, ontologically and metaphysically, a random, non-determined occurrence. Of course, with other commitments, the empirical stochastic nature of radioactive decay may be used to infer non-deterministic random occurrences. We may be unable to precisely determine certain occurrences, though those occurrences are precisely determined.

An inability to calculate, compute, or predict a future outcome has no bearing on whether or not the system is deterministic. This is so whether the inability is because of practical concerns, such as poor empirical recording equipment, or because there are non-trivial epistemic limits to what we can empirically know. To illustrate, consider a maximal string. Each digit is not calculable from the previous numbers in such a string. As far as computational calculation is concerned, each additional digit is not computable from the previous ones. Thus, computationally, each digit is not 'dependent' on the previous digits. If a physical occurrence seems to generate maximal series, and it is noted that verifying a maximal series is not possible, but assuming such processes exist, then difficulties arise. An example of such a process may be radioactive decay. Those that truly consider it a non-deterministic process are saying that it could potentially generate an infinite maximal sequence. If each member of the maximal string represents a physical event, each physical event is not calculable from any previous events. However, that does not mean that the event is not 'dependent' on antecedent events, or that it is 'causeless', or 'not determined by previous events'. It is computationally independent, and that is all that can be said. That we cannot determine that event (which is stating that we cannot generate the next member of a maximal series) does not mean the event is 'causeless', that occurred for 'no reason', that it is not dependent on the past.

In the context of our explanatory endeavour, there are maximal series, and therefore, it is logically possible that there are physical events, which, within this endeavour, are not dependent on past events. This is all within the context of an explanatory endeavour only. Ontological commitment and metaphysical statements must be carefully considered, if a form of epistemic limitation is accepted.

If behaviour is not considered distinct from ontology, then phenomenal realism may not be relegated to an inessentialist view, as our objective or empirical accounts will necessarily be incomplete. Thus, there are no conflicts with non-inessentialist ontological items compromising the causal closure of the empirically knowable world. This non-conflict, however, requires that behaviour and ontology are not separated, and the resulting limitations on empirical accounts are accepted.

5.3 Looking at hierarchical description

5.3.1 Symmetry, and the status of 'nothing but' facts

Empirical accounts, if considered 'complete', or considered to be reaching completeness, which is considered to be attainable in principle, will have a preferred set of ontological commitments and descriptions. Descriptions or facts based on these ontological commitments will have 'privilege' in an empirical account of a process or object. For instance, the microphysical facts may be considered the privileged facts. They will form the base upon which other facts may supervene, or they may form the base to which other facts are reducible.

An empirically based account will have certain ontological commitments that can be described. These descriptions may be considered as somehow accurate or complete descriptions. The ontological commitments themselves, as they are described in terms of interaction, may be considered true, or complete, or privileged facts. A statement of this notion of completeness is given by Pettit, in reference to his physicalist view: “the empirical world contains just what a true complete physics would say it contains” (Pettit 1993, 222).

‘Privileged’ description, seen as the base to which other descriptions and other facts reduce or supervene, does bring with it the difficulties of the status of the less privileged facts. These may be the reducible facts, or the supervenient facts. Where the privileged facts are taken to be complete, then the status of non-privileged facts is a complex issue. If they are seen as facts in themselves, such as the case where these facts are irreducible, or are seen to refer to ontologies outside those in the empirical account, then they have modal status. Such facts further specify how the world is, over and above the specification provided by the privileged facts. This is the case if the experiential colour facts of Mary are considered in this way. However, there could be arguments that the less-privileged facts are encompassed by the privileged facts, and so there are no modal constraints.

The separation into privileged and less-privileged facts is not simple. Consider the case of a pointillist picture. Perhaps an ontological commitment is made to types of dot, but there is no ontological commitment to ‘pictures’ *per se*. Thus, facts about dots are privileged, while facts which do not refer to the dots, but to the picture, are not. Both the facts about the picture and the dots are ‘real’, in that the less-privileged picture facts are valid facts. That they are reducible to facts about dots does not alter this; the dot facts are just more encompassing, as picture facts are reducible to them, but not vice versa.

Most empirical accounts have a hierarchy of facts, with privileged facts at ‘bottom’ which support reduction, or if not, at least act as a supervenience base. The less privileged facts are ‘high level’ facts of some sort. However, ‘high level’ facts are valid; but it is the case that their status as reducible facts, or supervenient facts, can lessen their status.

The manner in which it is conveyed that some facts have less privilege varies. But there is the sense in which two goals are achieved: the less privileged facts are accepted, and they are easy to accept, as they are less privileged. An example is the physicalism of Pettit, he describes ‘mental facts’ as facts which ‘come for free’. The degree to which this is coherent is the degree to which it is coherent to say that brush-stroke facts are privileged facts, and picture-facts ‘come for free’.

Yet, what does this mean? If the less privileged facts are reducible, it means that they are encompassed by the privileged facts. If the relationship is left at supervenience, then this may or may not be the case. How are facts about pictures, or mental facts, any less real than facts about dots or physical facts? Facts about pictures are more real to us than facts about dots, as are mental facts over physical facts.

The way in which certain facts are less real privileged rests on the completeness view of the privileged facts. Since the privileged facts explain everything, other ways of describing things, other facts, are reducible, and could be described as 'nothing but' facts. If these other facts were not reducible, they would not be 'nothing over and above' facts, and there would be modal concerns, as is argued in the case of Mary and her 'facts' about redness.

It is completeness that renders certain facts 'nothing but', or 'nothing over and above', or 'comes for free'. If completeness of privileged facts holds, then other facts which are not 'nothing but' cause modal concerns, as they are futher world fixers. Completeness means that other facts, of whatever form, cannot impinge on this completeness, and must not be actual 'facts' in the sense that the privileged facts are genuine facts.

These less-privileged facts must be necessarily related to the privileged facts. If the privileged facts can obtain in the absence of the less-privileged facts, those facts are not 'nothing but', or 'nothing over and above' facts. They would further specify the world, with respect to the privileged facts.

Consider a supervenience relation between two sets of facts, where one set is deemed 'privileged', as it is the supervenience base for the other set of facts. Consider the case where the supervenience relationship is necessary. In such a case, if one set of facts fails to obtain then so does the other. In the case of a division between 'physical' and 'mental' facts, the particular 'physical' facts upon which the 'mental' facts supervene cannot obtain in the absence of the 'mental' facts. If the 'physical facts' necessarily entail the 'mental' facts, then it is necessarily the case that an absence of the mental facts entails the absence of the associated physical facts.

This is not to say that general 'physical' facts are dependent on general 'mental' facts; it is not to say that there has to be mental facts along with physical facts always; it is a statement of a specific case. It could be said that the supervenience relationship is non-directional. The 'physical' facts supervene on the 'mental' facts as much as the 'mental' facts supervene on the 'physical', in that case. The supervenience relationship is necessary. However, one set of facts is assigned status because they form supervenience base, and that gives those facts privilege over and above the supervenient facts. This assigning 'status' to one set of facts over the other is, in this case, not required. However, it is this 'status' that provides for a sense that certain sets of facts are less 'real' than other sets of facts.

This implied asymmetry is part of the definition of reduction. Reduction is a necessary relationship, and reducible facts are deemed to have less 'status' than the facts to which they are reduced. Yet, the status of reducible facts is still not a trivial issue, as "there is no way of keeping the dots unchanged without keeping the shapes unchanged, no way of changing the shapes without changing the dots" Pettit (1994, 254). The relationship is necessary, and thus the dependency is symmetric, even if one set of facts is seen as more 'fundamental' or encompassing.

In the supervenience relationship, however, the hierarchical status of facts is not necessary. Neither is it necessary in the identity relationship, that 'mental' facts may be identical to 'physical' facts has no asymmetry. In the specific case of functional and mental facts, McGinn has argued that the implied asymmetry in functionalist views is disingenuous because it lessens the status of the 'mental' facts (McGinn 1991). Identity theorists, who may be realist about mental phenomena, have somewhat an easier time than property dualists, such as Chalmers. This is because the status of the mental in identity views is 'just identical' to the physical, and so does not have the same explicitness of Chalmers' dualist phenomenal properties. But the identity relation does not make the mental any more 'just physical' than it makes the physical facts 'just mental' in a specific instance. Pain is identical to neural firing, which is to say that neural firing is identical to pain, unless there is an implied asymmetry. Such an asymmetry may state the identity relationship always in one direction: pain is identical to neural firing.

The status of two sets of necessarily related facts is not a simple one, be that relationship one of reduction, identity, or supervenience. Yet, there are degrees of lesser status given to sets of facts in these cases. Where the relationship between two sets of facts is contingent, then there is more than an implied asymmetry. There are supervenience base facts, and supervenient facts, and the former may obtain independently of the latter, at least in a logically possible sense. Yet, in this case, the supervenient facts have considerable status, as they are further modal specifiers with respect to the 'privileged' facts upon which they supervene. Thus, the asymmetry does not make the supervenient facts 'lesser' facts.

The concept of 'status' of sets of facts all rests on the concept of a privileged and complete fundamental set of facts with its attendant fundamental set of ontological commitments. These are seen to specify the world exactly. So facts are either 'nothing but' if they are reducible, 'come for free' if they are related via identity, and if they are non-reducible, they are supervenient. If the relationship is contingent, then they describe a metaphysically distinct ontological category.

Where terms such as 'physical' facts or 'mental' facts are used to refer to all such facts within a world, then asymmetry does arise. However, where specific instances of 'physical' facts and 'mental' facts are considered, there is no asymmetry. There is no asymmetry in facts about tables and facts about four pieces of wood, a flat board, and some nails. Because a world with pieces of wood and nails may not contain tables, is not to say that 'table' facts are 'lesser'. Similarly, it is not appropriate to say that 'mental' facts are 'lesser' because there could be 'physical' facts without 'mental' facts. Largely, the set of privileged ontological commitments drives the asymmetry. In the table analogy, 'wood' and 'nails' would be privileged ontological kinds, whereas 'table' is given no ontological status, being 'nothing over and above' 'wood' and 'nails'. In that case, if there were no table, there would not be that particular instance of 'wood' and 'nails'. And if is true to say that the ‘table’ was no more than ‘wood’ and ‘nails’ then the specific instance of ‘wood’ and ‘nails’ is no more than a ‘table’.

This ‘lessening’ of the status of fact is common. Pettit claims that ‘mental’ facts are “nothing over and above” ‘physical’ facts, that they ‘come for free’. Just as this table is “nothing over and above wood and nails”, and “comes for free”.

Daly suggests that perhaps there is the sense in which two sets of facts, if necessarily related, must not be distinct, and so ‘mental’ facts must be ‘nothing over and above’ physical facts (Daly 1995). The attempt to ‘lessen’ a set of facts that are only contingently related to another set is not as common. The contingency keeps them distinct, and being distinct, the status of one set is not ‘lessened’.

5.3.2 The status of ‘bridge laws’

Where there is argued to be an epistemological explanatory gap, as there is in the case of physical facts and mental facts, the situation is different. If the mental facts were truly ‘nothing but’, and reducible to, the physical facts, there would be no need to explicitly state that mental facts are necessarily related to physical facts. However, as there is an epistemic gap, this needs to do be stated. However, there are those who consider the nature of such statements as “the mental facts are necessarily related to the physical facts”. Perhaps that statement can be treated as a ‘fact’; then one can ask whether it is a physical fact.

In the case of a reduction relationship, then the statement of necessary relation does not seem itself to be a fact; there seems no need for bridge-laws between certain facts and the facts to which they reduce. However, if there is a necessary, but non-reductive relationship, then there is credence to considering that the necessary relationship is itself a bridge law, or a fact of some sort. Horgan deals with this issue of bridge laws, and treats them as metaphysically necessary facts in themselves, and in the case of the relationship between the mental and physical, would not consider the bridge-law a physical fact (Horgan 1978).

Consider the relation between the physical and mental as a ‘bridge law’. There are reasons for not treating ‘bridge laws’ as physical laws or facts. If that bridge law were a physical fact, then a physical fact would be expressing something that is not encompassed by the physical facts. Thus, facts such as, ‘the mental supervenes on the physical’, is not a physical fact, if it is considered a ‘fact’ at all. Whether or not such bridge laws are ‘physical’ facts is a debated issue. But it does seem that, whatever the status of such bridge laws, they cannot be expressed in the terminology of physics. These are separate issues: whether bridge laws are physical facts, or merely inexpressible in physics terminology. In Pettit’s physicalism, however, this separation is not much of an issue, as “the empirical world contains just what a true complete physics would say it contains” (Pettit 1993, 222). Would a ‘true and complete physics’ claim that (a) there are mental facts, and (b), that there are bridge-laws (or relations) between the mental and physical facts? If the existence of mental facts is not expressible within the terminology used in physical facts, then to what extent is it meaningful to argue that mental facts 'come for free' given the physical facts?

Physicalism may claim that the relationship between mental and physical facts is not one of law-like connection. Thus, the difficultly of non-physical bridge-laws, or bridge-laws which if considered physical facts, apparently refer to non-physical facts, is avoided. Crane argued, based on the concept of a 'true and complete physics', that bridge laws were necessary, as the physical facts alone would not be enough to fix the mental facts (Crane and Mellor 1990). These bridge-laws could not be physical facts; therefore, physicalism is false. However, if there are no law-like connections, then this argument is avoided.

A non law-like connection, which is essentially removing the status of 'fact' from statements of 'bridge-laws', does solve certain difficulties. There seems no need to relate picture facts to dot facts in a way that needs to invoke 'laws'. Here, however, the situation becomes increasingly complex. It can be argued that in the dot/picture case there are hidden assumptions of the form: if certain dot-configurations obtain, certain picture-configurations obtain. Does this mean that there is some status of 'fact' given to the systematic connection between dots and pictures?

The mental/physical case is somewhat different, in that it does seem that without statements of relation between mental and physical facts, the physical facts themselves would not be seen to fix the mental facts. There is a connection, a relation, be that given status as 'law' or not. Thus, there is an issue to address; namely, where does this non-law, or systematic connection, come from, and why does it obtain? Crane's argument against physicalism, on the basis of non-physical bridge-laws, has been argued by Daly to stand, even if these bridge laws are no more than systematic connections, because such connections, even if not 'laws', must be assumed.

In the case of a contingent relation between the mental and the physical, however, the status of bridge laws cannot be lessened to mere systematic connection, as in the case of dots and pictures. They are facts, and the are non-physical facts, because they relate two realms, which are contingently connected.

5.3.3 The Implications for some issues

The important implication of a view such as this is simply that there is no credence given to 'lesser' facts. There is no separation into a hierarchy of levels, with 'fundamental' facts at bottom, and 'merely derived', 'nothing but', or 'come for free' facts floating above them. And since there is no splitting of levels, there are no fixed 'laws' that relate one level to another.

That is not to say that two sets of facts are not related in some way. They may even be related by reduction, with the caveat that the 'reduced' facts are not 'lesser'. Picture facts may be encompassed by dot facts, but there are still picture facts. The only necessarly relationship is one of co-reference. Dot facts and picture facts refer to the one supposed object. But as there is no stratification into levels, it is not appropriate to view two sets of facts as independent, with a 'bridge law' which relates them.

The equality of facts, and the avoidance of a hierarchy of levels, and the avoidance of the concept of 'lesser facts', has one important implication for the inessentialist notion. Consider facts which are considered 'nothing but' other facts. Perhaps they supervene on other facts. Now, in the privileged hierarchical view, the 'privileged' facts are genuine facts, and have genuine causal efficacy. The supervenient facts are 'merely derived', and if even acknowledged as something more, do not have causal efficacy. They are said to have 'qua' causation; they can cause in virtue of that to which they supervene. The 'mental' does not cause, because it supervenes upon/is reduced to/is identical to the 'physical', which causes. The 'mental' causes only in virtue of the fact that the 'physical' causes, and the 'mental' is related to the 'physical.

If there is no privilege, the entire 'qua' causation debate is avoided, because there is no privileged level to which concepts of 'causation' are ascribed. Note, this view says nothing about the metaphysical facts of causation. It does state that there is no sense to 'qua' causation, or 'merely derived' causation. As there is no hierarchy and no asymmetry, there is no meaning given to 'qua' causation, because that requires asymmetry and hierarchy. There is no irreducibilist conception of causal efficacy, and thus there is no sense in which properties at other levels can only count as causal in a derived sense.

A view such as this, where there is no hierarchy or privilege, can allow for genuine causality within many different accounts, without the need to find a privileged account with genuine causality. An example of a view such as this is that of Velmams (Velmans 1990). Since an empirical third person account is not considered privileged, it does not conflict with the seemingly causal efficacy of first person phenomena (Velmans 1993b),. Thus, velmans sees genuine causality within two accounts, ensuring that the mental is not 'inessential' in his view. As is mentioned here throughout, for this to be coherent, the notion of a 'complete' empirical privileged account must be dropped, and this is so in Velmans views (Velmans 1995).

5.4 Conclusion

One way to solve the difficulties of inessentialist phenomenal realism, while maintaining phenomenal realism, is to reconsider the 'inessentialist' notion. This need not require interactionism. What it does require, however, is that the notion of a complete and privileged 'base' account is dropped. What is left then, is a view which allows for a multitude of different empirical, and non-empirical, accounts, where these accounts are not placed in a strict hierarchy. Because there is no one privileged account, there are no strong reasons to accept only one set of ontological commitments. Thus, a liberal view of natural kind realism is the result. As there is no hierarchy, there is less reason to categorise accounts of causation with regard to supervenient mental states as 'qua' or 'merely derived' causation. Within an account which invokes 'mental states', they may be treated as causally efficacious. This does not entail any violation or conflict with the causal closure of one set of privileged fundamental facts, as a complete and accurate set of 'privileged' facts is not admitted as being a coherent notion.

In such a view, two sets of facts, where neither set is placed on a hierarchical scale, may still be related by reduction or supervenience. Some accounts may be reducible to other accounts, but it is not accepted that there is a 'base' account to which all accounts are reducible. There would be no explicit need to use phrases such as 'merely derived' in regard to facts, in such a view; there would be no requirement for considering certain facts as 'coming for free', for instance. Talk of the status of derived capacities, or capacities in virtue of being/supervening/related to other facts, of facts which 'come for free', of 'nothing but' facts, of 'nothing over and above' facts, would not be required. Relations between sets of facts are allowed, with the caveat that reduction, identity, or supervenience cannot be used in a way that assigns 'lesser' status to facts

This view is a pluralist view, which admits of irreducibly many properties and entities. An expression of this view comes from Bohr: "we must, in general, be prepared to accept the fact that a complete elucidation of one and the same object may require diverse points of view which defy a unique description" (Folse 1985, 179). Dupre has argued such a view (Dupre 1993) which has been defended by Daly (Daly 1996). In each case, the argument is against the assumptions of privileged facts and asymmetry, and the difficult issue of 'lesser' facts which follows (Daly 1995). Crane has argued against these two assumptions also (Crane and Mellor 1990). The view of McGinn with regard to necessary relations between sets of facts indicates that he rejects an asymmetric view in relation to sets of facts (McGinn 1991).

It is a view that is the opposite of the physicalism of Pettit, but not necessarily non-physicalist; it is compatible with non-reductive physicalism. It says there are many levels of description, but does not invoke specific 'laws' which related these levels to each other, though in specific instances, relationships can be described. Each set of facts is a set of facts in its own right. It is not 'fixed' or determined, by a set of facts at another level, thus there is no invoking of necessary bridge-laws between sets of facts.

As regards relating sets of facts, it is akin to the anomalous monist view of Davidson. However, the mental is a conceptual, not an ontological category for Davidson, whereas it is an ontological category in this view. Each set of facts about one seeming thing is related in virtue that they describe one seeming thing. Thus, in such a view, there may be 'mental' and 'physical' facts. In the accounts of 'mental states' and 'physical states', it may be that a ‘mental state’ obtains only in the presence of a ‘physical state’ and vice versa. An identity theorist may claim token identity. However, there are no fixed laws relating types of ‘mental state’ to types of ‘physical state’; there is no ‘type’ relationship. The relationship could therefore be called anomalous. As for necessary co-occurrence between ‘physical states’ and ‘mental states’, this is not an identity relationship of the identity theory sort.

As to a privileged ‘bottom’ level, there is none acknowledged. As to the relation between sets of facts, the relationship is anomalous, but it is compatible with a monist view. It is a pluralist anomalous monism, a non-reductive monism.

Chapter 6

Indexical

6.1 Introduction

The experience of redness is not capturable completely in functional terms. That is what is implied by the difficulties mentioned in the third chapter. Phenomenal experience has status that cannot be eliminated in favour of, or reduced to, functional description. This supports a phenomenal realist hypothesis. Inessentialism too has difficulties. These difficulties hinge on the view that the world contains what a complete empirical account would say it contains, or that functional description captures behaviour completely. Foregoing the notion of a complete and fundamental empirical description, and the completeness of functional accounts, is what is required. This may not be acceptable; if so, eliminativism results.

The phenomenal realist premise refers to something not capturable in functional or empirical terms. Phenomenal 'feel' remains, and Nagel's what-it's-likeness remains. There is, however, the temptation to deal with an aspect of what-its-likeness in empirical terms.

The empirical side of phenomenal experience is complex. Someone looking at someone else's experiencing redness need not be experiencing redness themselves. Looking at grey-matter in the brain is an experience of greyness, even if that brain is experiencing redness. The distinction between first and third person, which is the essential defining distinction of problems in philosophy of mind, is especially evident in such an empirical endeavor.

Such an empirical view is a third person view of first person phenomena. Either the first person is eliminated in favour of or reduced to the third person, or the third person view will leave something out. The extent to which empirical questions of phenomenal experience provide useful and accurate information is the extent to which there is coherence to an empirical side to phenomenology.

The first person concept encompasses both the experience and the experiencer. The phenomenal experience is the experience that is experienced. Pain is someone's pain, and that pain is not someone else's pain. A Pain is that particular pain, and not another pain. There is a phenomenal experience of pain, and nothing apart from that phenomenal experience of pain is that phenomenal experience of pain. It is the connection between phenomenal experience to the experiencer (if they are distinct), that is the difficultly for empiricism. The first person has a connection (or is) to his or her experiences, which the third person (empiricism) does not have.

There is an indexical aspect to phenomenal experience: experience has, or is, an experiencer. The first/third person distinction, given that it has not been bridged, implies that empirical aspects of phenomenal experience are limited in some way. I wish to address one such limitation. The limitation regards the location of phenomenal experience.

6.2 Location

If you and I swapped brains, where would you be? You would be where I was, and I would be where you were. If someone removed your brain and placed it in a life support vat with a transceiver, placing a further transceiver in your head, where would you be? In the first instance, you would be where your brain is, and in the second case, you would be where your body is. So there is some flexibility. In the second case, your looking at your brain in a vat would be the same as someone else looking at that brain in a vat. It seems what 'we', the 'experiencer', are wherever we think we are. Dennett and Sandford use these examples to argue an eliminativist case (Dennett (1982) and Sandford (1982), both in (Hofstadter and Dennett 1982)). Because of the arbitrariness of the location of the experiencer, they argue, there are no further facts of 'location of experience' over and above the empirical and functional aspects of persons. The ambiguity of the placement of the subjective aspect strengthens the eliminativist case.

What Dennett and Sandford are assuming is this: if there are experiences, they must be somewhere. Specifically, if there are experiences, then they must be *empirically detectable* somewhere. They are assuming that first person experiences must have a third person empirically verifiable aspect, and that this aspect has a third person location.

Their argument is that this supposed fact of placement could vary arbitrarily, as in their brain swapping examples. The question is raised, if there are first person experiences where are they? Because they are not to be found, they are concluded against. The question they ask is if there are first person experiences where are their third person empirically detectable aspects? The aspect they considered was location. They assumed that an empirical aspect to first person phenomena would be empirically discernable location.

Their assumption, however, is incorrect. We don't have to find empirical third person aspects of first person experiences. In third person terms, first person experiences *do not have to be anywhere.* There does not have to be an empirical aspect to first person experiences. Thus, the argument that assumes there must be, and then provides examples where placement is seemingly very flexible if not arbitrary, are not eliminativist.

There is no empirical placement of experience. From the third person point of view, experiences are neither anywhere nor nowhere. They are neither in heads nor outside heads. All this states is that no empirical location can be assigned to experiences. The empirical question of location cannot answer the question of where experiences are. There are a number of reasons for this.

Experiences have a spatial and temporal location, of varying precision and duration, from the first person viewpoint. Since the first person cannot be ignored, there is an *a priori* reason to accept, for now, that the when and where of phenomenal experience is valid issue. It can only be a valid issue, however, relative to the experiencer (or the experience, if experiencer and experience are conflated). Thus, *a priori*, we can only say that experiences have a location from the first person, for now. The third person location is an additional issue.

To locate the experiences in a third person manner is to locate the experience relative to someone else who is not the experience/experiencer. Such a task may be called the third person placement of first person phenomena. It involves a 'perspectival switch' between the third person and first person aspect. If this switch is not made explicit, it is easy to argue for difficulties and arbitrariness in the 'placement' of experience.

One method of dealing with the location of experience is to declare that experiences are where they seem to be. However, this does not say that empirically, experiences are where they seem to be, as stating that experiences are where they seem to be does not involve a third person perspectival switch; it merely states that experiences are experienced as being somewhere. This is all first person. This is not a statement of direct realism: there is no empirical third person fact impliced.

If it is stated that no empirical question of location can be asked, then there is no authority upon which to contradict the statement that experiences are where they seem to be. Velmans has argued at length for a view such as this (see (Velmans 1993a) and (Velmans 1991)).

It is easy to mistakenly read "experiences are where they seem to be" as involving a perspectival switch. If it is read in this way, then it can be countered simply by referring to the brain in a vat scenario, or commenting on virtual reality systems. Yet the statement does not involve a perspectival switch. It says that first person experiences are, from the first person, where they seem to be. Redness of roses is out there, in the world, from the first person. There is no third person claim, either implicitly or explicitly, in this statement.

Consider a neuroscientist attempting to locate my experience of redness when I look upon a red rose. The task is this: someone, who is not having, or is not a particular given experience, is to locate the experience of someone else who is having, or is, a particular given experience. When I see a red rose, it is out there in the world. I am the experiencer/experience, and that is where it is experienced to me, the experiencer/experience.

The redness of the red rose can only be in the world to those who experience the redness of a red rose. This is where experience lies, to the experiencer. The ‘location’ of the experience, from the first person point of view, is ‘in the world’.

There is no situation in which someone else can alter the location of my experience by appealing to authority, empirical or otherwise. I am taking it as being evident that no amount of *a posteriori* knowledge will change the first person location of experiences. There is no authority that can convince me that the experience of redness of roses is *not* ‘out there’. Still, the neuroscientist is looking at my brain, attempting to ‘locate’ experience. What meaning can this be given? She cannot share my experience, since she is not that experiencer/experience. And if she could, she would agree that the redness is out there, in the world. She could have a similar experience (by looking on the same rose), but she would also agree with me as to the location of the redness.

Dennett asked ‘where is experience?’ In first person terms, he said it is wherever we think it is. He then conflates the first and third person by using an implicit perspectival switch. Then he argues that ‘experience’ cannot be anywhere, and thus eliminativism ensues, because of the assumption that it must be empirically discernable as being somewhere if eliminativism is false.

There is no reason why anyone could declare that the redness experience is actually ‘in my head’ or anywhere apart from where the experiencer/experience itself reports it to be. Any such attempt at an answer is an appeal to third person authority. It can either override the first person or concur with the first person. But it is unlikely that third person explanation would deem redness in the world, since the third person places the mechanism of the arising of experience as dependent on the head.

It is false that my experiencing of a red rose is ‘in my head’ or anywhere else apart from out there, where the rose is, and the rose is in the world. I would be delusional if I claimed otherwise. This is my first person point of view. I am saying that redness is out there, in the world, to me. I am not saying that there is redness out there in the world in a third person context divorced from my point of view.

Consider an impossible situation, a situation that is incoherent and could never occur. The neuroscientist locates the redness experience that I am having of the red rose. The ‘redness’ of the red rose is experienced as being out there, in the world, from my point of view. She finds this experience in a particular region of my brain. This is a momentous event: the seat of consciousness found! She tells me, the experience of redness is here, in region 5 of the brain. This would be an important epistemological advance[1]. My experience of the red rose out there in the world actually turns out to be somewhere else. Yet, I would still see the redness of the red rose in the world. So now there are two answers: the neuroscientist has located the redness experience in my head, and I

[1] there is an ultimate contradiction in having complete authority of first person states given to third person empiricism, and it is nicely described in (Smullyan 1982).

experience the redness experience in the world. There are two different answers and they conflict. Which one is correct? Which answer is more privileged? If experiences could be located in a third person sense, there would be no answers forthcoming, only conflict.

6.3 The location of correlates

Emprical methods do not try to locate phenomenal experience. They attempt to find third person empirical correlates of experiences which are not directly empirically measured. Thus, empirical methods find empirical correlates of experiences via empirical events that are indicative of experiences. Empirical methods may correlate neural events with experiences via the report of experiences, (which are empirically discernable) for instance.

The question can be asked, "what areas of the brain, when active, bear a direct correlation with phenomenal experience, as reported from the first person?" This is a valid question, but it is not the question "what are the areas of the brain in which phenomenal experience is created/located". A lot of what goes on in our heads (as described by neuroscientists) does not bear a direct correlation to our phenomenal awareness. There is activity in my head of which there are no strong experiential correlates. Some of the processes in our heads do correlate with phenomenal experiences in a stronger fashion. There are areas which, if active, are accompanied by experiences in a strongly correlated maner. This is not to say that phenomenal experience is located in the latter areas, and not the former.

There is a tendency when discussing, or experimentally searching for correlates to enter into confusions over location. This results from the temptation to draw a line between 'unconscious' and 'conscious' processes, between the processes that bear a strong correlation to phenomenal experience and those that do not. The question "Are we aware of visual activity in V1?" is a question which can be read as having a perspectival switch: it is an ambiguous question. "Do we have phenomenal experiences in strong correlation to activity in V1" is less ambiguous.

We are not aware of any neural activity in our heads, we are aware of experiences. We do not experience neural firing as seen from an empirical viewpoint. We are aware of phenomenal experiences, which may be correlated, strongly, with neural activity. In the pragmatic search for correlates, these vagaries in language (which seem to be part and parcel of publications in the neural correlates field) make no practical difference, but they can lead to philosophical confusion. There may be the tendency to deem all processing one side of a particular correlation threshold as 'conscious', and the rest as 'unconscious', thus pushing the seat of consciousness further into the head[2].

[2] See (Crick and Koch 1995; Kolb and Bruan 1995) and (Crick 1994) for overview of the experimental work of finding 'cleaner' correlates of phenomenal visual experience further and further 'up' the neural

Our brains and bodies have activity that correlates with phenomenal awareness. I am not aware of any activity that a third person may empirically find. I may experience pain when a pin pricks my hand, but I am not aware of pain receptors firing, I am aware of pain. I am aware of pain in my hand, relative to me. The third person cannot locate that pain in my hand or anywhere else. Neither can I locate that pain anywhere else in third person terms.

There are understandable reasons why correlates of phenomenal experience may tend to be seen as the seats or locations of phenomenal experience. This is just an example of the retreat of phenomenal experience into the head. That direct realism is untenable suggests that there are no phenomenal properties in the external world. Our understanding of scientific ontologies means that roses no longer have the property of redness (in the sense of the experiential property of redness). But this does not mean we can say that experienced redness is not out there, because it is, from the first person. If it is assumed that experiences must have a third person location, then we are forced to push them back into the head.

Views, which are realist about the mental, have varying degrees of bias towards this issue of location. Wide supervenience relations do not have the problem of redness in the world, as the experience of redness can supervene widely, and on more than just the head. Narrow supervenience relations (which focus on the head), would, however, place experiences as in the head, if the third person placement of experience was considered coherent. Identity theories, which equate phenomenal experiences with neural firing in the head, also have this difficulty if third person placement is sought. However, since third person placement is not coherent, the relations which are 'narrow' do not have this problem. Statements of relation between phenomenal properties and neural properties cannot be read as making a claim towards the third person placement of experience.

This has relevance the bridging of the explanatory gap. Third person views, currently, cannot explain why certain activity in the brain 'painful' while other activity isn't anything at all, in the phenomenal sense, because that would be akin to locating phenomenal experiences. We can, however, correlate neural activity with 'painfulness'.

There may be a location bias, if one bridged the gap between certain neural firing and pain. The bridge, therefore, must not allude to placement. From the third person point of view, there are observations, which are correlated with first person reports. The first person cannot make third person claims about the location of their experience, and neither can the third person make claims about the location of the first person experience. No philosopher or neuroscientist ever has felt the pain of a blow to the big toe anywhere else but the big toe, regardless of the strong correlation between that experience of pain and firing neurons.

processing chain.

6.4 Measuring place and time

The answer to the question of where and when experiences occur is not answerable in third person terms. The first person has authority, and so visual experiences are usually in the world. This does not say anything as regards perceptual realism. Visual experiences are in the world in a first person sense only; there is no mention of the third person. The first person location of experiences does not entail anything as regards supposed third person placement of those experiences. There is no conflict. One can accept both the first person authority on the location of experiences, and yet not worry about experiences being 'in the world', since they are neither anywhere, nowhere, nor everywhere, from the third person.

This is not to say that they have a location, but it cannot be empirically determined. The attempt to find them from the third person is itself incoherent. The first person aspect of phenomenal experiences does not have a third person aspect. The third person aspect it does have is no more than correlates, as empirical methods can provide no more than that.

The third person process whereby an assignment of location (both where and when) is given is not absolute. Empirical methods do not actually give a single spatial or temporal location. They provide an offset read against some chosen spatial or temporal basis.

In order for a spatial assignment to be given, a coordinate system is required. In third person terms, the world, for the purpose of spatial or temporal assignment, is (currently) considered in an atemporal manner, as a vast realm of abstract readings from standard measuring devices. These abstract ideal devices provide the basis for accepting the spatial and temporal assignment provided by concrete devices.

The standard measuring devices can be thought of as rulers and clocks. All assignments, all locations, in third person terms, are descriptions of relationships between different sets of rulers and clocks. Rulers and clocks can be organised in different ways to provide different coordinate systems. The measurements have to be in terms of relationships between rulers and clocks since the measurements on a single ruler and clock are arbitrary[3].

A solitary spatial or temporal index is useless. The assignment is always a relation.

[3] This is the question of sameness, of similarly. Why can we treat a free-fall inertial reference frame as 'ideal'? Where does the sameness of whatever it is that ensures that there is a standard between clocks and rulers in differing locations? There is an underlying assumption of sameness. It is possible that addressing this assumption directly may be seen to require an ontological commitment to something that provides this sameness, and this is avoided because of 'ether' fears. Interestingly, Einstein would sometimes intentionally cause a stir when we would address the question of this assumption and its potential ontological aspects by using the term 'ether': "According to the general theory of relativity space without ether is unthinkable; for in such space there not only would be no propagation of light, but also no possibility of existence for standards of space and time (measuringrods and clocks), nor therefore any spacetime intervals in the physical sense"; from an address entitled "Ether and the Theory of Relativity", delivered on May 5th, 1920, in the University of Leyden.

There is the selection of one particular ruler and clock which provides the relationship to other rulers and clocks, and that selection depends on a vast background context within which empirical questions are asked[4].

This does not mean that there is no fixed coherence to particular time and place measurements. The third person view is atemporal, but not acausal. The particular indexical assignments to place and time are arbitrary, but their relation is not. That there is no fixed assignment of place and time to particular events does not mean they are not ordered. The ordering of events is the only meaning that the term 'causal' has in this empirical sense; it refers solely to the ordering of events.

This ordering (a 'causal' ordering) of events means that there is an ordering of events which is a fixed matter for all empirical frames of reference. There are events that are antecedent to other events, and this is a fixed fact. Because it is a fixed fact, the ordering of the arbitrary assignment of a time and place indexicals will be the same, independently of the frame of reference or the co-ordinate system chosen. The ordering of such events is a matter that is agreed for all third persons involved in empirical measurement. With a fixed sequence of events, the term 'causal' can be invoked in the sense of claiming that a particular event may have a causal antecedent in another event, or group of events. But there is nothing further in the definition of 'causal' in this context[5].

The ordering of events in these cases is something within the complete agreement of any third person outlook is a matter of ordering, only. The actual indic(es assigned to these events will vary. It is just that if the assignment of indices is coherent, the sequence of indices will be consistent with all other possible choice of indices; there will be no disagreement as to ordering.

There is no sense, however, in which there is third person agreement as to actual indexical assignment. There is no agreement as to when and where events, even causally ordered, took place. It is a matter of agreement that certain events are ordered in a certain way. But there is no possibility of an agreement as to when or where beyond this.

[4] This is what is meant by relativity. To speak vaguely, time is not relative, duration is; and place is not relative, but distance is. Actually, 'time' and 'place' have no actual third person meaning aside from relational measurements that can be assigned

[5] There is no agreed account of, or meaning which can be attributed to 'causal' beyond an event casting a 'shadow' into the past ('light cones') to encompass the events which could have been its causal antecedents. In deterministic systems there is no meaning to causal at all. The future can be seen to cause the past in as much as any particular instant determines all other instances, future and past. As for true non deterministic systems (aside from probabilstic or pseudo random systems) what does 'non deterministic' mean? Does it mean that there are uncaused events, 'first' causes? "All philosophers, of every school, imagine that causation is one of the fundamental axioms or postulates of science, yet, oddly enough, in the advanced sciences such as gravitational astronomy, the word "cause" never occurs. ...It seems to me...that the reason physics has ceased to look for causes is that, in fact, there are no such things. The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm" (Russell 1963, 132).

Empirical measurement of location provides arbitrary indices, and can only find agreement within empirical endeavors as to the ordering of events.

There are events that are not so ordered. In such a case, there is no fixed causal ordering to such events. In such a case, there is no possible agreement whatsoever between empirical accounts. All assignments of time and place indexicals will neither agree specifically, nor agree as to general ordering of these events.

A fixed, privileged or fundamental basis for relative measurement would allow the question of 'when' and 'where' to be answered in the sense of providing a specific location. Such a specific location would be a complete answer of the 'location' question, in that all empirical accounts would agree. If something akin to God-given time, or a formal 'center' to the world was accepted, then there would be sufficient meaning in 'place' and 'time' in the third person sense to begin to attempt to combine this with first person reports of experience location. With God-time, or a center to the universe, there could potentially be empirical agreement of placement. This 'center' would need to be both spatial and temporal: a center of space and a center of time.

The third person placement of experience rests upon the choice of a basis for measurement. As with all empirical third person endeavors, this ultimately rests upon empirical observation. To bring in actual measurements from clocks and rulers is to introduce the concept of observers who can read these various measurement devices. Each possible observation is labeled by readings on rulers and clocks, and these identify where that observation occurred, as specified in a coordinate system, and as reported by a particular observer. The readings are relative, in that the when and where of an event is considered by comparing readings on clocks at that event, and at the observation point. Without invoking observers, there is still the choice of a basis from which to base relational measurements to other rulers and clocks.

With those details out of the way, the difference between the sense of location in the first and third person sense can be simply stated. The first person placement of phenomenal experience is about a specific given place and time. The third person placement process, however, answers the location question in terms of 'distance' and 'duration': offsets from an agreed arbitrary indexical assignment. Distance is related to place, and duration to time, but they are not the same. Distance and duration need a place and time from which they are measured, and in empirical accounts, the choice of this basis is arbitrary.

Our experiences have a location, not a distance-offset, though they can have a distance, as in spatial extension, but this extension will be fixed to us. Distance in the third person sense, is not the fixed experience of particular distance from the first person. Our experiences seem to occur at particular times, though they have temporal extension. Duration, in third person terms, is dissimilar to first person experience of duration in a like manner.

The above does not say that there is no place/time in third person terms. It says that empirical methods answer that question in terms quite different to our experience of place, time, distance and duration. It does not say that empirical events do not have a specific non-relational place or time. It says only that the answers to the location question are not given in those terms.

There is, therefore, a difference in the treatment of the location issue in the first and third person cases. Our first person experience of place and time, as in an experience of redness being in a specific place, happening 'now' (or at a specific time), are simply statements which cannot be given a meaning in empirical terms. It is not the fault of empiricism, therefore, that it cannot answer these questions.

6.5 Finding time

Considering the temporal location of experience makes the difficulties clearer to see. There is no essential difference in the temporal and spatial attempts at location. To contest this point is to invoke a privileged ontological 'now', a God given present moment. Showing the difficulty in the temporal case is enough to demonstrate the difficulty in both the temporal and spatial cases.

The temporal case is the question of 'when' experiences occur. This is slightly different from 'where' experiences occur, as in the latter case, a distinction can be made between the experience and the experiencer, and there is then a sense of two locations: 'I' am here, experiencing redness 'out there'. This difference, however, is not important to the point I wish to make.

What I shall discuss can be described as the temporal form of the case of pain in the toe being in the toe. I mentioned before that pain in the toe is in the toe, and there is no sense in which it can be placed anywhere else. Now it is time to consider a similar case, but one that considers not spatial, but temporal location.

Consider the timing of phenomenal experience. We can rely on first person reports and attempt to match them with a third person measurement device. Every third person temporal index is a relational measurement in the context of some coordinate system. In order to find out when an experience occurs, it is necessary to have the first person report this. At some point the test subject says, "I am experiencing this sensation now". As far as the first person is concerned, this is a report of the first persons own sense of the present moment. It is a report of an experience, which has temporal aspects. This sense of "experiencing the sensation now" is not grounded in third person terms at all. The subject need not be looking at a clock to find out when the experience occurred. It is independent of third person assigned indices and it is not a relational measurement claim.

The claim of "experiencing this sensation now" is like saying, "I am standing here". It is not a report that, alone, provides any information that has meaning in empirical terms. It is not a relational measurement. 'Here' is not any specific place, from an empirical point of view. Experience does not have a third person spatial or temporal index at all, from the first person point of view. It cannot, therefore, be related to any third person measurement apparatus directly. There needs to be further constraints on interpreting the report "experiencing this sensation now".

What is available to the experimenter is the time of the report of the experience to which an index is assigned. Empirically, there is no meaning which can be understood from "this experience is happening now", because 'now' cannot be given an index. So that aspect of the claim has to be left aside. To imbue "this experience is happening now" with third person meaning is to give 'now' a temporal index and that can only be given relative to some other report of 'now' that is considered the basis for measurement. This other 'now' is usually the report of the experimenter. The subject says, "the experience is happening now", and the experimenter looks at the clock and says, "the experience occurred then".

The choice of timing basis for each party will allow compatibility of the arbitrary choice of basis for temporal measurement. Each person will each agree that 'now' was 'then', and assign indexicals from there. They will do this with regard to arbitrary events, in the third person sense. They agree that some events will be in the 'future', and some are consigned to the past. They will agree because there can be a fixed ordering to (certain) events[6]. But there is no fixed empirical answer to the duration between events (or the distance between events). There is sequence, only. However, we have experienced duration and distance, and we agree on empirical numerical assignments.

The experienced sense of "experiencing a sensation now" is left aside in favour of the practical empirical task of assigning an index to a report of an event, which is assumed to be near-enough concurrent with that event. What is also left aside, then, is the first person sense of duration. Empirical methods provide a relational measurement between two events, and this measurement is arbitrary. But the empirical account does make the assumption that the report of the experience is near enough co-occurrent with the experience.

The timing of experiences, then, can be seen to be accurate as far as the first person report of experiences is accepted as being accurate. And from our first person point of view, we can report experiences as they occur, or very soon afterwards.

But, empirically, there is difficulty in dealing with such first person reports of experienced duration. There is no scientific, third person meaning to experienced duration. Indeed, there is no empirical meaning to duration beyond arbitrary relational assignment within a given co-ordinate system. Consider what 'rate' time goes at. From the first person, we have some experiences of this 'rate' of time. But empirically, no meaning

[6] There are events which cannot be ordered. If one accepts a speed limit of influence (and whatever 'causal' means is therefore restricted to this speed limit), then there are events distant enough spatially, and near enough temporally, that no influences could have traveled between them.

can be given to this; our only understanding is first person. Our concept of ‘rate’, third person terms, means ‘speed’. But what ‘speed’ does time go at? One second per second always, no matter how ‘short’ or ‘long’ a second because ‘second’ is an arbitrary relational measurement empirically provided. We have experiences that we learn to correlate with ‘seconds’, and it is these, and only these, which provide our understanding of the ‘duration’ of ‘seconds’.

Empirically, what could we measure the ‘rate’ of time against? This is, of course, a meaningless question, because this is not the sort of question empiricism can answer. Empiricism can order, only. This, it does not provide a ‘place’ or ‘time’, it provides a relational measurement of empirical ‘distance’ and ‘duration’ quite different from first person experiences of distance and duration.

It would seem, therefore, that reports of experienced place, time, distance, and duration are inherently difficult to deal with from the third person. Empirical third person methods cannot in any meaningful sense, ‘locate’ experiences at all, either temporally, or spatially, nor can empiricism provide a meaning to experienced duration or distance. It can provide correlates, however.

6.6 Experimental work on timing

In the context of timing experiments, Libet did interesting work of this type. Libet was concerned with the timing of experience in the sense of the temporal location of experience, an issue that was revealed as more interesting than previously imagined (Libet, Wright, Feinstein, and Pearl 1979). Earlier experiments were concerned with correlates to willed motor action, such as flexing a finger. This correlate included a temporal dimension, which was considered more interesting. An EEG trace provided the temporal correlations. Subjects were asked to perform a simple ‘freely willed’ action, such as flexing a finger. The results indicated that there was significant build up of neural potential up to a second before the behavioural act took place

Our first person experience of such a freely willed action is that the decision to flex a finger is essentially concurrent with the act itself; there is little or no subjective time delay between decision and action in such a case. An empirical question could be asked as to ‘when’ the phenomenal experience of choosing took place. This question is akin to the third person question of ‘where’ pain experience is. In the former case, the first person view is that it is wherever the damaged part is (usually not in the brain), and in the latter case, the answer is that the choosing occurred in or around the time of the behavioural event.

Libet’s experiments continued the theme of third person temporal delay. He used the somatosensory cortex as the ‘seat’ of correlates. He used electrode recording directly from the brain, rather than EEG's[7]. Libet measured the time delays between the application of a stimulus to the skin, the neural correlates in the somatosensory cortex, and the reporting of conscious awareness of the stimulus by the subjects.

The first person sensation is that the awareness of stimulation is concurrent with the application of the stimulus, whereas Libet's experiements showed that, empirically, there was from half to one second delay between stimulus and awareness of stimulus. Variations on the experiments included dispensing with direct skin stimulation, and stimulating the somatosensory cortex directly.

With the empirically measured delay, it was possible, working within that half to one second window, to 'mask' stimulation of the skin by directly stimulating the cortex (Libet, Wright, Feinstein, and Pearl 1992). Thus, empirically, one could interfere with the experience of a sensation after the physical conditions that would initiate the experience occurred. This lead to what is called 'backwards referral' (Libet, Wright, Feinstein, and Pearl 1979) as the skin stimulus was later than cortical stimulus, but was masked by that stimulus. From the first person point of view, the sensation of the skin stimulus is concurrent with the sight of that stimulus being applied. From an empirical point of view, however, there is a delay. Thus, there is a different timing and ordering of experiences to that which empiricism would suggest (Libet 1981).

These experiments lead to other interesting experiments, some of which are discussed by Dennett (Dennett 1991), including the one I will describe. Because of the difference between empirical timing and first person experience, the empirical timing sequence can be exploited. With flexing a finger, the experience of the decision is concurrent with the flexing, but the empirical timing places these events as separated by up to a second or so. One can record directly from the potential build up of the finger flex, and use that as the carousel trigger rather than the actual button pressed with a finger flex. This is essentially measuring directly from the cortex to catch the action of a finger flex. Experiments have been carried out where a subject is told that they are to press a button to make a carousel turn a certain amount, while the carousel is actually being triggered by the potential buildup in the cortex. The button, in this experiment, does nothing, though this is unknown to the subjects. The experience of the situation is that the carousel seems to 'read ones mind', as it advances *just before* one presses the button. However, since the potential buildup in the cortex is for the action of flexing a finger, the subject cannot decide to refrain from pressing the button once the carousel advances. The carousel advances *just before* the button is pressed, but not long enough before to allow the subject to change his or her mind.

In such experiments, an attempt may be made to empirically 'locate' the phenomenal

[7] It never ceases to amaze me that people actually agree to have experimenters stick needles into their brain or leave electrodes in their head after brain operations just because some researcher wants to do some rather cool experiments.

experience of 'freely willing a simple action' a second or so before the first person experience of willing that action. This would involve an implicit perspectival switch between the first and third person, and would merely reveal conflict between the first and third person cases.

Such an attempt at empirically providing the timing of experiences is akin to providing the location of experiences based on empirical work. First person: "I seem to have a pain in my toe"; third person: "actually no, you have that pain in your head, based on empirical results". First person: "I decided to flex my finger pretty much immediately before I flexed it"; third person: "no, you didn't. You actually made that decision a second or so before your finger flex, based on empirical results". In both cases, the issue is one of correlation. Nobody disagrees that pains in the toe are in the toe, regardless of what the empiricist says. The timing case is the same, and thus cannot be dealt with in a different way. However, it is dealt with differently.

The build up of potential a second or so before the behavioural action can be seen as a correlation of 'freely willing an action', just as neural activity could be seen to correlate with 'pain in the toe'. But as the first person 'pain' and the neural correlate are in different places, from the first person viewpoint, so to can the neural correlates of 'freely willing an action' and the first person sense of willing that action be in different places temporally. Location means time and space, and both the spatial and temporal indices of location can differ in the first person and third person cases, though there is a correlation between them.

There is no need to convince the first person that their perception of 'pain' is incorrect, and that it is actually in the head, not in the toe. Similarly, there is no need to convince the first person that they do not actually decide to flex their finger pretty much concurrent with flexing it, but that they made that decision a second or two ago.

The difficulty with the timing case is that many interesting things can happen; sensations can be masked, motor decisions can be empirically detected in advance of the execution of the motor action to a degree that subjects believe that machines are 'reading their minds'. This seems to indicate that there are unconscious and non-experiential decisions being made, and that our awareness just piggybacks along. But in saying this there are several jumps back and forth between the first and third person points of view. From the first person point of view, just as pains in the toe are in the toe, we make decisions to press buttons and immediately press them.

Consider the claim that we are not conscious of the decisions we make. This is based on timing experiments. It is a claim about first person experience based on third person experiments. But it is an empty claim, as we are conscious of the decisions we make: I decide to flex my finger, and it flexes. So what is the claim? The claim attempts to locate phenomenal experiences where empirical events are detected. It attempts to locate the experience of 'deciding' a second or so before the first person experience of deciding. And because this is not so, from the first person view point, the decisions are thus deemed unconscious, and our experienced ‘decisions’ are simply an awareness of decisions already made. This is simply analogous to telling someone that their pain in the toe is really in their head: it makes no sense, because there is a perspectival switch. Just as the phenomenal experience of ‘pain’ is not where empirically discernable events take place, neither is the phenomenal experience of ‘freely willed action’ where (in a temporal sense) empirically discernable events take place.

Libet’s experiments do not force us to push our phenomenal experience of willed action any further into the past—say, one second—than we experience them to be, anymore than the neural correlates of pain require us to push our phenomenal experience of pain back into the head.

Libet considered the question of whether mental events are preceded by their physical causes, as this would have relevance for the question of intentional action and our sense of will. In conclusion, Libet suggests that perhaps consciousness proceeds physical causes, and that consciousness itself may act only as a veto in certain cases. Since there is build up of potential before awareness, and backwards referral, it is not our conscious awareness of intention that is the actual intention, as awareness of intention precedes the strongest correlate of intention; given this, he reasons, consciousness is carried along, only occasionally to play its veto (Libet 1985). What we experience as the decision to flex a finger Libet would say is actually the experience of a decision already made. The actual decision we do not experience.

The question “do mental events precede their physical causes” involves a perspectical switch between the first and third person perspective. The third person works with correlates and the first with phenomenal awareness, and these, being separate, do not enter into conflict with the other. In order to answer this question, it would be necessary either for the third person to override the first person, or vice versa. The question, from the first person point of view, is answered simply in the negative. From the third person point of view it cannot be answered, although a view which refers to correlates only, and accepts first person authority, may answer it.

To answer the question “are mental events preceded by their physical causes” would involve the temporal placement of first person awareness, in a manner like placing ‘pain’ in the brain, rather than in the toe. Thus, the issue as to whether consciousness is merely a spectator of physical potentials previously instigated, with a veto role at most, or more strongly associated with physical causes is seen to involve perspectival switches which ever way it is answered.

Because it involves a perspectical switch, it is not a question that can be answered. Thus, as with neural firing in the brain, while there is pain in the toe, there must be experienced causally efficacious decisions, and empirical delays and buildups of potential. One cannot override the other: the first person cannot conflict with the empirical evidence, and neither can empiricism claim that phenomenal experience is a one second after the fact spectator. The first person experience of motor decisions and motor executions are compatible with empirical timing correlates.

The potential gain from these experiments is increased, and not diminished, by the fact that the third person location of first person experience is not a valid pursuit. Indeed, these temporal experiments reveal that the notion of first and third person timing of experiences is a vast and complex issue. And just as correlations between the first and third person need not find conflict over the spatial dimension, as in pain 'in the toe' and neural activity 'in the head', neither are there similar constraints over the temporal dimension.

There need be no reason why it is not valid to correlate neural activity with phenomenal experience even if the report of awareness of that phenomenal experience is after the neural activity.

It is the understandably pragmatic approach of assuming that the first and third person notions of timing *must agree* that leads to beliefs such as neural build up before the report of awareness must not correlate to awareness, and hence to notions that there is a problem with regard to whether or not consciousness is preceded or proceeded by its physical causes (Libet 1985), or to the belief that our pains in the toe are actually pains in the head.

This is irrelevant if eliminativism is so

The Libet timing work assumes a degree of phenomenal realism. Libet himself is a strong phenomenal realist. If, however, phenomenal experiences are eliminated, then the intricacies of backwards referral, and the difficulties of aligning third and first person facts is lessened.

These experiments try to match empirical work with experiences themselves, not merely with the report of experiences. Thus, it is assumed, to varying degrees, that there is a line between experiences and the report of experiences; that experiences are more than the report of experiences. If, however, eliminativism is so, then there are no experiences in that way, as there is no experience aside from the judgement of experience, as it is the judgement that is the experience.

Churchland has argued, within the computational context, that Libet's experiments do not cause difficulty for a computationalist view of mind. Libet's experiments do not cause difficulty because memories and judgements can be reordered. Libet's experiments do, however, cause difficulty for a Cartesian theatre model, but such a model is not one computationalists would subscribe to. Churchland argued that 'backwards referral' does not cause difficulty by showing how such reordering within a computationalist system can accommodate this referral (Churchland 1981b).

Dennett, who does not subsribe to a Cartesian theatre model, can accommodate backwards referral quite easily within his multiple drafts model (Dennett and Kinsbourne 1992). In that model, there is no ‘end point’, or line, beyond which mental events become ‘conscious’. He does not actually need an explicit re-ordering module in his account; he does not need an Orwellian or Stalinesque theatre.

Dennett’s view is explicitly eliminativist. It is the judgement of an experience that is the experience. Thus, there is no ‘experience’ without judgement (Dennett 1979). Therefore, it is meaningless to say that an experience occurred, and then a judgement was made, as there is no distinction. The judgement ‘fixes’ the experience. In such an account, judgement and memory are all that are important; when memory is, to use Dennett’s term, ‘probed’, or judgements made, the experience is fixed. In this context Dennett has argued that dreams may be a case in which there is no ‘probe’ until after the dream (Dennett 1976). Thus, dreams are not ‘real’ experiences at all, they are memories. However, this just serves to point out that there are no ‘real’ experiences in Dennett’s view, just memories and judgements. In his account, we simply judge that we had an experience concurrent with seeing a skin stimulus a second ago; from our point of view, this will be the experience, and it will seem as if there is backwards referral.

6.7 Final Remarks

It does not mean anything in third person terms to state when and where phenomenal experiences take place. It is not a question that empirical methods can ask or answer. The aspects of ‘when’ and ‘where’ that has meaning from the first person point of view cannot be given an empirical meaning. What empiricism can provide are correlates of experience. But the locations of correlates are just the locations of correlates, not of the experiences themselves. That does not mean that the experiences are somewhere else, because empirically, they are not anywhere. The when and where have a first person meaning only. To say that pain in the toe is in the toe does not say anything about what may empirically be found, and it does not place experiences ‘out in the world’ in the empirical sense. As reports of experiences of pain in the toe, must be taken to indicate that the experience is (experienced as being) in the toe, so to for reports about the timing of experiences. It is incorrect to argue that our awareness of decisions are merely awareness of decisions unconsciously made, on the basis of an assumption that first and third person sense of timing must agree.

Empiricism is not required to move the spatial location of first person experience back into the head, neither is it required to push the temporal location of first person experience back into the past or indeed, forward towards the future.

Chapter 7

Conclusion

7.1 Concerning functionalism

It is not likely that we could have built an adding machine if we could not have done addition ourselves. If we could not have done addition in principle, it would not be effectively computable to us. But we can add; addition is effectively computable. It is not likely that we would fail to build an adding machine, if we could add ourselves. If we can do addition, it is effectively computable, and so it should be possible to describe addition algorithms, and build adding machines. Addition is effectively computable to us, and if it were impossible in principle to provide an algorithm or build an adding machine, then our brains would be a special sort of thing. Assuming adequate resources, what is computable is effectively computable, and what we can compute, other things, not necessarily persons, can compute also. This follows from the results of Turing, and the thesis of Church. It provided reason to believe that a functionalist account of mind is possible. However, it is also the case that everything can be given a functional description; and it is the case that, even if there are uncomputable processes generating maximal sequences, this cannot be verified. Thus, appropriate or not, functionalist accounts are possible for all behaviour that is empirically knowable: empirical results allow for functionalist accounts, and admit a computational description, regardless of the underlying nature of what caused or generated these empirical observations.

Abstract functions can be used as descriptions of behaving things, though these abstract descriptions may describe a variety of things. One description suits many things; and one thing can conform to many descriptions. Universal computation is built on this fact. Function (functional role) can be used as the identity criteria of mental states. But it cannot be used as an explanation of what these mental states are. Mental states may be the mental states they are in virtue of functional role, but this is not what they are specifically. Radical functionalism makes this mistake. Functional role, then, leaves something out; it leaves out what mental states are. Thus, it is to be expected that functionalism does not ‘fix’ qualia. A further constraint is required, and physicalist functionalism can provide this. Radical functionalism (and so-called ‘Strong’ artificial intelligence), cannot work. That which supports functional role is important. Any account which answers both issues, that of identity criteria, and the ontological issue of what mental states are, in abstract functional or behavioural terms, does not work.

Assuming a radical functionalist view proves the incoherence of functionalist views that do not have a physicalist, or other ontological constraint. If all aspects of our mental lives are determined by function alone, we could not, by hypothesis, know what it is that has this function. We could not make any ontological commitments. But of course, we must make some. We must at least admit that things are not abstract, whatever things are (there are no specific ontological commitments, however). Ignoring this difficulty leads to strong modal realism, which, in the context of the radical functionalist hypothesis, is epistemically unjustified. Radical functionalism is independent of ontological commitments, thus it says everything about nothing. Mental states are not determined by function alone.

7.2 Concerning inessentialism

We have descriptions and theories of the world. We could take it that a particular account (say, physics) could in principle describe and encompass all behaviour, all function, all action, and everything of causal relevance. But then, if a phenomenal realist view is held, phenomenal properties are behaviourless causeless things. Eliminativism is, however, an option.

If phenomenal realism is held, inessentialism results. Our phenomenal realist view has two parts. It has the behaviourless causeless core-epistemic part, and the talking, claiming, arguing part. They must be distinct, if inessentialism is so. Being distinct, they can disagree. Postulating extra rules which make them align and agree does not alter the fact that they are distinct. What is needed is connection, not alignment; and this connection needs to be necessary, not merely empirically necessary (necessity applied to our world). Chalmers, and other inessentialists, consider it that, in this world, these two parts align, so an extra ingredient is not needed. The alginment between these two aspects is either assumed, or argued to be the case empirically, in this world. However, there is a case noat mentioned in the literature. It is an empirically possible case of misalignment. This possibility overrides the premise of inessentialism. Therefore, either phenomenal realism, or inessentialism must be dropped. Eliminativism is one option. Alternatively, we can take it that these two parts are aligned necessarily; this overrides inessentialism. Interactionism is not the only option. One of the foundations of the inessentialist view can be revoked: that a fundamental description or account could, in principle, describe all behaviour and everything of causal relevance accurately and completely. Such a ‘complete’ view is held by some philosophers, but is unheard of in physics, having fallen out of favour this century.

Inessentialism does not arise if complete descriptions are impossible. If this route is taken, we are left with lots of descriptions. There is no longer a strong in-principle basis for declaring some descriptions as ‘merely derived’, or certain facts as ‘nothing but’ facts. And a chair is still a chair, and there is nothing ‘merely derived’ about a chair. It is not just some wood, as a picture isn’t just some dots. Mental states are not ‘nothing but’, ‘merely derived’ or ‘nothing over and above’ some other description of that which has mental states. A physicalist functionalism need not be inessentialist, if this view is taken. The contention is that, contra Pettit, the world does not contain what a single, fundamental, and in principle true and complete physics says it contains, as it is taken that there is no ‘true and complete’ physics. The condition is that the world is always more than or different to the empirical world; that the empirical world is an approximate view of the world. If this view is not taken, the reductio arguments herein allow for two other alternatives: eliminativism or interactionism.

7.3 Concluding remarks

The conclusions are, (1), radical functionalism is incoherent, and (2), inessentialism is incoherent. The resulting options are, a physicalist-functionalism which is explicitly eliminativist, an interactionist dualism, or a monist view with the characteristics of the monist view described herein. The third option has similarities to the views of McGinn and Davidson. It should be noted that the third option does not count against a physicalist-functionalist account of mental states, however.

The specific point of incoherence in the second conclusion, that inessentialism is false, concerns the contradiction inherent in divorcing the core-epistemic knowledge of phenomenal experience and the ‘judgement’ knowledge of phenomenal experience. For this reason, this conclusion is compatible with strongly eliminativist views such as Dennett’s, for example. However, Dennett’s view is more a radical functionalist, than a physicalist functionalist view, so there are difficulties. But a physicalist functionalist eliminativism is possibly the ‘cleanest’ option, as there is no argument for a phenomenal realist premise in this thesis; this was assumed in order to see what coherent view would result.

If the monist view—without the concept of a ‘complete’ account—is taken, a physicalist functionalist view is possible, but it will not be complete. Specifically, the functionalist identity criteria of mental states will not be complete. Thus, what makes a mental state the state it is, is not merely functional role. Neither is a particular mental state just a physical state. Both these statements follow from removing the concept of a ‘complete’ account. The difficulties with this are that the issues of the identity of mental states, and the issue of what they specifically are, cannot be treated as distinct. Accounts of what makes a mental state the state it is may crosscut accounts of what mental states are. With regard to phenomenal experience specifically, the identity criteria may need to refer to ontological aspects. For instance, part of what makes a mental state of experienced greenness the state that it is, is that it is an experience of greeness. Similarly, what a mental state is specifically, may need to refer to functional or behavioural aspects. The incoherence of inessentialism entails that the ontological facts of phenomenal experience crosscut the behavioural and functional facts. This all results from the lack of an in-principle complete fundamental 'base-level' account upon which all other accounts supervene or to which they are reduced. Though there may be many differing accounts, as there is no 'base-level' account, these differing accounts need not have any relation. Relations are possible in specific instances: particular accounts may be reducible to other accounts. There is no entailment of psycho-physical laws between accounts of physical and mental phenomena.

In summary, the conclusion is that there is no complete description of ones coffee-tasting action; and thus, no complete description of coffee-tasting action which does not refer to the taste of coffee. Behaviour and function are not independent of ontology. This thesis can be seen as an argument for descriptive pluralism. The descriptions are dependent on one another as they co-refer, but none are privileged, and none irrelevant. There are no specific ontological commitments as conclusion. There is, however, an assumption of phenomenal realism, and it is shown that in the context of radical functionalism, eliminativism with regard to qualia does not work. This does not rule against a physicalist functionalist eliminativist view, however.

Bibliography

Armstrong, D. M. (1968). *A Materialist Theory of Mind.* Routledge and Kegan Paul.

Armstrong, D. M. (1982). Metaphysics and supervenience. *Critica 42*, 3–17.

Averill, E. W. (1990). Functionalism, the absent qualia objection, and eliminativism. *Southern Journal of Philosophy 28*, 449–467.

Baker, L. R. (1993). Metaphysics and mental causation. In *Mental Causation.* Oxford University Press.

Barrow, J. D. and F. J. Tipler (1996). *The Anthropic Cosmological Principle.* Oxford: Oxford University Press.

Bisiach, E. (1988). The (haunted) brain and consciousness. In *Consciousness in Contemporary Science.* Oxford University Press.

Block, N. (1981). Psychologism and behaviourism. *Philosophical Review 90*, 5–43.

Block, N. (1990). Consciousness and accessibility. *Behavioural and Brain Sciences 13*, 596–98.

Block, N. and J. A. Fodor (1972). What psychological states are not. *Philosophical Review 81*, 159–81.

Boswell, J. (1791). *The Life of Samuel Johnson.* Harmondsworth: Penguin books reprint, 1986. (C Hibbert ed.).

Burge, T. (1986). Individualism and psychology. *Philosophical Review 95*, 3–45.

Burge, T. (1991). Individualism and the mental. In *The Nature of Mind.* New York: Oxford University Press.

Cam, P. (1985). Phenomenology and speech dispositions. *Philosophical Studies 47*, 357–68.

Campbell, C. (1970). *Body and Mind.* Doubleday.

Chaitin, G. J. (1988, July). Randomness in arithmetic. *Scientific American*, 80–85.

Chaitin, G. J. (1995). A new version of algorithmic information theory. *Complexity 1*(4), 55–59.

Chalmers, D. (1996a). *The Conscious Mind: In Search of a Fundamental Theory.* Oxford University Press.

Chalmers, D. (1996b). Does a rock implement every finite-state automaton? *Synthese 108*, 309–33.

Chalmers, D. J. (1994). On implementing a computation. *Minds and Machines 4*, 391–402.

Chatin, G. J. (1987). *Algorithmic Information Theory.* Cambridge University Press.

Church, A. (1936). An unsolvable problem of elementary number theory. *American journal of mathematics 58*, 345–363. (Reprinted in Davis, 1965).

Church, A. (1941). The calculi of lambda–abstraction. In *Annals of Mathematical Studies no. 6.* Princeton University Press.

Churchland, P. M. (1981a). Eliminative materialism and the propositional attitudes. *Journal of Philosophy 78*, 67–90. Reprinted in A Neurocomputational Perspective (MIT Press, 1989).

Churchland, P. M. (1982). Is 'thinker' a natural kind? *Dialogue 21*, 223–38.

Churchland, P. M. (1989a). *A Neurocomputational Perspective.* MIT press.

Churchland, P. M. (1989b). Reduction, qualia and the direct introspection of brain states. *Journal of Philosophy 82*, 8–28.

Churchland, P. M. (1996). The rediscovery of light. *Journal of philosophy 98*, 211–28.

Churchland, P. M. and P. S. Churchland (1981). Functionalism, qualia and intentionality. *Philosophical Topics 12*, 121–32.

Churchland, P. S. (1981b). On the alleged backward referral of experience and its relevance to the mind body problem. *Philosophy of Science 48*, 165–81.

Churchland, P. S. (1983). Consciousness: The transmutation of a concept. *Pacific Philosophical Quarterly 64*, 80–95.

Cole, D. (1994). Thought and qualia. *Minds and Machines 4*, 283–302.

Cole, D. J. (1991). Artificial intelligence and personal identity. *Synthese 88*, 399–417.

Crane, T. and D. H. Mellor (1990). There is no question of physicalism. *Mind 99*, 185–206.

Crick, F. (1994). *The Astonishing Hypothesis.* New York: Scriber's.

Crick, F. and C. Koch (1995, 11 May). Are we aware of neural activity in primary visual cortex? *Nature 375*, 121–123.

Daly, C. (1995). Does physicalism need fixing? *The Philosophical Quarterly 55*(3), 135–141.

Daly, C. (1996). Defending promiscuous realism about natural–kinds. *The Philosophical Quarterly 46*(185), 496–500.

Davidson, D. (1970). Mental events. In *Experience and Theory*. Humanities Press.

Davidson, D. (1980). The material mind. In *Essays on Action and Events*. Oxford University Press.

Davis, L. (1982). Functionalism and absent qualia. *Philosophical Studies 41*, 231–239.

Dennett, D. C. (1976). Are dreams experiences? *Philosophical Review 75*, 151–71.

Dennett, D. C. (1979). On the absence of phenomenology. In *Body, Mind, and Method*. Kluwer.

Dennett, D. C. (1981). Wondering where the yellow went. *Monist 64*, 102–8.

Dennett, D. C. (1982). Where am I? In D. R. Hofstadter and D. C. Dennett (Eds.), *The Mind's I*, pp. 217–231. London: Penguin Books.

Dennett, D. C. (1988). Quining qualia. In *Consciousness in Contemporary Science*. Oxford University Press.

Dennett, D. C. (1991). *Consciousness Explained*. London: Penguin Books.

Dennett, D. C. (1995). Cog: Steps toward consciousness in robots. In T. Metzinger (Ed.), *Conscious Experience*. Ferdinand Schoningh.

Dennett, D. C. and M. Kinsbourne (1992). Time and the observer: The where and when of consciousness in the brain. *Behavioural and Brain Sciences 15*(2), 183–201.

Deutsch, D. (1997). *The Fabric of Reality*. Penguin.

Dupre, J. (1993). *The disorder of things: metaphysical foundations of the disunity of science*. cambridge: Harvard University Press.

Eddington, A. S. (1928). *The nature of the physical world*. Cambridge: Cambridge University Press.

Elitzur, A. C. (1989). Consciousness and the incompleteness of the physical explanation of behaviour. *Journal of Mind and Behaviour 10*(1), 1–20.

Feynman, R. P. (1982). Simulating physics with computers. *International Journal of Theorectical Physics 21*, 469.

Fodor, J. (1968). *Psychological Explanation*. Random House.

Fodor, J. A. (1990). Methodological solipsism as a research strategy in cognitive psychology. *Behavioural and Brain sciences 3*, 63–109.

Folse, H. J. (1985). *The Philosophy of Niels Bohr*. Amsterdam: North–Holland Physics Publishing.

Gardner, J. (1970, October). The fantastic combinations of John Conway's new solitaire game 'Life'. *Scientific American.*

Gardner, M. (1971, February). On cellular automata, self–reproduction, the garden of eden and the game 'Life'. *Scientific American.*

Gardner, M. (1979, November). The random number *omega* bids fair to hold the mysteries of the universe. *Scientific American*, 20–34.

Godel, K. (1931). On formally undecidable propositions of Principa Mathematica and related systems I. *Monatshefte fur Mathematik und Physik 38*, 173–198. (Reprinted in translation in Davis, 1965).

Hardcastle, V. G. (1993). The naturalists versus the skeptics: The debate over a scientific understanding of consciousness. *Journal of Mind and Behaviour 14*, 27–50.

Hardcastle, V. G. (1996). The why of consciousness. *Journal of Consciousness Studies 3*, 7–13.

Hofstadter, D. R. and D. C. Dennett (1982). *The Mind's I.* London: Penguin Books.

Horgan, T. (1978). Supervenient bridge laws. *Philosophy of Science 45*, 227–249.

Horgan, T. (1984). Functionalism and token identity. *Synthese 59*, 321–38.

Jackendoff, R. (1997). *Consciousness and the Computational Mind.* MIT Press.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly 32*, 127–136.

Jackson, F., R. Pargetter, and E. W. Prior (1982). Functionalism and type-type identity theories. *Philosophical Studies 42*, 209–25.

Jackson, F. and P. Pettit (1988). Functionalism and broad content. *Mind 97*, 318–400.

Kaplan, D. (1978). On the logic of demonstratives. *Journal of Philosophical Logic 8*, 81–98.

Kemp Smith, N. (1965). *Immanuel Kant: Critique of Pure Reason (translation)* (Norman Kemp Smith ed.). New York: St. Martins Press.

Kim, J. (1978). Supervenience and nomological incommensurables. *American Philosphical Quarterly 15*, 149–56.

Kim, J. (1979). Causality, identity and supervenience in the mind-body problem. *Midwest Studies in the Mind–Body Problem 4*, 31–49.

Kim, J. (1984a). Concepts of supervenience. *Philosophy and Phenomenological Research 45*, 153–76.

Kim, J. (1984b). Epipehnomenal and supervenient causation. *Midwest Studies in Philosophy 9*, 247–70.

Kim, J. (1992a). “Downward causation” in emergentism and nonreductive physicalism. In *Emergence or Reduction?: Prospects for Nonreductive Physicalism*. De Gruyter.

Kim, J. (1992b). The nonreductivist's trouble with mental causation. In *Mental causation*. Oxford University Press.

Kirk, R. (1979). From physical explicability to full-bodied materialism. *Philosophical Quarterly 29*, 229–37.

Kirk, R. (1996). Strict implication, supervenience, and physicalism. *Australasian Journal of Philosophy 74*, 244–57.

Kleene, S. C. (1936). General recursive functions of natural numbers. *Mathematische Annalen 122*, 727–742. (Reprinted in Davis, 1965).

Kolb, F. C. and J. Bruan (1995, 28 September). Blindsight in normal observers. *Nature 377*, 336–338.

Kripke, S. (1979). A puzzle about belief. In *Meaning and Use*. Boston: Reidel.

Kripke, S. A. (1972). *Naming and Necessity*. Harvard University Press.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly 64*, 354–61.

Levine, J. (1988). Absent and inverted qualia revisited. *Mind and Language 3*, 271–87.

Lewis, D. (1966). An argument for the identity theory. *Journal of Philosophy 63*, 17–25.

Lewis, D. (1980). Psychophysical and theoretical identifications. In *Readings in the Philosophy of psychology*. MIT Press.

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Basil Blackwell.

Libet, B. (1981). Timing of cerebral processes relative to concomitant conscious experience in man. In G. Adam, I. Meszarcos, and E. I. Banyai (Eds.), *Advances in Physiological Science*. Pergamon.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences 8*, 529–66.

Libet, B., E. W. Wright, B. Feinstein, and D. K. Pearl (1979). Subjective referral of the timing for a cognitive sensory experience. *Brain 102*, 193–224.

Libet, B., E. W. Wright, B. Feinstein, and D. K. Pearl (1992). Retroactive enhancement of a skin sensation by a delayed cortical stimulus in man: Evidence for a delay of a conscious sensory experience. *Consciousness and Cognition 1*, 367–75.

Lowe, E. J. (1995). There are no easy problems of consciousness. *Journal of Consciousness Studies 2*, 266–71.

Lycan, W. G. (1987). *Consciousness*. Cambridge, Mass.: MIT Press.

McGinn, C. (1981). A note on functionalism and function. *Philosophical Topics 12*, 169–70.

McGinn, C. (1989). Can we solve the mind–body problem? *Mind 98*, 349–66.

McGinn, C. (1991). Functionalism and phenomenalism: a critical note. In *The Problem of Consciousness*. Blackwell.

Moore, G. E. (1962). *Philosophical Papers*. New York: Collier Books.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review 4*, 435–50.

Nagel, T. (1979). Subjective and objective. In *Mortal Questions*. Cambridge University Press.

Nagel, T. (1986). *The View from Nowhere*. Oxford University Press.

Nelkin, N. (1989). Unconscious sensations. *Philosophical Psychology 2*, 129–41.

Niddith, P. H. (1975). *John Locke: An Essay Concerning Human Understanding* (Niddith, P H ed.). Oxford: Clarendon Press.

Pettit, P. (1993). A definition of physicalism. *Analysis 53*, 213–23.

Pettit, P. (1994). Microphysicalism without contingent micro-macro laws. *Analysis 54*, 253–57.

Post, E. (1936). Finite combinatory processes. formulation I. *Journal of Symbolic Logic 1*, 103–105. (Reprinted in Davism 1965).

Putnam, H. (1960). Minds and machines. In S. Hook (Ed.), *Dimensions of Mind*. New York University Press.

Putnam, H. (1967). The mental life of some machines. In H. Castaneda (Ed.), *Intentionalisy, Minds and Perception*. Wayne State University Press.

Putnam, H. (1975a). The meaning of ‘meaning’. In *Mind, language, and reality*. Cambridge University Press.

Putnam, H. (1975b). Philosophy and our mental life. In *Mind Language and Reality*. Cambridge University Press.

Putnam, H. (1985). Reflexive reflections. *Erkenntnis 22*, 143–153.

Quine, W. V. (1966). On mental entities. In *On the Ways of Paradox*, pp. 208–214. New York: Random House.

Rey, G. (1986). A question about consciousness. In H. Otto and J. Tuedio (Eds.), *Perspectives on Mind*. Kluwer.

Robinson, H. (1976). The mind–body problem in contemporary philosophy. *Zygon 11*, 346–360.

Rorty, R. (1965). Mind-body identity, privacy and categories. *Review of Metaphysics 19*, 22–54.

Rosenthal, D. M. (1990). The independence of consciousness and sensory quality. In *Consciousness*, pp. 329–359. Ridgeview.

Russell, B. (1905). On denoting. *Mind 14*, 479–93.

Russell, B. (1963). *Mysticism and Logic.* George Allen & Unwin.

Ryle, G. (1949). *The Concept of Mind.* London: Hutchinson.

Sandford, D. H. (1982). Where was I? In D. R. Hofstadter and D. C. Dennett (Eds.), *The Mind's I*, pp. 232–241. London: Penguin Books.

Schrödinger, E. (1958). *Mind and Matter.* Cambridge: Cambridge University Press.

Searle, J. (1990). Is the brain's mind a computer program? *Scientific American 262*, 20–5.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioural and Brain Sciences 3*, 417–424.

Searle, J. R. (1992). *The Rediscovery of the Mind.* MIT Press.

Segal, G. (1989). Seeing what is not there. *Philosophical Review 97*, 189–214.

Shoemaker, S. (1975). Functionalism and qualia. *Philosophical Studies 27*, 291–315. Reprinted in Identity, Cause, and Mind (Cambridge University Press, 1984).

Shoemaker, S. (1984). Churchland on reduction, qualia, and introspection. *Philosophy of Science Association 2*, 799–809.

Silberstein, M. and J. McGeever (1999). A taxonomy of emergence. The Philosophical Quarterly, to appear.

Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review 68*, 141–56.

Smullyan, R. M. (1982). An epistemological nightmare. In D. R. Hofstadter and D. C. Dennett (Eds.), *The Mind's I*, pp. 415–426. London: Penguin Books.

Stewart, I. (1988, 10 March). The ultimate in undecidability. *Nature* (10), 115–116.

Tienson, J. L. (1987). Brains are not conscious. *Philosophical Papers 16*, 187–93.

Tipler, F. (1989). The omega point as *eschaton*: Answers to Pannenberg's questions for scientists. *Zygon 24*, 241–42.

Toffoli, T. (1982). Physics and computation. *International Journal of Theoretical Physics 21*(3–4), 165–175.

Turing, A. M. (1936). On computable numbers with an application to the entscheidungsproblem. Volume 42, pp. 230–265. (Reprinted in Davis, 1965).

Velmans, M. (1990). Is the mind conscious, functional, or both. *Behavioural and Brain Sciences 13*(4), 629.

Velmans, M. (1991). Consciousness from a 1st–person perspective. *Behavioural and Brain Sciences 14*(4), 702–726.

Velmans, M. (1993a). Common–sense, functional theories, and knowledge of the mind. *Behavioural and Brain Sciences 16*(1), 85–86.

Velmans, M. (1993b). Consciousness, causality and complementarity. *Behavioural and Brain Sciences 16*(2), 409–416.

Velmans, M. (1995). The limits of neurophysiological models of consciousness. *Behavioural and Brain Sciences 18*(4), 702–703.

von Neumann, J. (1966). *Theory of Self–Reproducing Automata.* University of Illinois Press. Edited and finished by Burks, A W.

White, S. (1996). Curse of the qualia. *Synthese 68*, 333–68.

Wittgenstien, L. (1953). *Philosophical Investigations.* Oxford: Blackwell.