



Title	Dualism, Kripke's Modal Argument, Private Knowledge, and Other Problems
Author(s)	Malik, Luke
Citation	大阪大学, 2014, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/33848
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

*Dualism, Kripke's Modal Argument, Private
Knowledge, and Other Problems*

Luke Malik

Contents

<i>Preface</i>	<i>p.3</i>
 <i>Chapter One: Dualism</i>	
- <i>Section A: Introduction to Dualism</i>	<i>p.10</i>
- <i>Section B: Kinds of Dualism</i>	<i>p.43</i>
- <i>Section C: The Arguments for Dualism</i>	<i>p.68</i>
 <i>Chapter Two: Kripke's Argument for Dualism, Private Knowledge and Other Problems</i>	
- <i>Section D: Kripke's Argument for Dualism</i>	<i>p.95</i>
- <i>Section E: Private Knowledge and other problems</i>	<i>p.137</i>
- <i>Section F: Conclusions</i>	<i>p.237</i>
 <i>References</i>	 <i>p.248</i>

Preface

This dissertation is split into two chapters. Each chapter has three sections. The first chapter is an introduction to the topic of mind-body dualism, its context and related themes. For example, a notion of “functionality” is introduced in order to understand why the problem of consciousness is considered so “hard” and why it, therefore, tends to dualism. The topics, context and themes introduced are relevant to the discussion of the latter part of this work. However, the main function of the first chapter is expository and introductory. No substantive claims, beyond those made by others in the field are made or defended, and no defence of dualism is offered or intended on the author’s part. Indeed, the first section of the first chapter in many areas highlights animosity to dualism. This section is intended to be taken as a jumping off point into the debate surrounding dualism and its relation to other theories and theorists of mind. Notable contributors to the debate and their relations to each other are introduced. For example, Daniel Dennett, David Chalmers, David Lewis, John Searle, Joseph Levine, Saul Kripke, Avshalom Elitzur, Sir John Eccles, Benjamin Libet, Colin McGinn, George Sperling, Ned Block, Paul Feyerabend, Paul Churchland, and Richard Rorty are all introduced as contributors to the context in question. Themes covered in relation to these authors range from the animosity noted lately, to functionalism, to the “easy” and “hard” problems of mind, to the “explanatory gap”, to “qualia”, to scientific speculations and evidences for dualism, to “cognitive closure”, to the traditional terms of the debate, etc. These are some of the authors and themes that define the context in which dualism is discussed.

The second section attempts to define dualism. I utilize William Seagers’ definition of

physicalism in order to do this. That definition is used to define monism in general and dualism is defined in antithesis to monism. This definition of monism provides three dimensions. Dualism can vary in the ways it sets itself up against monism by taking a particular view along each of these dimensions. For example, denying the first of these conditions, *completeness*, that says everything is of one type, may lead to the view that considers *everything* as both mental and physical, “panpsychism”. A dualist appraisal of the second condition, which is *closure* and deals with questions of causality or law-like relatedness, leads to interactionism, parallelism or epiphenomenalism. With respect to the first, Karl Popper, Avshalom Elitzur and Sir John Eccles are mentioned; with respect to the second, Fechner, Leibniz and Spinoza; associated with the third position, Thomas Huxley, Frank Jackson, William Seagers and Benjamin Libet. Not all of these authors are, strictly speaking, dualist, but they help introduce certain positions that dualists can take up. For example, strictly speaking, Karl Popper is a pluralist, but his work introduces “interactionism”. Denying the third condition of monism, which is *resolution* in Seagers system, more commonly, reduction, is the archetypical way to establish dualism. This leads to an array of positions depending on where the denial of reduction is made. Donald Davidson, for example, denies the conceptual reduction, but not the ontological reduction, which makes his system dualistic in one sense and monist in another. John Searle is also mentioned in the same breath, here, for he denies ontological reduction on the one hand, but advocates metaphysical reduction on the other. What is so murky about all these positions and theories, from Popper to Searle, is that the extent or degree to which each position is dualist is contestable. So, for sure, there may be tension between the categorizations offered in relation to each theorist and position; John Searle, for example, would deny he had *any* dualistic leanings. I don’t see this as detrimental, though, for two

reasons: First, nothing substantive is being argued here, it's an exposition of the philosophical contexts in which dualism seems to emerge, and the whole context can become muddled and obscured by the different levels at which dualism can be perceived and denied. Second, nothing in the substantive views to follow depends on the categorizations made here. In fact, insofar as I define a dualism from assumptions held in the second chapter, it is based on a rather minimal set of assumptions, which are independently argued for there.

The third section introduces the main arguments for dualism. These arguments, in essence, challenge the third condition of physicalism, resolution in Seagers' language, reduction in common philosophical parlance. Complete denial of this possibility is needed if a real dualism is to be argued for. To this end, Frank Jackson's *Knowledge Argument* is considered first and David Chalmers' *Conceivability Argument* is introduced next. These are the big hitters. Last the *Inverted Spectrum Argument* and the *Absent Qualia* argument are briefly introduced, with respect to Sydney Shoemaker and Ned Block. These are "possibility" arguments so I consider them reducible to Kripke's modal argument; for me, the biggest hitter of all, the logic of which is introduced in section one, and the psychological aspects of which will be harnessed at the beginning of chapter two to provide an assumption upon which dualism can rest. Jackson's argument and Chalmers' argument are looked at in the most detail in this section. Again, no substantive defence is provided or intended. A version of Jackson's argument is reformulated in section two of chapter two, it is independently defended there. In this section, some of the major objections to the argument are considered only. In relation to Chalmers argument, again, no great critique or defence is provided. However, the reader should note, I will reject his

argument as incautious and gratuitous in the first section of chapter two. This is where Kripke's modal argument is explicated in greater detail, with a focus on its psychological aspect, in order to establish the possibility of a base upon which to rest the rather minimal version of dualism hypothesized in chapter two. Again, then, the last section of chapter one should not be seen as an argument for or against dualism, rather just an introduction to the arguments *for* dualism.

In chapter two the original and substantive work of this dissertation begins. The workings of Kripke's modal argument for dualism are grounded. To do this, I introduce the idea of genuine and disingenuous modal possibilities, basically the former are logical possibilities and the latter are not. The idea of a "qualitative analogue" is also introduced. We may think that we can talk about a disingenuous possibility, but really that's impossible because such states of affairs are logically impossible, and so can't be referred to. A qualitative analogue tells us what we are really talking about in such circumstances. This restores clarity to the terms we use and meanings we express. This notion will be important in what follows. In Kripke, a confluence of modality, logic and psychology come together. Kripke's views on how modality is determined are accepted. The underlying logic has already been reviewed in section one of chapter one and is accepted, too. Talk of genuine and disingenuous possibility respect this logic. The logic involves the following kinds of formula, $P \rightarrow \Box P$, $\neg P \rightarrow \Box \neg P$, etc.¹ where, for example, P is an essentialist statement of identity. However, the psychological aspects of Kripke's work

¹ Notation such as "box", \Box , and "diamond", \Diamond , are common in logic. The box stands for necessity and the diamond for possibility. Given P abbreviates a proposition, $\Box P$ stands for it is necessary that P. Other useful symbols are the "negation sign", \neg ; "implication", \rightarrow ; "and", $\&$; and "or", \vee ; following convention. Universal, (x) , and existential, $(\exists x)$, quantifiers are also used.

are more difficult to grasp so far as he says so little about them. In fact, it is even noted that some philosophers ask just why Kripke even mentioned such factors at all. However, these aspects play a major role in his argument for dualism because they allow one to understand ignorance, error and confusion in relation to one's modal judgements and related statements. As said, not much detail is given in Kripke. I fill in the gaps and, I think, this represents the most original part of this thesis. I use Stephen Kosslyn and Nelson Goodman to construct a theory of the imagination that reconstructs Kripke's thinking in terms of the theory I articulate. This theory is a pictorial theory of the imagination (but it needn't be pictorially based). It allows one to understand just why it is that Kripke's dualistic intuition cannot be ruled out—basically because no psychological error or confusion can be identified. So far as this theory is taken to be a theory of the imagination, Kripke's argument is given a strong psychological flavour. I continue, once it is established that Kripke's intuition stands, to contrast Kripke's argument with Chalmers' Conceivability Argument, that latter is rejected as incautious and gratuitous. Finally, caution is called for. I argue that Kripke's argument does not establish that dualism is true, but establishes that dualism cannot be ruled out. This allows dualism to be formalised further and opened up to other moves and strategies that might rule it out or falsify it. This is the work of section two of chapter two: I will attempt to falsify the kind of dualism generated by the result of Kripke's argument.

The next section, then, begins with a succinct and ubiquitous statement made of the intuition that Kripke thinks cannot be ruled out. Strictly speaking, this does not mean the statement is accepted as true, but it is accepted as a hypothesis. This hypothesis is called (K). Next, Thomas Nagel's idea that one can know what it's like to be another is

developed and a relation, (R), specified. The basic condition underlying (K) and (R) is also made explicit, i.e. consciousness. The assumption of consciousness is identified in a statement, (O). A minimal dualism is exemplified by this set of assumptions (O), a necessary condition; (K) a sufficient condition; and (R), the relation inspired by Nagel. These assumptions lead to absurdity when conjoined with a statement that captures a physicalist perspective on knowledge, which is represented by (P). That is, the conjunction of (O), (R), (K) and (P) tend to absurdity. Given that one of the assumptions must be dropped, I examine which the dualist must give up. (O) cannot be given up because it is necessary to dualism. (R) cannot be given up for anti-sceptical, existential, ethical, and scientific reasons—the latter are heavily reliant on recent neuroscience, for example, the work of Vittorio Gallese, and the phenomenological turn in this field, represented, for example, by Francisco Varela. Dualism needs to accept (K), or some form of (K), because dualism needs sufficient reason to think it true. This means that (P) must be abandoned by the dualist, that is, she must accept that knowledge is not determined physically. The set of assumptions for dualism that follow are (O), (R), (K) and $\neg(P)$. In order to see if this set of minimal assumptions can be ruled out, I try to derive an absurd conclusion. It seems, one can conclude, that this set of assumptions leads to a kind of *private knowledge*, and this threatens absurdity, and that might be reason enough to abandon dualism. However, taking inspiration from the work of the neuroscientists mentioned above, the conclusion to private knowledge may be averted disarming the threatened absurdity. Furthermore, the conjunction of (O), (R), (P) and $\neg(K)$, an alternative to the set of dualist conditions lately noted, tends to absurdity. It seems, then, that (O), (R), (K) and $\neg(P)$ can be legitimately accepted by the dualist. However, I continue to see if these assumptions and, therefore, dualism can be undermined by other

arguments. For example, Davidson's Anomalous Monism and Searle's Biological Naturalism are examined as alternatives. These positions, however, are found wanting. Last, causality is considered. Dualism is attacked for being either epiphenomenal or interactionist. I consider Chalmers "Paradox of Judgement", which leads him to explanatory epiphenomenalism, which in turn threatens causal epiphenomenalism, and dismiss his argument, I formulate the idea of a *quantitative analogue* in order to do this. I return to Avshalom Elitzur and examine his reasons for interactionism, but reject these, too. In terms of the choice offered by these two protagonists, the choice between epiphenomenalism and interactionism is a false choice. I generalise this conclusion by considering how causality becomes problematic for alternatives to dualism. Indeed, it often tends to *absurdity*. This leads me to argue that a Humean account of causality be adopted, i.e. a psychological one as against a substantive one, or an account of causality suggested by Lynne Rudder Baker, who advocates that causality must be subordinated to explanation, which sits quite well with my rejection of Chalmers explanatory epiphenomenalism. I conclude we haven't found reasons to rule dualism out: no psychological confusions, no descent into private knowledge, no major alternative views compel, and no problems from causality. The dualist hypothesis, thus, stands.

The last section aims to revise the arguments of chapter two and draw the paper to an end with the conclusion that dualism has not been ruled out as a theory of the subjective aspects of consciousness. A final remark or two will deal with what, if anything, has been learned about physicalism.

Chapter One: Dualism

Section A: Introduction to Dualism

Dualism has for the most part of the last century and this century with the rise of naturalism been on the back-foot. It has been criticized for its perceived commitments to a variety of notions. For example, Daniel Dennett, one of the most famous critics of dualism, talks about dualism in the following disparaging manner: ²

If I learn that somebody is an idealist, or a dualist, my initial working assumption is going to be that this person holds a **forlorn** view--since the "refutations" of idealism and dualism are well known.³

[D]ualism is **not a serious view to contend with**, but rather a cliff over which to push one's opponents. ⁴

My main objection to dualism is that it is an **unnatural** and **unnecessary** stopping point--**a way of giving up**, not a research program. That is quite enough for me.⁵

² The boldface in the following quotes is mine. It is used as a highlighter.

³ Daniel Dennett, (1992), Back From The Drawing Board, extracted <http://ase.tufts.edu/cogstud/papers/backdraw.htm> 15/11/2013

⁴ Daniel Dennett, (1978),*Brainstorms*, (Cambridge, Massachusetts: The MIT Press), p.252

⁵ Daniel Dennett, (1993a), "Living On The Edge", in *Inquiry* 36, extracted <http://cogprints.org/270/1/liveedge.htm> 15/11/2013

I diagnose dualism as a sort of **false crutch**.⁶

[I]ntrospection exposes the **vacuity** of dualistic "investigation," and denying it leaves dualism with no avenues of exploration.⁷

One can see, then, that dualism gets something of a rough press in the literature. It is worth considering exactly why dualism is held in such disregard. But in order to do this one must first consider what the primary object of contemporary dualism is. To start with I turn my attention to David Chalmers. Chalmers is a well-known advocate of the position. Chalmers famously made the distinction between the (relatively) easy problem of mind and the hard problem. According to him, the (relatively) easy problem of consciousness relates to the *functionally definable* attributes of consciousness. Something that is functionally definable is something which is defined according to its functions. For example, anatomically speaking, we might ask, what is a heart? If we study the heart we find it has a particular function, which is to pump blood around the body. This explanation of its function is sufficient to answer the question asked. Understanding the processes associated with the function will provide us with everything that we need to say about the heart in order to understand what it is. Chalmers gives the following example:

To explain the gene, for instance, we needed to specify the mechanism that stores and transmits hereditary information from one generation to the next. It turns out that DNA performs this function; once we explain how the function is performed,

⁶ Ibid.

⁷ Ibid.

we have explained the gene.⁸

Chalmers holds that the same goes for a range of cognitive abilities and functions.

Chalmers gives some examples:

[T]he ability to discriminate, categorize, and react to environmental stimuli; the integration of information by a cognitive system; the reportability of mental states; the ability of a system to access its own internal states; the focus of attention; the deliberate control of behaviour; the difference between wakefulness and sleep.⁹

In his book Chalmers, also, seems to allow for the functional definition and explanation of *intentionality*. A typical example of such a functional definition is provided below.

The form that Lewis gave the functional definition of mind is a pretty good example, here is the recipe that one needs to follow if one wants to form a functional definition of the mental:

- (0.1) Let T stand for the names of some theory and let O stand for the other terms of the theory. And let them stand in relation to each other in a very long sentence.
- (0.2) Then we can replace T terms ($t_1...t_n$) with variables ($x_1...x_n$). Leaving us with a long open sentence containing variables and other terms and their causal relations.

⁸ David Chalmers, (1995), "Facing up to the problem of Consciousness", in *Journal of Consciousness Studies* 2(3), 200-219, extracted <http://consc.net/papers/facing.html>, 15/11/2013

⁹ Ibid.

(0.3) Then bind said variables with an existential quantifier.¹⁰

This gives one the “Ramsey sentence” for the theory.¹¹ And, at this point, it is said, we have a functional definition of the terms that are bound by the existential quantifier.¹² The variables stand for anything that will make the sentence true, if anything does. In relation to the theory of psychology containing theoretical terms of the following kind (belief, perception, desire etc.), proceeding in the aforementioned way gives one something like the following:

(1.1) John has a belief that p and that is caused by the perception that q and together with the desire that r causes action a .

(1.2) John has x and x is caused by y and together with z causes action a .

(1.3) $(\exists x)(\exists y)(\exists z)$ John has x and x is caused by y and together with z causes action a .¹³

Chalmers sees this method of explanation as *reductive*: Throughout the higher-level sciences, reductive explanation works in just this way, he says.¹⁴ Well, sometimes it fails. Chalmers says that the hard problem of consciousness has to do with *experience*:

¹⁰ Cf. David Lewis, (1972), “Psychophysical and Theoretical Identifications” in *Australasian Journal of Philosophy*, 50, 249-58

¹¹ A Ramsey sentence is a way to quantify over “theoretical entities” first developed as a way to formalize talk about non-observable atomic particles such as “electrons.”

¹² Cf. John Searle, (1994), *The Rediscovery of Mind*, (Cambridge, Massachusetts: The MIT Press), especially pp.41-42

¹³ Cf. John Searle, (2004), *Mind: A Brief Introduction*, (New York and Oxford: Oxford university Press), pp.44-45

¹⁴ Op cit., Chalmers, (1995)

The really hard problem of consciousness is the problem of experience...As Nagel...has put it, there is something it is like to be a conscious organism. This subjective aspect is experience.¹⁵

These subjective aspects are of the following kind:

[T]he felt quality of redness, the experience of dark and light, the quality of depth in a visual field. Other experiences go along with perception in different modalities: the sound of a clarinet, the smell of mothballs. Then there are bodily sensations, from pains to orgasms; mental images that are conjured up internally; the felt quality of emotion, and the experience of a stream of conscious thought.¹⁶

The reason the hard problem of consciousness centres upon the subjective aspect of experience is that no functional explanation of the aforementioned features of consciousness is sufficient to tell us all we want to know or understand about them:

The problem persists even when the performance of all the relevant functions is explained.¹⁷

What makes the hard problem hard and almost unique is that it goes beyond

¹⁵ Ibid.

¹⁶ Ibid.

¹⁷ Ibid.

problems about the performance of functions.¹⁸

There is an explanatory gap (a term due to Levine 1983) between the functions and experience, and we need an explanatory bridge to cross it.¹⁹

The explanatory gap referred to here was made famous by Levine who used a famous argument by Kripke to make his point.²⁰ Here is that argument and Levine's analysis of it. Kripke, it is commonly held, argues that we can intuit the disassociation of pain and the firing of c-fibres in some counterfactual situation.²¹ Therefore, it is false that necessarily pain is the firing of c-fibres. Therefore, the identity between pain and the

¹⁸ Ibid.

¹⁹ Ibid.

²⁰ See also Ray Jackendoff, (1987), *Consciousness and the Computational Mind*, (Cambridge, Massachusetts: The MIT Press) on the "explanatory gap"; Jean-Michel Roy, Jean Petitot, Bernard Pachoud and Francisco Varela, (1999), "Beyond the Gap: An Introduction to Naturalizing Phenomenology", in (Eds.) Jean Petitot, Francisco Varela, Bernard Pachoud, and Jean-Ichel Roy, (1999), *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*, (Stanford, California: Stanford University Press), pp.1-83

²¹ It really doesn't matter if one can't intuit this oneself. The fact of the matter remains that Kripke, Ned Block, David Chalmers, Joseph Levine, John Searle, et al. can. The point of Kripke's argument, and what Levine finds compelling, is that this intuition cannot be shown to be false by what Kripke refers to as the "usual moves and analogies". Kripke, (1980), *Naming and Necessity*, (Cambridge, Massachusetts: Harvard University Press) p.155 That Kripke thinks that he can intuit the dissociation is clear from the last line of *Naming and Necessity*, see. Ibid. p.155. That Levine thinks that Kripke has such an intuition is clear from the first paragraph of the following work. J Levine (1983), "Materialism and Qualia: The Explanatory Gap" in *Pacific Philosophical Quarterly*, 64, 354-361, p 354 If one can't intuit it, then one must explain where Kripke's intuition goes awry. But that's what Kripke is trying to show can't be done, *his* intuition cannot be spirited away. That's what lies at the heart of his argument. This is shown much more clearly in chapter two with the focus on the psychology of the intuiting party. Cf. David Papineau, (2009), "Kripke's Proof That We Are All Intuitive Dualists" extracted <http://www.kcl.ac.uk/ip/davidpapineau/Staff/Papineau/OnlinePapers/Kripke's%20Proof.htm> 27/10/2013 for the idea that Kripke's intuition is more widespread than suggested in this footnote.

firing of c-fibres is false. Therefore, it is necessarily the case that pain is not the firing of c-fibres.²² We may counter that we can intuit the disassociation of heat and molecular motion to the conclusion the identity between heat and mean molecular motion is false, and any further metaphysical conclusions thereof. However, Kripke says, that in the latter case the intuition can be explained away because what is really being imagined is the disassociation of the *sensation* of heat and mean molecular motion and not the separation of heat *proper* and mean molecular motion. This move is not available, however, in the former case. For in the former case pain *proper* just is the *sensation* of pain. Therefore, the metaphysical conclusion is left untroubled by argument.

Kripke argues that if objects are identical, then they are necessarily identical. Kripke also argues that if two proper names, α and β stand in the following relationship, $\alpha = \beta$, then it follows that necessarily $\alpha = \beta$. This can be summed up by the following proposition:

$$(K) \alpha = \beta \rightarrow \Box(\alpha = \beta)$$

There are two qualifications: (a) this kind of analysis only applies to so-called *genuine properties*, a genuine property is basically an essential property, which is important to note when essentiality rather than identity is stressed;²³ (b) The second qualification is that the identity (or essentiality) in question is between those that are named, not the

²² To be fair to Kripke, he only says that this or something the intuition that implies this cannot be ruled out.

²³ A Kripkean example is, I am essentially the product of my father's sperm and my mother's egg, though, of course, not identical to them.

names.²⁴

A proof of (K) may go like this: if α and β pick out the same object in counterfactual situation, P1, and one object can never fail to be identical to itself, then $\alpha = \beta$ is true in P1. If this is true of P1, then it is true of Pn, for if α and β pick out the same object in any counterfactual situation, Pn, given one object can never fail to be identical to itself in Pn, then $\alpha = \beta$ is true in Pn. But what is true in all counterfactual situations is necessarily true. Therefore, we can conclude that necessarily $\alpha = \beta$, or $\Box(\alpha = \beta)$

We can also understand Kripke as arguing that if two proper names, α and β stand in the following relationship, $\alpha \neq \beta$, then it follows that necessarily $\alpha \neq \beta$. This can be summed up by the following proposition:

(K') $\alpha \neq \beta \rightarrow \Box(\alpha \neq \beta)$

A proof of (K') may go like this: if α and β pick out distinct objects in counterfactual situation, P1, and two distinct objects can never fail to be non-identical, then $\alpha \neq \beta$ is true in P1. If this is true of P1, then it is true of Pn, for if α and β pick out the same object in any counterfactual situation, Pn, given two distinct objects can never fail to be non-identical in Pn, then $\alpha \neq \beta$ is true in Pn. But what is true in all counterfactual situations is necessarily true. Therefore, we can conclude that necessarily $\alpha \neq \beta$, or $\Box(\alpha \neq \beta)$

²⁴ See G. W. Fitch, (2004), *Saul Kripke*, (Montreal and Kingston: McGill-Queens University Press), p.90

Here is an example involving identity: ‘the Duke of Wellington’²⁵ and ‘Arthur Wellesley’ pick out the same object in counterfactual situation, P1, and one object can never fail to be itself, so ‘the Duke of Wellington is Arthur Wellesley’ is true in P1, and ‘the Duke of Wellington’ and ‘Arthur Wellesley’ pick out the same object in the counterfactual situation, Pn, so given one object can never fail to be itself, ‘the Duke of Wellington is Arthur Wellesley’ is true in Pn. Thus ‘the Duke of Wellington is Arthur Wellesley’ is true in all counterfactual situations. But what is true in all counterfactual situations is necessarily true. Therefore, ‘necessarily the Duke of Wellington is Arthur Wellesley’.

Here is an example involving non-identity, ‘Arthur Wellesley’ and ‘Barack Obama’ pick out distinct objects in counterfactual situation, P1, and two distinct objects can never fail to be non-identical, so ‘Arthur Wellesley is not Barack Obama’ is true in P1, and ‘Arthur Wellesley’ and ‘Barack Obama’ pick out distinct objects in counterfactual situation, Pn, so given two distinct objects can never fail to be non-identical, ‘Arthur Wellesley is not Barack Obama’ is true in Pn. But what is true in all counterfactual situations is necessarily true. Therefore, ‘necessarily Arthur Wellesley is not Barack Obama’.

The analysis can be extended to the analysis of all *rigid designators* and the relations of identity and essential properties thereof. A rigid designator is a designating term that picks out the very same object in all counterfactual situations. In the above schematic, then, α and β can be understood as standing for rigid designators.

²⁵ If you object that ‘the Duke of Wellington’ is not a name, then please change to ‘The Iron Duke’, which was Arthur Wellesley’s nickname.

Given this idea of necessity, we can examine the following set of statements²⁶:

- (1) Heat is the motion of molecules.
- (2) Felt-heat is the motion of molecules.
- (3) Pain is the firing of c-fibres.
- (4) Pain is felt-pain.

By the reasoning above we can conclude:

- (5) Necessarily heat is the motion of molecules.

That is because ‘heat’ and ‘the motion of molecules’ are rigid designators and they pick out the same object in P1, and given one object cannot fail to be itself, ‘heat is the motion of molecules’ is true in P1 and the same is true of Pn . Given what is true in all counterfactual situations is necessarily true, then (5) follows from (1).

²⁶ Not all expressions starting with “the” are definite descriptions, for example, consider ‘The Hague’, the *name* of the capital of Holland, or ‘The Edge’, the name of U2’s guitarist, ‘The Smiths’, the *name* of an English band, ‘the assassination of President Kennedy’, the *name* of an event, etc. Although “the” is frequently used in the identity statements given these are not definite descriptions, that is, these are naming expression and, therefore, rigid designators. This is because “The” prefixes the *names* of sensations, processes, people, etc. Compare, for example, with the expression, “the assignation of President Kennedy”. This designates an event rigidly because the term “President Kennedy” is a rigid designator. Kripke variously talks about “C-fibre stimulation” and “the stimulation of C-fibres” which mean the same thing for him, and each rigidly designates because ‘C-fibres’ he understands to be a rigid designator: “The same holds for the term ‘C-fibres stimulation’, provided that ‘C-fibres’ is a rigid designator, as I will suppose here... Thus the identity of pain with the stimulation of C-fibres, if true, must be *necessary*”. Op cit. Kripke (1980), p.149. Kripke also says, “if ‘C-fibres’ is not a rigid designator, simply replace it by one which is, or suppose it used as a rigid designator in the present context.” Ibid. p.149 Levine talks of “the firing of C-fibres” and assumes rigid designation applies. Op cit. Levine (1983), p.354.

However, let us consider (2). We can draw the following conclusions:

(6) If felt-heat is the motion of molecules, then necessarily felt-heat is the motion of molecules

(7) It is false that necessarily felt-heat is the motion of molecules.

That is because whether I have the sensation or not the motion of molecules would produce all the effects of heat in the physical vicinity, i.e. melting stuff, etc. Therefore, by modus tollens,

(8) It is false that felt-heat is the motion of molecules.

And, given 'felt-heat' is a rigid designator and 'the motion of molecules' is a rigid designator, then

(9) It is necessarily false that felt-heat is the motion of molecules.

The reasoning is: 'felt-heat' and 'the motion of molecules' are rigid designators and they pick out distinct objects in P1, and two distinct objects can never fail to be non-identical, 'heat is not the motion of molecules' is true in P1, and is true in Pn. But what is true in all counterfactual situations is necessarily true. Therefore, 'necessarily felt-heat is not the

motion of molecules'. That is (8) is true and (2) is false.

By similar reasoning, which I pass over here, we can justify the following propositions:

(10) It is false that necessarily pain is the firing of c-fibres.²⁷

Therefore,

(11) It is false that pain is the firing of c-fibres.²⁸

Allowing the conclusion, by the reasoning above:

(12) It is necessarily false that pain is the firing of c-fibres

Last, assuming (4),²⁹

²⁷ This is just Kripke's intuitions. Intuitions are often instinctive without explicit grounds, that's why they may not be called reasons, but mere intuitions. The OED defines "intuition" in the following manner: "the ability to understand or know something immediately, without conscious reasoning". It also defines "intuitive" thus: "based on what you feel to be true; instinctive". This is the sense in which the statement is taken as true by many.

²⁸ The reasoning here is this $a = b \rightarrow \Box(a = b), \neg\Box(a = b)$, therefore, $\neg(a = b)$.

²⁹ This proposition is generally accepted by most philosophers of mind and scientists because most philosophers of mind accept that pain is a feeling or sensation. It is definitely accepted by Kripke. See S Kripke (1980), p.151, for example. For a view that sees felt-pain as merely sufficient to pain see Thomas Polger, (1999), "Kripke and the Illusion of Contingent Identity", (APA Eastern Division (Boston, Massachusetts) Colloquium Paper), extracted <http://homepages.uc.edu/~polgertw/Polger-APA1999.pdf>, 25/11/2013

(13) Necessarily pain is felt-pain.

From (12) and (13) we may further conclude that:

(14) It is false that necessarily *felt-pain* is the firing of c-fibres

This allows Kripke to argue in the following way:

(T1) We might have the intuition that there are counterfactual situations in which heat and the motion of molecules are separated. However, if we imagine the latter, then what one is imagining is the proposition that (8) expresses, one is not imagining the falsity of (1) in some world.

(T2) We might also have the intuition that there are counterfactual situations in which pain and the firing of c-fibres are distinct, that is, we may intuit the truth of (10). Yet, when we imagine the disassociation of pain and c-fibres firing, we cannot be said to be imagining *merely* that (14).

The conclusion is that when we are intuiting the possibility of disassociation between heat and molecular motion, we are confusing our notion of heat *proper* and felt-heat. The same conclusion cannot be drawn for our disassociation of pain and c-fibres firing, for pain *proper* is nothing but felt-pain. (11), therefore, cannot *be explained away* by equivocation

and seems to be a metaphysical truth.

To be sure, according to Levine, the intuition that heat can be disassociated from molecular motion is explained in the way that Kripke says. The reason why this dissolution is satisfactory to “us” is because all that there is to know about heat is its *functional*, or *causal*, role in the environment. This includes the role it plays in producing felt-heat. (Of course this can’t actually be true without an answer to the mind-body *interaction* problem, but Levine doesn’t address this difficulty.) The connection between heat *proper* (molecular motion) and felt-heat is a contingent relation. And the reason why a parallel separation cannot be had in relation to the disassociation of pain *proper* and the firing of c-fibres is that all that there is to know about pain is not satisfied by a functional, or causal, explanation of pain’s role in the environment. Thus the firing of c-fibres as one explanation of pain, even if such an explanation includes explaining the role c-fibres play in producing felt-pain, does not suffice, for it leaves out explaining the *quality* of pain, precisely the *sensational* aspects of pain. This gives the intuition in question some legitimacy. Yet, says Levine, the intuition cannot ground a *metaphysical* conclusion. The most the intuition can do is ground an epistemological conclusion. We can sum up Levine’s position thus:

- I. The intuition that pain can be disassociated from the firing of c-fibres has motivation.
- II. The intuition finds traction because the functional, causal, explanation of pain does not explain the quality or sensational aspects of pain. Therefore, an *explanatory gap* remains.

III. However, an epistemological possibility,³⁰ which the intuition can be seen as centring, does not suffice to establish the metaphysical conclusion drawn.³¹

Chalmers is making reference to this explanatory gap, which has, as he points out, everything to do with the gap between functional, or causal, explanations and the perceived inadequacy of such explanations to explain the quality of the subjective aspects of experience.

Returning to Chalmers, he notes questions go left unanswered:

Why is the performance of these functions accompanied by experience? A simple explanation of the functions leaves this question open.³²

Why doesn't all this information-processing go on "in the dark", free of any inner feel?³³

Here the usual explanatory apparatus of the sciences, reductive explanation via functional explanation, fails:

³⁰ An "epistemic possibility" is a possibility *for me* based on what I know (what I don't know is a better way to put it). So, for Shota, who is seven, it's possible that water is made of the same stuff as the sky because *he knows no better*, and that makes it an epistemic possibility for Shota.

³¹ See, Ibid. Levine (1983)

³² Op Cit., Chalmers, (1995)

³³ Ibid.

To explain experience, we need a new approach. The usual explanatory methods of cognitive science and neuroscience do not suffice. These methods have been developed precisely to explain the performance of cognitive functions, and they do a good job of it. But as these methods stand, they are only equipped to explain the performance of functions. When it comes to the hard problem, the standard approach has nothing to say.³⁴

Chalmers often stresses that one way to see that functional explanations will not do is to consider the conceptual (or logical entailments) involved:

The facts about experience cannot be an automatic consequence of any physical [or functional] account, as it is conceptually coherent that any given process could exist without experience. Experience may arise from the physical [or related functions], but it is not entailed by the physical [or related functions].³⁵

Indeed, this is why the questions above are leftover. To sum up then, the functional aspects of consciousness can be given functional explanations using the reductive methods of explanation pertinent to the sciences. The *non-functional* aspects of consciousness cannot be given functional explanations using the reductive methods of the explanations pertinent to the sciences. The aspects of consciousness in need of the former kind of explanation constitute the (relatively) easy problem of consciousness. The aspects of consciousness in need of the latter kind of explanation constitute the hard problem of

³⁴ Ibid.

³⁵ Ibid.

consciousness. Such aspects are characterised by Chalmers as the subjective aspects of experience, or what an experience feels like, or just what it is like to feel or have an experience.

These subjective aspects of experience have been addressed by other authors too. Elitzur, for example, makes a distinction between the qualitative and quantitative aspects of reality.

When dealing with consciousness, science miserably fails in what has always been its hallmark of success, namely, reducing qualities to quantities.³⁶

Elitzur asks us to consider a set of statements, of which one is chosen here, to make his point. Consider the following statement: ‘Red differs from blue’.

Such a statement he says can be given a quantitative analysis thus:

Both red and blue light are electromagnetic waves, differing only in their wavelengths: 700 nm for red and 400 nm for blue. Consequently, different cones in our retina react differently to these wavelengths due to different amino-acid sequences of their rhodopsin.³⁷

³⁶ Avshalom Elitzur, (2009), “Consciousness Makes a Difference: A Reluctant Dualist’s Confession”, in A. Batthyany & A. C. Elitzur (eds.), *Irreducibly Conscious: Selected Papers on Consciousness*, (Heidelberg and New York: Winter), and extracted <http://a-c-elitzur.co.il/site/siteArticle.asp?ar=675/11/2013>, p.4

³⁷ Ibid. p.4

He says of the preceding example and others he provides:

In all these examples, qualitative differences between percepts turn out to be basically quantitative.³⁸

Continuing, he contrasts these explanations to the explanations required for the subjective aspects of experience, which he (as others do) refers to as *qualia*:

[T]hese explanations do a good job with percepts, rendering them (through neuroscience and chemistry) physical events, [but] some intriguing phenomena that accompany these percepts are left out. These are pure qualities, *qualia*.³⁹

Qualia (“quale” in singular) are those aspects of our experience that cannot be communicated yet we know they are there.⁴⁰

Pressing home his point with the following paradox:

Suppose that, with sufficiently advanced technology, you obtain the fullest real-time description of what goes on in my brain – every neuron, synapse and neurotransmitter molecule – when I see a red rose...Paradoxically, the problem now becomes worse: You know better than I do what goes on in my brain when

³⁸ Ibid. p.5

³⁹ Ibid. p.5

⁴⁰ Ibid. p.5

I perceive red, and yet, that doesn't bring you any closer to my quale of red.⁴¹

The subjective aspects of experience, what some call the what it's like of experience, can thus, also, be understood as the qualitative aspects of experience, those which are irreducible to the quantitative explanations of science; indeed, Elitzur tells us above that they are incommunicable. His prime example is the subjective experience of seeing red.

It is this subjective aspect of consciousness, its seemingly purely qualitative aspect, and considerations that deem it to be, for example, 'incommunicable' which attracts the bad press noted above. Citing Dennett, again, we find the following additional terms (in boldface) associated with dualism:

The phenomena of consciousness are an admittedly dazzling lot, but I suspect that dualism would never be seriously considered if there weren't such a strong undercurrent of desire to protect the mind from science, by supposing it composed of a stuff that is in principle **uninvestigable** by the methods of the physical sciences.⁴²

A family of compelling intuitions work to keep "the problem of consciousness"

⁴¹ Ibid. p.5

⁴² Daniel Dennett, (1991a), "The Brain and Its Boundaries", extracted <http://ase.tufts.edu/cogstud/papers/mcginn.htm>, 15/11/2013

systematically **insoluble**...⁴³

[Dualism] is a self-imposed **constraint**.⁴⁴

Now maybe there are **miracles**, but they are nothing science should ever posit in the course of business.⁴⁵

[P]ositing something which one has reason to believe must be **inexplicable** is going too far.⁴⁶

I'd pose a more lenient demand: that the dualist offer any articulated, non-**vacuous** explanation of anything in the realm of psychology or mind-brain puzzles.⁴⁷

It is quite clear that Dennett thinks dualism is a cul-de-sac, without scientific merit, which places the full stop of mystery where the ellipses of scientific research need to be filled in by doing the science.

Sir John Eccles the esteemed neuroscientist and dualist disagrees, he offers us what he believes to be a quasi-scientific theory of dualism. To be sure he recognises that the theory

⁴³ Ibid.

⁴⁴ Op cit. Daniel Dennett, (1993a)

⁴⁵ Ibid.

⁴⁶ Ibid.

⁴⁷ Ibid.

is *untestable*, but believes this is made up for by the theories *explanatory* power. Libet presents us with a concise elaboration of Eccles late thinking:

Eccles realized that some sort of field would have to be postulated to account for the integrative aspects of the mind. For the elements in the brain that give rise to the field, Eccles...postulated the existence of organized bundles of neurons that he called ‘‘psychons’’...Each psychon could represent a mental event or process. Eccles, in collaboration with Beck, proposed that synaptic probability for release of its neural transmitter is affected by random quantum inputs...Such inputs could not be detected by any physical measurement and could thus be a mental action that is not externally apparent. A field of appropriate psychons, acting together, would produce an integrated mental experience. Eccles admitted that such a process was not testable (personal communication).⁴⁸

This untestability is countered by the explanatory power of the theory advanced:

[H]e argued that the hypothesis had **explanatory** power, and, as such, he believed it had some usefulness and even validity.⁴⁹

However, Dennett is vehement in his rejection of the doctrine of dualism, particularly with respect to the mystery which seems to be the inevitable outcome of the subjective

⁴⁸ Benjamin Libet, (2006), ‘‘Reflections on the Interaction of the Mind and Brain’’ in *Progress in Neurobiology* 78, 322-326, p. 324

⁴⁹ Ibid. p.323

aspects of consciousness; he writes:

I find [the] thesis not just **incredible** and **ludicrous**. As a fellow philosopher, I find it **embarrassing**.⁵⁰

Here Dennett is making a pointed attack on Colin McGinn. Dennett continues his attack thus:

McGinn's central thesis is that the problem of consciousness is systematically **insoluble** by us (Martians or demigods might have better luck). Our brains just weren't meant to get a grip on this tough problem, but—there, there, it's all right—we mustn't draw the conclusion from the fact that we can't understand it, that the mind is **intrinsically mysterious**.

Colin McGinn introduced the notion of *cognitive closure* to the philosophy of mind.⁵¹ Property P is cognitively closed to X if it is impossible for X to grasp the concept(s) associated with P. An example of this is as follows: the properties of quantum physics are cognitively closed to chimpanzees because it is impossible for chimpanzees to grasp the associated concepts. So far as McGinn rejects dualism and eliminativist materialism as extremes, he endorses the “naturalistic” view that there is some property P that explains the mind-body relation, yet at the same time he holds that humans are cognitively closed

⁵⁰ Op cit. Daniel Dennett, (1991a)

⁵¹ Colin McGinn, (1989), “Can We Solve the Mind-Body Problem”, in *Mind, New Series*, Volume 98, Issue 391 (July 1989), 349-366.

with respect to property P. He comes to this conclusion for the following reasons: (a) by *introspection* there is no conceptual way to make this property out; (b) by observation, theory, or inference to the best explanation there is no conceptual way to make this property out. Therefore, McGinn concludes humans are cognitively closed to P, that is, the solution to the mind-body problem is cognitively closed to humans and the human sciences. Though the case may be different for alien minds, hence the Martians and Demigods aforementioned in Dennett.

But Dennett ascribes *ignorance* to McGinn, in particular he attributes *scientific ignorance* or *neglect* to McGinn for his conclusions:

McGinn has carefully—and correctly—deduced just the properties that scientific concepts must have if they are to offer a genuine explanation of consciousness, but then he **neglects** to look to see if any such concepts have been developed by the relevant sciences.⁵²

Yet, I think, Dennett is a little unfair. McGinn's position is much less superficial than Dennett seems to think it is and rich in the philosophical tradition (well, perhaps, that's where Dennett thinks McGinn goes wrong). In a little more detail, McGinn's thinking can be stated in the following manner: Understanding that the mind has limits is *not* an eccentricity. McGinn locates the idea that human knowledge is limited squarely in the philosophical tradition. Here are some examples: Plato argued that everything has an ideal form, however, we are unable to grasp such forms; Descartes conjectured the world might

⁵² Op cit. Dennett, (1991a)

be the work of an evil demon and wondered how we might cut through such doubt; Kant argued that the thing-in-itself was beyond our capacity to know; Wittgenstein thought the limits of thinking were constrained by language and its logical structure; Rorty set limits culturally; and so on. For McGinn, Chomsky's biological limits are a good place to start. Following Chomsky, McGinn places the limits of knowledge in the biological sphere and it is *species*-relative. These limits are not external to the knower and not an essential element of the subject-matter of knowledge but internal to the architecture of the knower's cognitive capacities. We may view the limits of our knowledge along a spectrum. At one end of the cognitive spectrum, things come easy to us, for example, the acquisition of a language; towards the other end things are "hard-won", for example, what can be learned from the physical sciences; at the diametric end of the spectrum we find things which we are "chronically unable" to make progress on. There are a number of signs which mark this extremity: Intractability over time, recurrence of the same questions and fashionable answers over time, etc. Philosophy, at least in its intractable forms, is symptomatic. Such chronically intractable problems are dealt with in four usual ways: (a) reductive explanations; (b) non-reductive explanations; (c) non-natural or magical explanations; (d) elimination. The problem of consciousness is a prime example, compare, for example, physicalism; dualism; occasionalism; eliminativism for familiar ways of dealing with consciousness. The intractable nature of the problem of consciousness, its "hardness" is due not to the subject matter itself, but the nature of the mind that is trying to grasp it. Our competence with language is central to understanding such limits. "Discrete infinity" or recursion is crucial to this thought. Discrete infinity is the property of making sense in our language via discrete units which may be added to recursively to create more discrete

units ad infinitum and which, for Chomsky, is innate in us.⁵³ Chomsky speculates it is involved in the way we can understand numbers and how they work, that is, because we can adapt this kind of linguistic competency to such subject matter. McGinn speculates that a combination of such linguistic abilities and the capacity to apply such abilities to our spatial representations allows us to come to know the physical world. However, so far as consciousness is concerned, consciousness cannot be subsumed under such linguistic competency. Therefore, consciousness and whatever intractable philosophical problems we find lie beyond the scope of our understanding.⁵⁴ Interestingly, though, we do possess knowledge about consciousness. This is stored in our DNA. It is just that such information cannot be represented by us cognitively. McGinn sees his *speculations* as a *naturalistic hypothesis*.⁵⁵ If this is true, perhaps, it is not as unscientific as Dennett argues and McGinn can be understood to be a physicalist in some sense. Dennett, however, will require that such a hypothesis is falsifiable if it is to count as a *scientific* hypothesis.

To return to Dennett's attacks on dualism, Dennett himself gives us a definition of the easy and hard problems of consciousness in order to undermine it precisely in terms of its unscientific nature:

This alleged division between experience and function is often mapped onto the

⁵³ See Noam Chomsky, (1965), *Aspects of the Theory of Syntax*, (Cambridge, Massachusetts: The MIT Press)

⁵⁴ See, for example, Colin McGinn, (1999), *The Mysterious Flame: Conscious Minds in a Material World*, (New York: Basic Books), especially pp.54-62

⁵⁵ Colin McGinn, (1993), *Problems in Philosophy: The Limits of Inquiry*, (Oxford: Blackwell Publishers)

distinction between the ‘hard and ‘easy problems of consciousness. Under this view, the hard problem is answering the question of how phenomenal experience arises from physical events in the brain, whereas the easy problems are characterizing the mechanisms supporting cognitive functions.⁵⁶

He argues, with his co-author Michael A. Cohen, that the hard problem is in fact the *impossible* problem of consciousness and, again, relates this to the unfalsifiable and, therefore, unscientific nature of the problem. Dennett and Cohen’s argument proceeds by imagining the “perfect experiment”:

Consider perhaps the most drastic experiment possible, the ‘perfect’ experiment: imagine that, in the future, surgeons are able to isolate the parts of the visual cortex that represent colour while wholly preserving their activation patterns. After this surgery, the areas involved in colour perception (visual area V4, inferotemporal cortex etc.) behave normally but are simply unable to project to higher brain areas: perfect isolation. Although the colour areas are isolated, all other visual areas (e.g. motion, luminance, object recognition etc.) are untouched and project to higher-level regions in a normal manner.⁵⁷

⁵⁶ Michael A. Cohen and Daniel C. Dennett, (2011), “Consciousness Cannot Be Separated From Function”, in *Trends in Cognitive Science*, 15, 358-364, p.358

⁵⁷ Ibid. p.361

Now what will happen if the subject of this surgery is presented with a red apple? The dualist, or theorist who posits the dissociation of consciousness (the subjective aspects of experience) and function, will hold that:

When presented with a red apple there will be normal activation of the colour areas of the brain but without projections to higher-level areas. Other areas of the brain (e.g. object representation and identification, language production etc.) will function normally, so the patient will be able to report that he or she sees an apple but an apple that has no colours.⁵⁸

That is:

[Theorists] that posit dissociation between consciousness and function would necessarily assume that participants of the ‘perfect experiment’ are conscious of the apple’s colour but simply cannot access that experience. After all, the conditions these theories stipulate for phenomenal consciousness of colour are all met, so this experiment does not disprove the existence of isolated consciousness; it merely provides another particularly crisp example of consciousness without access.⁵⁹

Dennett and Cohen say that this precisely highlights the problem:

⁵⁸ Ibid. p.361

⁵⁹ Ibid. p.362

However, there is a crucial problem with this logic. If this ‘perfect experiment’ could not definitively disprove dissociative theories, then what could? The subject manifests all the functional criteria for not being conscious of colour so what would ground the claim that the subject nevertheless enjoys a special kind of consciousness: phenomenal consciousness without access consciousness?⁶⁰

The distinction between access- and phenomenal-consciousness is Ned Block’s.⁶¹ Blind-sight patients, for example, are said to have a phenomenal-consciousness that exceeds accessible consciousness, that is, they are phenomenally conscious of more than they can access functionally.⁶² The original evidence for this distinction is drawn from the work of George Sperling.⁶³ Sperling is taken to have shown that phenomenal consciousness does in fact exceed accessible consciousness. He asked subjects to look at a 4 by 3 grid, each cell of which contained a letter, for a limited amount of time. He then asked subjects to recall as many of the letters as possible. On average, subjects could only manage to recall 4-5 letters, although they claimed to have seen many more or all of the letters. Sperling, then, repeated the experiment. A grid of letters was presented to a group of subjects for a short time. This was followed by a high-pitched tone, medium pitched tone

⁶⁰ Ibid. p.362

⁶¹ Ned Block, (1995), “On a confusion about a function of consciousness” in *Behavioural and Brain Sciences*, 18, 227–47

⁶² See *ibid.*

⁶³ George Sperling, (1960), “The Information Available in Brief Visual Presentations”, in *Psychological Monographs: General and Applied*, 74, (11, Whole No. 498), 1-29

or low pitched tone. The high-pitched tone was meant to cue the subjects for a report about the top line, the middle pitched tone to cue the subjects for a report about the middle row, and a low pitched tone a report about the bottom row. The subjects seemed to recall about 3 letters. The generalisation was that subjects, thus, retained about 9 letters in a kind of sensory store. (For example, if a given subject was cued to remember the top line and 3 letters were remembered, but the subject could have been cued to remember the second row and 3 letters remembered, and so for the bottom line, then in total about 9 letters would have been stored in the sensory store, 3 accessed at any given time.) The conclusion being that subjects took in about 9 letters even though only able to report 3 at any given time. Another way to put this is that subjects were conscious of 9 letters, but only aware of 3 at any given time. Or, again, subjects were *phenomenally*-conscious of 9 letters, but *access*-conscious of only 3 letters at any given time. Sperling himself drew a distinction between *sense*-memory, which later came to be called *iconic*-memory, and *working*-memory. The interpretation of the experiment is still up for grabs, and in fact thinkers like Dennett and Cohen try to interpret the findings in a way that does not bifurcate consciousness.⁶⁴ The findings remain hotly contested.⁶⁵ There is independent support for Sperling's conclusions, for example, from the work of Benjamin Libet who found that: "Cerebral cortical activities, in response to a somatosensory stimulus, must proceed for about 500 ms in order to elicit the conscious sensation".⁶⁶ Further, that

⁶⁴ Cf. Ian B Phillips, (2011), "Perception and Iconic Memory: What Sperling Doesn't Show" in *Mind & Language*, Vol. 26, No. 4 September 2011, 381–411

⁶⁵ Ned Block, (2011), "Perceptual Consciousness Overflows Cognitive Access" in *Trends in Cognitive Sciences*, December 2011, Vol. 15, No. 12, 567-575, p.567

⁶⁶ Benjamin Libet, (2002), "The Timing of Mental Events: Libet's Experimental Findings and Their Implications" in *Consciousness and Cognition* 11, 291–299, p.291

stimulation of a shorter duration resulted in “unconscious detection” and that unconscious detection could be made conscious through further stimulation. “Antedating” the experience was proposed in order to account for the lack of delay in the subjective experience of the stimulus. The key point here is that phenomenal consciousness seems to be present in the mind before it can be accessed.

However, Dennett and Cohen’s conclusion is that:

[D]issociative theories are inherently **unfalsifiable** and beyond the scope of science, because inaccessible conscious states are intrinsically off-limits to investigation.⁶⁷

It should be noted that Fahrenfort and Lamme have directly contested the unfalsifiability claim to which Dennett and Cohen have, in turn, responded.⁶⁸

Why all this to-and-fro? John Searle, who does not deny the subjective aspects of consciousness, blames a set of traditional connotations which swim about the central notion of subjective feel for obscuring the philosophy of mind:

The weird feature about this entire discussion is that materialism inherits the

⁶⁷ Op cit. Cohen and Dennett (2011)

⁶⁸ Johannes J. Fahrenfort and Victor A.F. Lamme, (2012), “A True Science of Consciousness Explains Phenomenology” in *Trends in Cognitive Sciences*, Vol. 16, No. 3, pp.138-139; Cf. Michael A. Cohen and Daniel Dennett, (2012), Response to Fahrenfort and Lamme: Defining Reportability, Accessibility and Sufficiency in Conscious Awareness” in *Trends in Cognitive Sciences*, Vol. 16, No. 3

worst assumption of dualism. In denying the dualist's claim that there are two kinds of substances in the world or in denying the property dualist's claim that there are two kinds of property in the world, materialism inadvertently accepts the categories and the vocabulary of dualism. It accepts the terms in which Descartes set the debate. It accepts, in short, the idea that the vocabulary of the mental and the physical, of material and immaterial, of mind and body, is perfectly adequate as it stands.⁶⁹

Indeed, this is something that Dennett often enough stresses, too:

Ill-considered Cartesian leanings that once could be gracefully tolerated or ignored are now positively distorting the imaginations of theorists, who will not be able to take the next step in creating a theory of consciousness without coming to terms with these **residual metaphors** and **images**.⁷⁰

Eliminative materialists, like Richard Rorty, Paul Feyerabend and Paul Churchland make similar points. For example, Feyerabend refers to our conventional ways of talking about sensations and intentionality as a mere "historical accident".⁷¹ Paul Churchland equates this historical discourse with a theory, "folk psychology", and argues that as a theory it can be scientifically eliminated just the way other older scientific theories were replaced

⁶⁹ Op cit. John Searle, (1994), p.54

⁷⁰ Daniel Dennett, (1993b), "Caveat Emptor", in *Consciousness and Cognition*, vol. 2, no. 1, 1993, 48-57, extracted <http://cogprints.org/276/1/caveatem.htm>, 15/11/2013

⁷¹ Paul Feyerabend, (1993) "Materialism And The Mind-Body Problem" in Christensen, M. S., and Turner, D. R., (eds.), *Folk Psychology and the Philosophy of Mind*, Psychological Press, pp.3-17

with better theories.

Eliminative materialism is the thesis that our common-sense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience.⁷²

Richard Rorty, another kind of eliminativist, argues for the realignment of mental terms, historically associated with mental phenomena, with non-mental phenomena on pragmatic grounds, although he does not argue for the elimination of mental discourse as such.⁷³

Searle's point is a little different, however, he is arguing that materialists, *eliminativists too*, are guilty of sharing the same vocabulary as their dualist opponents and that this is why the debate seems to be irresolvable.⁷⁴

As we can see the antipathy to dualism and the nature of subjective experience, and arguments that tend to dualism can take a torrid form. This squarely seems to centre upon

⁷² Paul Churchland, (1981), "Eliminative Materialism and the Propositional Attitudes", in *Journal of Philosophy* 78, 67-90. p.67

⁷³ See Richard Rorty, (1965), "Mind-Body Identity, Privacy, and Categories", in *Review of Metaphysics* 19, 24-54; and Richard Rorty, (1970), "In Defense of Eliminative Materialism" in *Review of Metaphysics* 24, 112-121.

⁷⁴ Searle lists the shared vocabulary in Op cit. John Searle, (2004), p.81

the nature of the subject in question and its relations to our understanding. Thus, to summarise, *by its very nature* the subjective aspects of experience do not reduce to the functional or quantitative explanations of science. An aspect of reality is, therefore, thought to be left seemingly *inexplicable* and questions about it remain unanswered. It is this seeming inexplicability that garners the bad press associated with dualism and puts dualism on the back foot. Well, now that we understand what dualism is dealing with and the general discourse that surrounds it, let me try and define some forms dualism takes.

Section B: Kinds of Dualism

Dualism can be defined against monism. Dualism in the philosophy of mind typically opposes a specific sort of monism, *physicalism*, though it must also oppose other forms of monism, for example, idealism. If we use monism to define dualism we must first define monism. To help me do this, I want to introduce what William Seagers calls the Scientific Picture of the World (SPW):

Three interlocking features seem of central importance to the SPW: completeness, closure and resolution. Completeness is the doctrine that everything in the world is physical and as such abides by closure and resolution. Closure entails that there are no ‘outside forces’ – everything that happens, happens in accordance with fundamental physical laws so as to comply with resolution. Resolution requires that every process or object be resolvable into elementary constituents which are, by completeness, physical and whose abidance with laws governing these constituents leads to closure.⁷⁵

These three conditions can be used to define monistic pictures of the world, more generally, which are not necessarily physicalist. They may, for example, be used to define the barebones of both physicalism *and* idealism respectively. To do this we need to abstract the conditions in the following manner:

⁷⁵ William Seagers, (1999), “Generalised Epiphenomenalism”, extracted <http://www.uts.utoronto.ca/~seager/mywork.htm>, 10/01/2013, p.1

- (a) Completeness: Everything is x and only x.
- (b) Closure: Everything is governed by one set of laws, R, which only apply to things that are x.
- (c) Resolution: Everything can be *reduced* to the elements which constitute x or be eliminated.

This allows physicalism to be defined in the following fashion:

- (P1) Completeness says that everything is physical and only physical,
- (P2) Closure says that everything is governed by one set of laws, physical laws, which only apply to things that are physical, and
- (P3) Resolution says that everything can be reduced to elements which constitute the physical or be eliminated.

On the other hand, the three conditions cited may also be used to define idealism:

- (I1) Completeness says that everything is mental,
- (I2) Closure says that everything is governed by one set of laws, psychological laws, which only apply to things that are mental, and
- (I3) Resolution says that everything can be reduced to elements which constitute the mental or be eliminated.

Dualism can now be defined against monism:

(D1) Dualism denies completeness: It is false that everything is x and only x.

(D2) Dualism is not committed to closure: It is not certain that everything is governed by one set of laws, R, which only apply to things that are x.

(D3) Dualism denies resolution: It is false that everything can be reduced to the elements which constitute x or can be eliminated.

How one construes these denials will give rise to different and complex kinds of dualism. For example, (D1) may be thought to lead to the conclusion that everything is either physical or mental. This may mean, for example, everything is *both* physical and mental. But taking the disjunction in an exclusive sense says that everything is either physical *or* mental but not both. The first sense of the mental and physical may ultimately lead to a kind of panpsychism (whether dualist or otherwise), where that just means everything has both kinds of attribute, mental and physical. The latter sense may lead to an interactive dualism or some other form of dualism.

The second denial, encapsulated by (D2), may be viewed in a number of ways too. For example,

- i. Everything mental and physical is governed by one and only one set of laws;
or

- ii. everything mental is governed by one set of laws and everything physical is governed by another set of laws, or
- iii. everything mental is governed by one set of law and everything physical is governed by another set of laws and interactions between the mental and the physical are governed by yet another set of laws, or
- iv. everything mental is governed by one set of laws and everything physical is not governed by any set of laws, or
- v. everything mental is governed by no set of laws and everything physical is governed by one set of laws, or
- vi. nothing mental and nothing physical are governed by any set of laws.

The stance that a dualist might take on the lawful relations between the mental and the physical helps define three of the most prominent forms of dualism, that is, by how one understands the *causal* relations between the mental and the physical we get: Interactionism, parallelism, and epiphenomenalism. *Interactionism* understands the mental and the physical *interacting*, something like i and iii above. *Parallelism* understands the mental and the physical as interacting at the intra-structural level, but as inactive at the inter-structural level, something like ii. *Epiphenomenalism* understands the mental as causally inert with respect to either or both the intra- and inter-structural levels, or the physical as causally inert with respect to either or both stated levels, something like iv and v may be relevant. The following are examples of these positions (note that not all are strictly speaking dualistic).

Interactionism

Interactionism is *not* always a dualism. For example, consider Popper's theory of mind: Popper is neither a materialist (physicalist) nor a dualist. Popper is a *pluralist*. Nature is divided into what he calls *worlds* and, *at the least*, three of these: World 1: the *physical* world; World 2: the *mental* world; World 3: the *abstract* world.

In terms of mind we can restate these as "sub-universes" in the following terms: World 1: the world of the *brain*; World 2: the world of *subjective* thought processes and experience; World 3: the world of *objective content*. We can get a handle on this tripartite schema if we consider the following: 'I am writing'. There is a physical story to tell about this sentence to do with the ink, paper, light, reflection, eyes, neurons, etc. There is a mental story to tell about this sentence to do with the thought I had writing the sentence, your thought reading it, our subjective experience of its colour, etc. And there is an objective or abstract story to tell about this sentence to do with its logical relations, for example: it entails, 'Therefore, it's false I'm not writing', and expresses the same proposition as 「私は書いている」.

Each world is considered real, where real is defined as causally active. A scientific conjecture is a good way to establish the reality of each world. For example, consider the following conjecture: 'All swans are white'. Primarily this is a thought: a world 2 object. But the truth or otherwise it expresses is an abstract object: a world 3 object, with a set of logical relations, propositional and invariant in translation, etc. World 2 works on and improves world 3, for example, subjective thought processes are used to establish the logical consequence of the aforementioned statement: 'If one swan is black, then it is false

that all swans are white. Therefore, the conjecture that all swans are white is falsifiable. Being falsifiable is the mark of good science'. World 3 works on world 1 but only mediated by world 2, for example, I may think: 'I know that if I can find one swan that is black I will falsify the conjecture in question. I'm a swan scientist and I guess this is what I should be trying to do'. This leads to an ongoing quest to examine every swan I can find in order to disprove the conjecture in question, a question which can only occur by applying my thoughts and senses to the physical world around me. But all this could only happen if the physical world, itself, had given rise to the biological or mental world, that is, if world 1 caused world 2 to emerge.

There are clear *causal* roles here for each world:

- (a) World 1 gives rise to world 2, the physical gives rise to the biological including the mind.
- (b) World 2 gives rise to world 3: the mind gives rise to abstract content and works on it.
- (c) World 3 inter-acts with world 1 through world 2: abstract content acts on the physical and works on it through the mind.
- (d) World 1 inter-acts with world 3 through world 2: modifying senses and thought leading to improvements in objective content.

The world, mind and abstract content are, thereof, very much causally *inter-active*. So far as the mind and body is concerned we have a form of interactionism. We can also note, each world has a set of *intra-active* relations too. For physical things act on physical

things, thought processes influence thought process from one individual to another not just in one individual, and thought contents inter-act logically. To sum up, Popper is not, strictly speaking, a dualist, but he is an interactionist. Prominent *dualists* who believe in interactionism are, for example, Descartes, Elitzur and Eccles.

Elitzur is a dualist who believes, somewhat regrettably he says, in *interactionism*. It is regrettable to Elitzur, who is a physicist by profession, because allowing qualia (the qualitative aspects of subjective experience) causal efficacy violates the fundamental laws of physics. Indeed, he thinks “any causal role played by qualia is not only redundant but forbidden” by physics.⁷⁶ Yet, he is a dualist who accepts interactionism. Here is his argument for interactionism: Elitzur contends that when we think of a quale and a percept we are *baffled* by their perceived non-identity. If this is true, and if physicalism is true, then physicalism will provide a physical explanation for this bafflement. Since this is a falsifiable assertion, it makes physicalism falsifiable. Elitzur believes that a thought experiment alone is sufficient to falsify the physicalist conjecture. He asks us to “Imagine intelligent beings that resemble us in every detail our physiology, neuroanatomy and chemistry, but have no qualia.”⁷⁷ Elitzur contends that such creatures, zombies, are quite possible. He even thinks that “zombies accord with physics more than the existence of non-zombies.”⁷⁸ Here he means that beings which lack qualia are more amenable to physics than beings which possess qualia because the former do not violate the fundamental laws of physics. So, Elitzur talks about zombies as if they are possible. And

⁷⁶ Op cit, Elitzur, (2009), p.11

⁷⁷ Ibid.

⁷⁸ Ibid.

he thinks that a zombie would explain its bafflement of the non-identity in question by physical explanation alone, whereas a human will not. This is exemplified by the fact that for a zombie a zombie is inconceivable, whereas for a human a Zombie is quite conceivable. In effect there is:

- (a) Bafflement₁ which is the sense of bafflement that arises physically, and
- (b) Bafflement₂ which is the sense of bafflement that arises from the qualitative aspects of consciousness.

Elitzur seems to be saying that:

- (c) A Zombie will express no sense of bafflement₂
- (d) A human will express a sense of bafflement₂ (exemplified by worries about spectrum inversion and the problem of other minds, etc.).

The inevitable conclusion is that qualia must exert causal powers on us: it is the qualitative aspect of subjective experience that is the *cause* of the sense of bafflement² in humans and not some *physical* aspect (which is the cause of the sense of bafflement in Zombies). Elitzur reckons if it were it would be physically locatable. Chalmers, in turn, disagrees with Elitzur's position.⁷⁹

Eccles is another prominent scientist who believes in an interactionist dualism, he states

⁷⁹ For example, see Chalmers (1997), p.183. More will be said on this in Chapter two.

his position thus:

It is a very strong dualism...Briefly, the hypothesis states that the self-conscious mind is an independent entity that actively engages in the reading out from a multitude of active centres in the modules of the liaison areas of the dominant cerebral hemisphere.⁸⁰

The mind, therefore, liaises with the brain. And, furthermore, the brain acts back on the self or mind. This two-way interaction is summed up in the following way:

[T]he self-conscious mind exercises a superior interpretative and controlling role upon the neural events by virtue of a two-way interaction across the interface between World 1 and World 2⁸¹

World 1 is the physical world and world 2 is the mental world, this nomenclature is, of course, from Popper. Of note, and in contrast to Elitzur, Eccles aims to provide a way to avoid violations of the fundamental laws of physics. The self is not a material entity that exists in time and space but a “probabilistic field”. Eccles argues that consistent with the laws of quantum physics such fields do not carry mass or energy but are, nevertheless, active. Thus, the mind, as a probabilistic field does not violate the fundamental laws of physics but acts, nevertheless. The loci of action are the “liaison” areas of the neural

⁸⁰ Sir John Eccles and Karl Popper, (2003), *The Self and the Brain*, (London and New York: Routledge), p.355

⁸¹ *Ibid.*, p.355

cortex.⁸²

Parallelism

There are a number of different forms of parallelism that can be identified. For example, Gustav Fechner, one the most famous advocates of parallelism, is identified as holding *three* distinct forms by Michael Heidelberger: First, *empirical parallelism*: this view holds parallelism as a “methodical rule for researching the mind-body relation, claiming that there is a consistent correlation between mental and physical phenomena”.⁸³ There is no *causal* commitment on this view. The parallelism may be said to actively fail to endorse any of the possibilities of lawful interaction between mental and physical phenomena. Second, *identity theory*: this view is sometimes also known as “dual-aspect theory”. The mental and physical are merely defined in terms of the perspective, first- or third- person through which they are perceived. However, to all intents and purposes, the two are identical. It is not dissimilar to viewing an object from two different angles, the different perspectives on the one thing do not render that thing two different things. Again, this view has no need to posit causal relations between the two aspects in question because the two aspects in question are identical and, therefore, not in need of causal explanations. This allows the theorist to posit a single set of universal laws. Third, *panpsychical parallelism*: This is basically the extension of the last view to the whole of reality. Fechner

⁸² Sir John Eccles, (1994), *How the Self Controls Its Brain*, (Berlin: Springer-Verlag), see especially p.56

⁸³ Michael Heidelberger, (2003), “The Mind-Body Problem in the Origin of Logical Empiricism: Herbert Feigl and Psychophysical Parallelism”, in (Eds.) Paolo Parrini, Wesley C. Salmon, and Merrilee H. Salmon, *Logical Empiricism: Historical and Contemporary Perspectives*, (Pittsburgh, PA: University of Pittsburgh Press), 233-262, p.237

advanced this position in later life. Heidelberger says that “His argument rested on the premise that the mental must not necessarily correlate to a nervous system; it could also be realized in other material systems.”⁸⁴

Leibniz philosophy of mind is also a parallelism so far as the mind-body relation is concerned. But we have to keep in mind that strictly speaking Leibniz is *not* a dualist, for Leibniz *monads*, which are “simple substances” are infinite in nature. In the terminology that we have been using, each substance is *closed* with respect to another: “Strictly speaking, one can say that no created substance exercises a metaphysical action or influence on anything else.”⁸⁵

Therefore, so far as the mind is one particular kind of substance and body is another kind of substance, both of which may be said to compose a person, they are closed with respect to each other. There is no sense in which they *inter-act*. However, according to their status as substances and their closed nature there is a sense in which each *intra-acts*. Yet, as such both act as if in accord with each other. “[B]odies act as if there were no souls (though there couldn’t be no souls); and souls act as if there were no bodies. And both act as if one of them influenced the other.”⁸⁶ The system Leibniz is building upon is his system of *pre-established harmony*. If there *seems* to be causal interactions between the mind and body, this is due to a pre-established harmony between the two types of substances

⁸⁴ Ibid. p.240

⁸⁵ Gottfried Leibniz, (1686), *First Truths*, (Trans.) Jonathon Bennett, extracted http://www.earlymoderntexts.com/f_leibniz.html p.4

⁸⁶ Gottfried Leibniz, (1686), *Monadology*, (Trans.) Jonathon Bennett, extracted http://www.earlymoderntexts.com/f_leibniz.html 01/11/2013, S81

in question.⁸⁷ The soul follows its own laws, and the body likewise follows its own laws. They fit together in a *pre-established harmony* existing between all substances since they represent one and the same universe.⁸⁸ The pre-established harmony is summed up like this: “each created substance is programmed at creation such that all its natural states and actions are carried out in conformity with all the natural states and actions of every other created substance.”⁸⁹ It is God that orders this pre-established harmony. Leibniz writes to this end thus:

I cannot help coming into this Notion, that God created the *Soul* in such manner at first, as that it *produces within it self*, and *represents in it self* successively, what passes in the *Body*; and that he has made the *Body also* in such manner, as that it *must of it self* do what the Soul wills. So that the Laws which make the Thoughts of the Soul *follow each other* successively in the Order of final Causes, and in the Order of its Perceptions arising within it self; *must* produce Images, which shall be *coincident*, and go *Hand in Hand* with the Impressions made by Bodies upon our Organs of Sense: And the Laws by which the Motions of the Body *follow each other* successively in the Order of efficient Causes, are likewise *coincident* and go

⁸⁷ Ibid. S80

⁸⁸ Ibid. S78

⁸⁹ Mark Kulstad and Laurence Carlin, (2013), "Leibniz's Philosophy of Mind", *The Stanford Encyclopaedia of Philosophy* (Winter 2013 Edition), Edward N. Zalta (ed.), extracted <http://plato.stanford.edu/archives/win2013/entries/leibniz-mind/> 17/11/2013

Hand in Hand with the Thoughts of the Soul, in such manner as that these Laws of Motion make the Body act at *the same Time* that the Soul Wills.⁹⁰

So Leibniz theory of mind is a parallelism, but it is more generally not a dualism in the sense that there may be any number of closed substances. Like Popper, then, he is a pluralist. Still, so far as mind and body are concerned, we may talk of a mind-body parallelism.

Spinoza may also be attributed a parallelism. It may be apt to attribute Spinoza a kind of *dual-aspect* parallelism in so far as Spinoza considered mind and body to be aspects of one underlying reality, God.

Spinoza writes:

P1. Thought is an attribute of God, or God is a thinking thing.⁹¹

⁹⁰ Samuel Clarke, *A Collection of Papers, Which Passed Between the Late Learned Mr. Leibnitz and Dr. Clarke, In the Years 1715 and 1716* (London: 1717), extracted <http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00236>, 15/01/2013

⁹¹ Baruch Spinoza, (1677), *Ethics*, (Trans.) R.H.M Elwes, (1883), (MTSU Philosophy Webworks Hypertext Edition), extracted <http://capone.mtsu.edu/rbombard/RB/Spinoza/ethica-front.html> 17/01/2013, Book II, P1

P2. Extension is an attribute of God, or God is an extended thing.⁹²

In Part I, we have already learned that:

P14. Besides God no other substance can be granted or conceived.⁹³

And ‘attribute’ has been defined as: “By ‘attribute’ I mean that by which the intellect perceives as constituting the essence of substance”.⁹⁴ Thus, thought and extension are essentials constituents of the one substance there is, God. God may be identified with each attribute as a thinking substance or an extended substance, nevertheless, there is an identity relation between: thinking substance = extended substance. “Substance thinking and substance extended are one and the same substance.”⁹⁵

This being the case:

P7. The order and connection of ideas is the same as the order and connection of things.⁹⁶

Such that:

⁹² Ibid. Book II, P2

⁹³ Ibid. Book I, P14

⁹⁴ Ibid. Book I, Def. 4

⁹⁵ Ibid. Book II, Note to P7

⁹⁶ Ibid. Book II, P7

Corollary: [W]hatsoever follows from the infinite nature of God in the world of extension, follows without exception in the same order and connection from the idea of God in the world of thought.⁹⁷

This extends to causality, too. However, it only makes sense to *explain* each order in terms of the attribute in question, according to Spinoza. That is, *intra-active* explanations explain the chain of causality in any given substance.

[S]o long as we consider things as modes of thinking, we must explain the order of the whole of nature, or the whole chain of causes, through the attribute of thought only. And, in so far as we consider things as modes of extension, we must explain the order of the whole of nature through the attributes of extension only; and so on, in the case of other attributes.⁹⁸

Thus, even if extension and thought are essential attributes of one substance and substance considered as either extended or thinking is identical, whereby extended/thought order, connection, and causes are the same, *explanation* of the respective order and causes of nature are restricted to the relevant attribute. In a general sense, since God is the one and only substance and thought and extension essential constituents of God, Spinoza's parallelism extends to a panpsychism as Fechner's does. In another sense, Spinoza is a property dualist insomuch as each attribute is explicable solely in its own terms and is in

⁹⁷ Ibid. Book II, Corollary to P7

⁹⁸ Ibid. Book II, Note P7

that sense causally irreducible.

Epiphenomenalism

Thomas H. Huxley is usually taken to be a primary proponent of epiphenomenalism:

Our mental conditions are simply the symbols in consciousness of the changes which take place automatically in the organism; . . . to take an extreme illustration, the feeling we call volition is not the cause of a voluntary act, but the symbol of that state of the brain which is the immediate cause of that act. We are conscious automata.⁹⁹

More recently, Frank Jackson has defended a version of epiphenomenalism. Jackson argues that certain mental properties, the qualitative aspects of experience, qualia, are not efficacious with respect to the physical world.¹⁰⁰ His defence involves countering some common attacks on epiphenomenalism in order to show that there are no “knockdown arguments” against holding the position. First, it just seems obvious that qualia effects behaviour. Pain, for example, most often than not leads to avoidance behaviours, right? Not necessarily. Jackson says that an overarching theory may explain why both the qualia and the behaviour occur together. The one needn’t be the cause of the other. Rather, it

⁹⁹ T.H. Huxley, (1992), "On the Hypothesis That Animals Are Automata," in Brian Beakley and Peter Ludlow (Eds.), *The Philosophy of Mind*, ed. (Cambridge, MA: the MIT Press, 1992), p.136

¹⁰⁰ Frank Jackson, (1982), “Epiphenomenal Qualia” in (Eds.) Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, (2004) *There’s Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson’s Knowledge Argument*, (Cambridge, Massachusetts: The MIT Press), p.46

may simply be “a consequence of the fact that certain happenings in the brain cause both.”¹⁰¹ Second, if qualia are not efficacious, why did they evolve?¹⁰² To this Jackson replies, they may have evolved as a concomitant of some other efficacious characteristic as a by-product.¹⁰³ Third, how about the problem of other minds? How can we know that others have qualia? Ultimately, Jackson thinks this problem is not a special problem for epiphenomenalism but one that is much broader. It effects, for example, interactionism, too.¹⁰⁴ Last, qualia seem to do nothing and explain nothing according to epiphenomenalism, so how do they fit into the scientific world picture? On this, Jackson takes a position very much like McGinn’s. He argues that we have a limited capacity to understand the why of qualia “For the simple reason that such knowledge and understanding is irrelevant to survival”.¹⁰⁵

William Seagers suggests the SWP leads to a “generalised epiphenomenalism”. He imagines a perfect simulation of the universe, a fully developed and mathematically complete detailed computer simulation of physical systems. This allows Seagers to define emergence in the following terms: “The simulation thought experiment can be used to provide a simple and clear definition of emergence. An emergent is anything that is not coded into the simulation. Thus a thunderstorm is an emergent entity since, I take it, we

¹⁰¹ Ibid. p.46

¹⁰² The evolutionary reaction to epiphenomenalism was made famous by William James, see William James, (1879), “Are We Automata?”, extracted 15/08/2013, from <http://psychclassics.yorku.ca/James/automata.htm>

¹⁰³ Op cit. Frank Jackson, (1982), pp.46-47

¹⁰⁴ Ibid. p.48

¹⁰⁵ Ibid. p.48

would not need, in addition to coding in the quarks, leptons and bosons and their properties, to add thunderstorms as such to our simulation code.”¹⁰⁶ Temperature, being alive, being hydrogen are other properties which would not need to be coded into the computer simulation. Contrasted to “radical emergence” which claims that the aforementioned properties cannot be reduced to those of physics, this leads to a “benign emergence”, which Seagers defines in the following manner: “Benign emergence is simply the claim that all features not coded into the simulation are subject to resolution under closure.”¹⁰⁷

Now Seagers argues this threatens generalized epiphenomenalism. He offers three arguments. First, the economy argument: all we need to duplicate the world is the most basic level of description, anything else is superfluous.

Imagine the fundamental physics simulation of peg approaching hole. There is no need to code into the simulation anything about squareness or roundness, or whether something is a peg and something else is a hole, or that the peg is moving towards the hole or anything else at a level of description above that of fundamental physics. Nonetheless the world of the simulation reveals that the peg won't go through the hole.¹⁰⁸

Second, the screening off argument: It is hard to distinguish a cause from a mere correlate.

¹⁰⁶ Op cit. William Seagers, (1999), p.7

¹⁰⁷ Ibid. p.8

¹⁰⁸ Ibid. p.9

Here is one way to do it:

(C1) $P(A|C \ \& \ B) = P(A|C)$ ¹⁰⁹, but

(C2) $P(A|C \ \& \ B) \neq P(A|B)$.¹¹⁰

Although this test is not fool-proof, it is often effective. It helps screen off B and establish C as the real cause of A. A nice example that Seagers gives is the false causative relation between cancer and coffee.

The screening off test reveals this since the statistics end up as follows:

$P(\text{cancer}|\text{smoking} \ \& \ \text{coffee}) = P(\text{cancer}|\text{smoking}) \neq P(\text{cancer}|\text{coffee})$.¹¹¹

He notes, though, “The weakness of the test is nicely revealed in this example too, for it is evidently *possible* that absolutely every coffee drinker should be a smoker and vice versa.”¹¹² Nevertheless, Seagers argues that screening off will relegate psychology, for example, the belief that I know the answer, to be replaced by the basic physics, in relation to, for example, raising my arm in class.

Third, the abstraction argument: explanation tends towards abstraction, but abstractions can’t really be said to have causative powers. He writes,

¹⁰⁹ This is read: the probability of A given C and B equals the probability of A given C.

¹¹⁰ Ibid. p.14

¹¹¹ Ibid. p.14

¹¹² Ibid. p.14

[W]hatever mathematical structure underlies the transition from neural to mental states, it will be a mathematical abstraction from the underlying properties... Insofar as mental states are seen as such mathematical abstractions they cannot be granted causal efficacy.¹¹³

They, like temperature, Seagers says will just be “mathematically convenient ways of thinking about the mass actions of the fundamental physical constituents”.¹¹⁴ From these three arguments, Seagers thinks that the basic physical description of the world will create a benign emergence leading to a generalized epiphenomenalism, not just minds, but thunderstorms, etc.¹¹⁵

There is even some scientific evidence for epiphenomenalism from the work of Benjamin Libet. First, we need to understand what an “RP” is. An RP is a "readiness potential". It is the "electrical indication of certain brain activities" preceding actual movements. Libet's study observed these RPs before voluntary acts were performed. Subjects were allowed to perform "capriciously", that is, without any restrictions on when they flicked a wrist. The result was, “RPs in these acts began with onsets averaging 550 msec. before

¹¹³ Ibid. p.21

¹¹⁴ Ibid. p.21

¹¹⁵ It should be noted that there is a sting in the tail, for Seagers ultimately suggests, given the mind dependent nature of what is emergent, the mind cannot be reduced. Op cit. William Seagers, (1999), pp. 23-24

activation of the involved muscle”¹¹⁶ The conclusion was that “The brain was evidently beginning the volitional process in this voluntary act well before the activation of the muscle that produced the movement.”¹¹⁷ What had to be checked next was when the subjects recognised the urge or wish to flex their wrists. The results showed: “For groups in which all the voluntary acts were freely spontaneous, with no reports of rough preplanning of when to act, the onset of RP averaged –550 msec. (before the muscle was activated). The W times [the times at which subjects became aware of the wish to act] for first awareness of wish to act averaged about –200 msec”¹¹⁸ *There was a 350ms gap between the RP and the recognition of the urge or will to act.* Recalculating for errors, the time grew to 400ms. And, Libet thinks it's even greater than this because the RP probably starts off at a place in the brain different from the place at which it is recorded.¹¹⁹ The act could be vetoed because one only needs 100ms to veto the act and subjects could do so and we all have experiences of doing so.¹²⁰ But, of course, veto, itself, may be involuntary.¹²¹ If this is so, epiphenomenalism threatens, for it seems that the brain is the centre of action and decision, of which we become aware at some later time. Libet, himself, rejects this and thinks that there is an element of free-will in play here, since he is keen to protect the attribution of responsibility to minded beings.¹²²

¹¹⁶ Benjamin Libet, (1999), “Do We Have Free Will”, in *Journal of Conscious Studies*, 6, No. 8-9, 47-57, p.49

¹¹⁷ Ibid. p.49

¹¹⁸ Ibid. p.50

¹¹⁹ Ibid. p.51

¹²⁰ Ibid. p.51

¹²¹ Ibid. p.52

¹²² See also, Benjamin Libet, (2005), *Mind Time: The Temporal Factor in Consciousness*, (Cambridge, Massachusetts: Harvard University Press)

Now, to return to the denials posited above, (D3) may be interpreted in a number of ways, too. For example, one may accept that (D3) is apt to apply at the level of concepts or properties, but deny that (D3) applies at any substantial level. That's commonly known as property dualism. In fact, generally, complicating matters are the distinctions between the levels at which the denials of monism are accepted. We noted a distinction between conceptual, property and substantial levels just now. Relatedly, *epistemological* and *metaphysical* categories may need to be distinguished. This means the conditions above may be read in an epistemological flavour, but read in another way at a metaphysical level. In point of fact, there may be many divisions to accept, causal, biological, linguistic, etc.

Take Donald Davidson's argument for anomalous monism:

[A]t least some mental events interact causally with physical events. (We could call this the Principle of Causal Interaction).¹²³

Followed by:

The second principle is that where there is causality, there must be a law: events related as cause and effect fall under strict deterministic laws.¹²⁴

¹²³ Donald Davidson, (1970), "Mental Events" in Donald Davidson, (2001), *Essays on Actions and Events*, (Oxford: Clarendon Press), p.207

¹²⁴ *Ibid.* p.207

Then, Davidson tells us that:

The third principle is that there are no strict deterministic laws on the basis of which mental events can be predicted and explained (the Anomalism of the Mental).¹²⁵

On the other hand:

Physical theory promises to provide a comprehensive closed system guaranteed to yield a standardized, unique description of every mental physical event couched in vocabulary amenable to law.¹²⁶

Davidson concludes that the identity between the mental and the physical is easy to establish:

The demonstration of the identity follows easily. Suppose *m*, a mental event, caused *p*, a physical event; then under some description *m* and *p* instantiate a strict law. This law can only be physical...But if *m* falls under a physical law, it has a physical description; which is to say that it is a physical event. An analogous argument works when a physical event causes a mental event. So every mental event that is causally related to a physical event is a physical event.¹²⁷

¹²⁵ Ibid. p.207

¹²⁶ Ibid. pp.222-223

¹²⁷ Ibid. p.223

We won't know what mental event pairs with what physical event (given there are no psychophysical laws), but every mental event is a physical event (if it is causally related to the physical world, which is assumed from the start):

[W]e see that it is possible to know that a mental event is identical with some physical event without knowing which one.¹²⁸

This theory is non-reductive at the level of psychology since the mental is anomalous, but it is monistic at the ontological level, since every mental event is identical with some physical event.

Now, compare John Searle's "Biological Naturalism". He writes:

Our world picture, though extremely complicated in detail, provides a rather simple account of the mode of existence of consciousness. According to the atomic theory, the world is made up of particles. These particles are organized into systems. Some of these systems are living systems, and these types of living system have evolved over long periods of time. Among these, some have evolved brains that are capable of causing and sustaining consciousness. Consciousness is, thus, a biological feature of certain organisms in exactly the same sense of 'biological' in which photosynthesis, mitosis, digestion, and reproduction are biological features of organisms.¹²⁹

¹²⁸ Ibid. p.223

¹²⁹ Op cit. John Searle, (1994), p.93

More succinctly: “consciousness is a feature of the brain and thus part of the physical world”.¹³⁰ But, and this is where a dualistic aspect seems to come into play:

[I]n the case of consciousness we can make a causal reduction but we cannot make an ontological reduction without losing the point of having the concept.¹³¹

So there do seem to be two distinct ontological spheres, there is a first-person ontology, which relates to consciousness and intentionality, and a third-person ontology, which relates to everything else. Doesn't this sound like a kind of dualism? Searle says no, because causal reduction to the physical is possible. Well, as with Davidson, it seems to me that Searle retains a dualistic stance at one level of discourse.

As we can see the forms dualism takes are various. But what are the arguments for dualism in the first place? Returning to the notions of completeness, closure and resolution, it seems to me that the most important arguments against physicalism, aim to undermine resolution because that also undercuts completeness. That is, the arguments for dualism tend to deny that the mental is, or is reducible to, the physical. It is these arguments that I now turn to.

¹³⁰ Op cit. John Searle, (2004), p.80

¹³¹ Ibid. p.83

Section C: The Arguments for Dualism

As said the most useful form of attack for the dualist is to deny that resolution is possible. If the mental is not reducible to the physical in any way, then the physicalist mirror of the world doesn't reflect everything. And one may legitimately think that the physicalist picture of the world is incomplete. Thus, lack of reduction implies lack of completeness. Of course, this kind of irreducibility must occur at a significant level if it is not to slide back into monism. Here are some arguments that aim for just this kind of denial.

The Knowledge Argument (Frank Jackson)

Frank Jackson formulates this argument in the following fashion:

Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room via a black and white television monitor. She specializes in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like 'red', 'blue', and so on. She discovers, for example, just which wave-length combinations from the sky stimulate the retina, and exactly how this produces via the central nervous system the contraction of the vocal chords and expulsion of air from the lungs that results in the uttering of the sentence 'The sky is blue.'

What will happen when Mary is released from her black and white room or is

given a colour television monitor? Will she *learn* anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had all the physical information. Ergo there is more to have than that, and Physicalism is false.¹³²

Let's sum up: At time t1 Mary knows all the physical information about the world and does not know what it's like to see red and at time t2 after encountering, say, a red apple she still knows all the physical information about the world but in addition she knows what it's like to see red. Therefore, Mary would learn something new if she saw red for the first time even if she had all the knowledge about the physical world at her disposal. Therefore, not all truths are physical truths.

Ludlow et al. formulate the argument in the following manner:¹³³

- (1) Every physical truth is such that Mary (before her release) knows that truth.
- (2) It is not the case that every truth is such that Mary (before her release) knows the truth.
- (3) Therefore, there is at least one truth that is nonphysical.

There are some responses to this argument some of which I briefly go over. Most time is spent on the property dualist's response because it helps in understanding the original argument more fully and it helps introduce David Chalmers to the debate – his

¹³² Op cit. Frank Jackson, (1982), "Epiphenomenal Qualia", pp.42-43

¹³³ Op cit. Peter Ludlow et al. (2004), p.14

conceivability argument is another one of the arguments I introduce in this section. First, then, there is just the plain denial that Mary does learn anything new. This response is held by Dennett.¹³⁴ More recently, Frank Jackson himself argues along these lines. He now argues that there is an illusion that Mary learns something new based on the way the information at t2 is presented to her:

[Mary] is in a new kind of representational state, different from those she was in before. And what is it to know what it is like to be in that kind of state? Presumably, it is able to recognize, remember, and imagine the state.¹³⁵

In effect, this is the ability response to the argument. This is summed up very nicely by David Lewis, who following Nemirow, responds in precisely this way to Jackson's argument:

The ability hypothesis says that knowing what an experience is like *is* the possession of these abilities to remember, imagine, and recognize. It isn't the possession of any kind of information, ordinary or peculiar. It isn't knowing that certain possibilities aren't actualized. It isn't knowing-that. It's knowing-how. Therefore it should be no surprise that lessons won't teach you what an experience is like. Lessons impart information; ability is something else.

¹³⁴ See Daniel Dennett, (1991b), *Consciousness Explained*, (New York, Boston and London: Back Bay Books, Little, Brown and Company), especially pp.398-406

¹³⁵ Frank Jackson, (2003) "Mind and Illusion", in Peter Ludlow, (2004), p.439

Knowledge-that does not automatically provide know-how.¹³⁶

Indeed, Jackson says, “We have ended up agreeing with Laurence Nemirow and David Lewis on what happens to Mary on her release.”¹³⁷ Although, he remarks he doesn’t know how they reached their conclusions. It seems to me though talking about an ability is more or less to talk about a function. Further, it has been argued by philosophers that none of the abilities talked about are necessary or sufficient for knowing what it’s like to see red.¹³⁸

Another perceived threat comes from the property dualist mentioned above. For example, we can explain the apparent lack of knowledge away if we take a view akin to anomalous monism. That is, we can say, “it is no surprise that the set of all physical truths does not imply the set of all psychological truths insofar as psychology is anomalous, that is, there are no strict and determinate psychophysical laws”. *Nevertheless*, such truths are physical. Recall what Davidson says:

[W]e see that it is possible to know that a mental event is identical with some physical event without knowing which one.¹³⁹

¹³⁶ David Lewis, (1988), “What Experience Teaches”, in Peter Ludlow et al., (2004), p.100

¹³⁷ Op cit. Frank Jackson, (1982), p.439

¹³⁸ See Michael Tye, (2000), “Knowing What it is Like: The Ability Hypothesis and the Knowledge Argument,” in M. Tye, (2000), *Consciousness, Color, and Content*, (Cambridge, Massachusetts: The MIT Press); Martine Nida-Rumelin, (1995), “What Mary Couldn’t Know: Belief About Phenomenal States”, in (Ed.) Thomas Metzinger, (1995), *Conscious Experience*, (Exeter, UK: Imprint academic), pp.219-241

¹³⁹ Op cit. Donald Davidson, (1970), p.223

Another way of generating this objection comes from Ludlow et al. in terms of supervenience, psychophysical conditionals and the necessary a posteriori. First, they explain supervenience in terms of the aforementioned psychophysical conditional:

Suppose we gather together all the physical truths of the world into one megatruth, *P*. And suppose we do the same with all the psychological truths to produce one psychological megatruth, *Q*. Now consider the conditional formed from *P* and *Q*—if *P* then *Q*—and call this conditional *the psychophysical conditional*. To say that the psychological nature of the world supervenes on its physical nature is to say or imply that the psychophysical conditional is necessarily true. Given the supervenience account of physicalism, therefore, the following conditional is true: if physicalism is true, the psychophysical conditional is necessarily true.¹⁴⁰

They then introduce Kripke's notion of the necessary a posteriori. For example, it is commonplace to think that water is necessarily H₂O even though it is not known to be so a priori because even if the relation in question is a necessary one, one needs to discover the truth of that relation empirically. The psychophysical conditional in question may also be necessary but not a priori in the same way, such that:

[I]t would be reasonable to say that *Q* is a distinct truth from *P*, even if they are

¹⁴⁰ Daniel Stoljar and Yujin Nagasawa, "Introduction", in Peter Ludlow et al. (2004), pp.14-15

necessarily connected.¹⁴¹

Physicalism is not defeated on this account because what Mary learns cannot be derived a priori from the set of all physical truths. This leads Ludlow et al. to remark that Jackson must be construed as supposing, in light of this problem and with reference to his later work: [N]ot only that: if physicalism is true the psychophysical conditional is necessary, but also that: if physicalism is true, the psychophysical conditional is a priori.¹⁴²

Given this assumption, Jackson can conclude, according to Ludlow et al:

[I]f physicalism is true, there is *no* other contingent truth from which the psychophysical conditional can be derived. For physicalism aims at completeness...[So] if physicalism is true, the psychophysical conditional is not only necessary but also a priori.¹⁴³

Therefore, if Mary did know all the physical truths about the world, she could have derived the psychological truths a priori. But she wasn't able to. So not all truths are physical truths or entailed by physical truths. The common ground here is that Jackson must deny, against Davidson, for example, that psychology and psychophysical laws are anomalous. Here is one reason that Jackson gives, which he introduces in relation to the reduction of the thermodynamic theory of gases to the kinetic theory.

¹⁴¹ Ibid. p.15

¹⁴² Ibid. p.15

¹⁴³ Ibid. p.15

Scepticism about gases having temperature and pressure threatens if we insist that we cannot go a priori from the molecular account of gases and the concomitant functional roles to gases have temperature and pressure.

This point is implicit in the well-known schematic account of why it is right to identify temperature in gases with mean molecular kinetic energy:

Temperature in gases is that which does so and so (a priori premise about the concept of temperature).

That which does so and so is mean molecular kinetic energy (empirical premise).

Therefore, temperature in gases is mean molecular kinetic energy.

He continues: “unless something like the first premise is a priori, eliminativism about temperature and pressure in gasses is inevitable.”¹⁴⁴ Here it might seem that Jackson provides grounds to think, psychophysical truths are a priori and, so, if there are psychological truths, even if discovered empirically, they should be able to be known a priori.

Well, Jackson *and* Chalmers argue that the physical description, if complete, *would* a priori entail knowledge of all the *facts*. They both use a “two-dimensional” framework in

¹⁴⁴ Op cit. Frank Jackson, (2003), p.424

order to argue for this and they have *jointly* defended this position.¹⁴⁵ I'll introduce Chalmers 2-D framework at this point in order to try and make sense of why they think like this.

Chalmers thinks that a statement, S, can be said to be true in two different ways. S can be said to be true relative to *a* world taken as actual. Epistemic possibility is assigned relative to S being true in that world, and necessity relative to S being true in any world taken as actual. On the other hand, S can be taken to be true relative to *the* world. Metaphysical possibility is, then, assigned relative to S being true in a world taken as counterfactual relative to *the* world and necessity relative to S being true in any world taken as counterfactual relative to *the* world. The intension of S is associated with a function that assigns a truth-value to S at w. So far as there are two ways to assign truth to S, there are two functions, epistemic-intensions (1-intensions) and subjunctive-intensions (2-intensions). If S is assigned truth in some possible world taken as actual by a 1-intension, then S is epistemically possible. If S is assigned truth in any possible world taken as actual by a 1-intension, then S is epistemically necessary. If S is assigned truth in some possible world taken as counterfactual by a 2-intension, then S is metaphysically possible. If S is assigned truth in every possible world taken as counterfactual, then S is metaphysically necessary. Often epistemic possibility and metaphysical possibility are inversely related. Consider the following statement:

(1) Water is not H₂O

¹⁴⁵ David Chalmers and Frank Jackson, (2001), "Conceptual Analysis and Reductive Explanation", in *The Philosophical Review*, 110, 315-361

This statement is, in the epistemic sense, *merely* possible. That is, there is some possible world taken as actual in which the statement is true. That is, there is a 1-intension that assigns (1) truth in a possible world but not all possible worlds. So it is 1-contingent. However, (1) is, in the subjunctive sense, impossible. That is, there is no possible world taken as counterfactual in which the statement is true relative to *the* world. That is, there is a 2-intension that assigns (1) not true in every possible world. Its falsehood is, therefore, 2-necessary. The truth or falsity of statements like (1) is discovered empirically. This is to say that they are a posteriori, so the truth or falsity of these statements is not known a priori. Previous to 1750 the molecular structure of water was not known and nothing but empirical research could have uncovered it. This suggests that statements like (1) are always 1-contingent. Chalmers calls this the 2-D thesis:

(2-D Thesis) If S is a posteriori, S is 1-contingent¹⁴⁶

This in turn suggests:

(2-D+) If S is not 1-contingent, S is not a posteriori

Assuming that ‘not a posteriori’ is equivalent to ‘being a priori’ and ‘not contingent’ equivalent to ‘being necessary’, that suggests:

¹⁴⁶ David Chalmers, (2004), “Phenomenal Concepts and the Knowledge Argument”, in Peter Ludlow et al., (2004), p.281

(2-D++) If S is 1-necessary, then S is a priori

Let's consider a new statement, a 1-necessary statement. What does it look like?

Continuous with Jackson's writings above it looks like this:

(2) Water is that which acts in the way water does

That is, where 'that which acts in the way water does' is associated with a set of descriptions including the various ways water tantalizes our senses and the roles it plays in the physical environment. This statement is, as said, in the epistemic sense, necessary. That is, in all possible worlds taken as actual the statement is true. A 1-intension assigns (2) truth in any possible worlds taken as actual. So it is 1-necessary. However, in the subjunctive sense, it is contingent. That is, there are some possible worlds taken as counterfactual in which the statement is false relative to *the* world. That is, there is a 2-intension that does not assign (2) truth in every possible worlds. It is 2-contingent on this basis. The truth of statements like (2) are not discovered empirically. This is to say that they are a priori as said above.

Now, there are a class of statements that have the same epistemic and subjunctive assignments of truth. Consider, then:

(3) Pain is felt-pain

This statement is, in the epistemic sense, necessary. That is, in every possible world taken

as actual the statement is true. That is, there is a 1-intension that assigns (3) truth in all possible worlds taken as actual. *And*, (3) is, in the subjunctive sense, necessary. That is, in every possible world taken as counterfactual the statement is true. That is, there is a 2-intension that assigns (3) true in every possible worlds relative to *the* world. We can note the truth of statements like (3) seem to be known a priori. Well, this would follow from the 2-D thesis, given (3) is not 1-contingent.

Presently, the relation, especially between (1) and (2) seems to be important to the knowledge argument. Chalmers and Jackson want to tell something like the following story: First, we know a priori that a statement like (2); empirical research and advances in the sciences allow us to fill in the blank space represented by the clause ‘that which acts in the way that water does’ and say what water is; since it is H₂O that is that which acts in the way that water does, water is H₂O. So the reason why once one has all the physical facts one can derive all the physical truths is because the blank spaces, the ‘that which acts in the way such and such does’, are filled in by the contingent (1-contingent) truths discovered empirically by the advanced sciences, given that the set of physical truths are fully-furbished with all the contingent (1-contingent) truths one can derive all the physical truths. Generally, I think, this is the correct view to take. And so, on a Kripkean account, too, where facts determine modality and the way in which they determine modality is a priori; the a posteriori discoveries of the facts allows the logical derivations to go through, for example, we know, $A \rightarrow \Box A$, $\Diamond \neg A \rightarrow \Box \neg A$, etc. a priori according to the logic involved. Thus the discovery of an a posteriori fact like $\neg A$, entails the conclusion that $\neg \Diamond A$.¹⁴⁷ So, even if $P \rightarrow Q$ is discovered a posteriori, given all the

¹⁴⁷ See, for example, section one of chapter one; “A” is an essentialist statement here.

physical truths, including contingent gap-fillers, one should be able to derive Q from P a priori if entailed by P. But, it seems, to Jackson and others, that Mary wouldn't be able to do the sums and make the required derivation. Next to Chalmers' argument.

The Conceivability Argument (David Chalmers)

The idea that $P \rightarrow Q$ can be used to kick start an argument that argues against physicalism working on a 2-D framework. Here it is:

1. $P \rightarrow Q$ is a posteriori
2. ($P \rightarrow Q$ is a posteriori) \rightarrow ($P \rightarrow Q$ is 1-contingent)
3. ($P \rightarrow Q$ is 1-contingent) \rightarrow ($P \rightarrow Q$ is 2-contingent)
4. ($P \rightarrow Q$ is 2-contingent) \rightarrow Physicalism is false
5. Therefore, physicalism is false.¹⁴⁸

The first premise accepts the relation between P and Q is discovered empirically. The second premise is based on the 2-D thesis. The third premise is related to the possibility of statements like (3) above. It can be argued for in terms of "semantic neutrality".¹⁴⁹

Chalmers says:

¹⁴⁸ It is assumed that P abbreviates physicalism in the simplest way so that it allows for everything that is entailed by it in the actual world and no more. See David Chalmers, (1997), *The Conscious Mind: In Search of a Fundamental Theory*, (New York and Oxford: Oxford University Press), p.42 Q abbreviates a qualitative truth.

¹⁴⁹ See David Papineau, (2007), "Kripke's Proof is Ad Hominem not Two-Dimensional", in *Philosophical Perspectives*, Vol. 21, 475-494

It is not true in general that 1-contingent statements are 2-contingent: counterexamples include ‘water is H₂O’, and ‘Hesperus is Phosphorus’. The reason is that expressions such as ‘water’ and ‘Hesperus’ have quite different epistemic and subjunctive intensions. However, the principle is true for statements including only *semantically neutral* expressions, whose epistemic intensions are the same as their subjunctive intensions.¹⁵⁰

But what exactly are “semantically neutral expressions”? Semantically neutral expressions are those expressions that have the same epistemic and subjunctive intensions. I have already provided an example above, ‘Pain is the sensation of pain’. Chalmers, who thinks the notion is quite intuitive, gives the following examples:

[W]e have a good grasp on the notion. For example, ‘water’ and ‘Hesperus’ are not semantically neutral; but ‘and’, ‘philosopher’, ‘friend’, ‘consciousness’, and ‘cause’ plausibly are.¹⁵¹

One test for semantic neutrality is this: semantically neutral expressions are those that are immune to twin-world thought experiments. Chalmers says, “One promising approach is to define such an expression as one that is not “Twin-Earthable”.”¹⁵² A twin-world example conceives with respect to a term, for example, ‘water’, that two speakers, here,

¹⁵⁰ Op cit. Chalmers, (2004), p.282

¹⁵¹ David Chalmers, (2006), “The Foundations of Two-Dimensional Semantics”, in (Eds.) Manuel Garcia-Carpintero and J. Macia, (2006). *Two-Dimensional Semantics: Foundations and Applications*, (Oxford: Clarendon Press) p.86

¹⁵² Ibid. p.86

Ryota and Shota, share the same epistemic sense of the terms, as associated with the set of descriptions that describe all the ways in which water plays on our senses, but that Ryota's term 'water' has a different extension to Shota's term 'water'. Ryota's term picks out H₂O, Shota's XYZ. A term like 'pain', on the other hand, is immune to such examples. There is *no* world in which, given Ryota and Shota share the same epistemic sense associated with the term, for example, the feeling of pain, the extension of the term as Ryota uses it is different from the extension of the term as Shota uses it; that is, the term picks out felt-pain in all possible worlds bar none.

This can be put in terms of epistemic intensions and subjunctive intensions. Consider, a set of possible worlds that may either be taken as actual or counterfactual. Two dwellers, Ryota and Shota are assumed. Each is associated with a distinct possible world. The term in question is 'water'. A function connects the term 'water' to whatever fills the watery role. Ryota is associated with a possible world in which a function connects the term 'water' to H₂O (actual water). Shota is associated with a possible world in which a function connects the term 'water' to XYZ (non-actual water). The set of possible worlds may be taken as actual. The set of possible worlds may be taken as counterfactual. Having the same *epistemic* intension is having the same psychological state. Having the same subjunctive intension is being related to the same extension. If we take the set of possible worlds as actual, the term has the same epistemic intension but a different subjunctive function. That is, 'water is whatever plays the water role' and either 'water is H₂O' or 'water is XYZ'. The first statement is necessary and the latter two are contingent. That is, 'water is whatever plays the water role' is 1-necessary, but, for example, 'water is H₂O' is 1-contingent. That is, in effect, the first is a priori necessary the latter is a priori

contingent. If we take the set of possible worlds as counterfactual, the term has the same subjunctive intension but a different epistemic function. That is, 'water is H₂O' is necessary, but 'water is whatever plays the water role' is contingent (because non-actual water may well play that role somewhere). The former is necessary, or 2-necessary, and the latter is contingent, 2-contingent. In other words, the former is a posteriori necessary, the latter a posteriori contingent. A twin-world example can be defined as playing on the necessary 1-intension and the necessary 2-intension. That is, Ryota and Shota share a necessary 1-intension relative to possible worlds taken as actual but do not share a necessary 2-intension relative to the world. So, Ryota and Shota's epistemic situation is the same, but not their relation to the actual world. Finally, Ryota and Shota share a similar psychology but not extension with respect to the term in question.

Now consider the following: Again, a set of possible worlds that may either be taken as actual or counterfactual, two dwellers, Ryota and Shota, are associated with a distinct possible world each. This time, however, consider the term 'pain'. It is not possible to distinguish functions such that the term 'pain' is associated with 'the phenomenal role of pain' and functions connecting Ryota to a possible world in which the term 'pain' picks out felt-pain (actual pain), playing the phenomenal role of pain, and Shota to a possible world in which 'pain' does not pick out felt-pain with something else playing the phenomenal role of pain, for whatever fills the phenomenal role of pain *is* felt-pain! Therefore, no twin-world example as defined above can be generated because Ryota and Shota *share* a necessary 1-intension relative to possible worlds taken as actual *and* they *share* a necessary 2-intension relative to *the* world. That is, Ryota and Shota's epistemic situation is the same and their relation to the actual world is the same. Allowing us to

conclude that Ryota and Shota share a similar psychology and a similar extension with respect to the term in question.

However, this method is not fool-proof:

This test works for many purposes. A semantically neutral term (in the intuitive sense) is never Twin-Earthable. But the reverse is not quite the case. For example, let L be an expression that functions to rigidly designate the speaker's height. Then any twin of L will have the same 2-intension (since a twin speaker will have the same height), but L is not semantically neutral. One might respond by watering down the requirements of physical and phenomenal duplication (perhaps to some sort of mental duplication), but similar cases will still arise: e.g. if M is an expression that rigidly picks out 1 if the speaker has visual experience, and 0 if not, then M is not Twin-Earthable even by this sort of standard, but it is nevertheless not semantically neutral.¹⁵³

Well, the test lets one get a handle on semantic neutrality; still, it remains fairly vague. So, ultimately, we must rely upon our intuitive grasp of the notion alone for the time being: "A precise formal characterization of semantic neutrality remains an open question for future research."¹⁵⁴ For present purposes all we need to note is, for example, if A is a neutrally semantic expression, and A is 1-necessary, then A is 2-necessary. For another example, if A and B are semantically neutral expressions, and $A \rightarrow B$ is 1-contingent,

¹⁵³ Ibid. pp.86-87

¹⁵⁴ Ibid. p.87

then $A \rightarrow B$, is 2-contingent.¹⁵⁵ This leads Chalmers to say, in relation to the argument posited above, “If P...and Q were semantically neutral, premise (3) would be true”.¹⁵⁶ However, although Chalmers thinks that Q is semantically neutral, he has doubts about P:

[I]t is arguable that P is not semantically neutral. It is plausible that terms for microphysical properties, such as ‘charge’, refer rigidly to intrinsic properties, but pick out those properties by virtue of the fact that they play a certain causal role in our world...If so, then at any given world, the epistemic intension of ‘charge’ picks out whatever property plays the relevant causal role in the world, while the subjunctive intension picks out the intrinsic property (charge) that plays the causal role in *our* world. And it is arguable that these intensions differ, since there are arguably worlds where the relevant causal role is played by a property distinct from the property playing the role in our world. If so, premise (3) is false.¹⁵⁷

There is a way out of this impasse. Either P is a semantically neutral expression or it is not. If it is, then (3) is true and the argument is valid. If it isn’t, then physical terms refer to a deeper physical intrinsic reality than P describes. It may even be phenomenological

¹⁵⁵ For example, if P is 1-possible and semantically neutral and Q is 1-possible and semantically neutral and $P \rightarrow Q$, then $P \rightarrow Q$ is 1-possible and semantically neutral and, therefore, $P \rightarrow Q$ is 2-possible, for all semantically neutral says is that give an expression like $P \rightarrow Q$ if it is semantically neutral it is both 1-possible *and* 2-possible (or 1-necessary and 2-necessary as the case may be) at the same time.

¹⁵⁶ Op cit. Chalmers, (2004), p.282

¹⁵⁷ Ibid. p.283

(suggesting a conscious universe all the way down to the ground floor):

[I]t is coherent to suppose that these properties have a special nature that is tied to consciousness. They might themselves be phenomenal properties, or they might be protophenomenal properties: properties that collectively constitute phenomenal properties when organized in the appropriate way.¹⁵⁸

It will be this mysterious reality that determines both the physical relations described by P, but more importantly the phenomenological properties captured by Q. Standard physicalism is false on both accounts.

The argument here has been put in terms of epistemic and subjunctive intensions. However, it may be put in terms focusing on *conceivability* alone. Here it is:

- (1) $P \ \& \ \neg Q$ is conceivable
- (2) If $P \ \& \ \neg Q$ is conceivable, $P \ \& \ \neg Q$ is 1-possible
- (3) If $P \ \& \ \neg Q$ is 1-possible, $P \ \& \ \neg Q$ is 2-possible.
- (4) If $P \ \& \ \neg Q$ is 2-possible, materialism is false.
- (5) Materialism is false.

¹⁵⁸ David Chalmers, (2009), “The Two-Dimensional Argument Against Materialism”, in (Eds.) Ansgar Beckermann, Brian McLaughlin, and Sven Walter, (2009), *The Oxford Handbook of the Philosophy of Mind*, (Oxford: Oxford University Press), extracted <http://consc.net/papers/2dargument.html> 17/11/2013, Cf. Bertrand Russell, (1927), *The Analysis of Matter*, (London: Kegan Paul), and Daniel Stoljar, (2001), “Two Conceptions of the Physical”, in *Philosophy and Phenomenological Research* 62: 253-281

Let me fill in the details. Premise one uses a specific sense of conceivability. Conceivability comes in different forms and is subject to three different kinds of specification. First, prima facie and ideal conceivability, that is, conceivability on first appearances and conceivability on ideal rational reflection: "S is ideally conceivable when S is conceivable on ideal rational reflection."¹⁵⁹ Second, negative and positive conceivability, that is, something is conceivable because it is not ruled out according to some specification, for example, it isn't ruled out prima facie, or it isn't ruled out on ideal reflection, or a priori, etc. On the other hand, something is conceivable because, more or less, we can imagine it to be the case:

Positive notions of conceivability require that one can form some sort of positive conception of a situation in which S is the case. One can place the varieties of positive conceivability under the broad rubric of imagination¹⁶⁰

Third, primary and secondary conceivability. The former applies to something we can conceive to be true in possible worlds taken as actual, the latter applies to something we can conceive to be true in a possible world taken as counterfactual. Note, the former kind of conceivability is a priori: "Primary conceivability is always an a priori matter."¹⁶¹ The latter kind is obviously a posteriori. The kind of conceivability that applies to the first

¹⁵⁹ David Chalmers, (2002), "Does Conceivability Entail Possibility", in (Eds.) T. Gendler & J. Hawthorne, (2002), *Conceivability and Possibility*, (Oxford University Press), pp.145-200, extracted <http://consc.net/papers/conceivability.html> 17/11/2013

¹⁶⁰ Ibid.

¹⁶¹ Ibid.

premise is ideal positive primary conceivability. That is that $P \ \& \ \neg Q$ can be imagined to be true on ideal reflection when some possible world is considered to be actual or it can be imagined to be true on ideal reflection a priori.

Premise 2 moves from conceivability to possibility in a specific way. So far as primary conceivability is concerned, we can say that if something is primarily conceivable there is a possible world which taken as actual is evidence for possibility of what is conceived. But here we must be sure to understand, so far as we are talking about what is *primarily* conceivable, we are also only talking about what is *primarily* possible, not what is secondarily possible: "primary conceivability does not entail metaphysical possibility: 'water is not H₂O' is primarily conceivable, but it is not metaphysically possible."¹⁶²

We have already talked about epistemic, or 1-intensions, and subjunctive, or 2-intensions. These apply to statements, or for Chalmers, sentences, in terms of intensions, then, "we can say that while the secondary intension of 'water is not H₂O' is false at w, the sentence's primary intension is true there". This may come to something like there is some possible world that provides evidence for or verifies the truth of a sentence like the one mentioned, but no possible world in which it is true. What is important is this, according to Chalmers, "Ideal primary positive conceivability entails primary possibility".¹⁶³ So the second premise only allows us to go from what is 1-conceivable to what is 1-possible.

The third premise comes down to the idea behind semantic neutrality:

¹⁶² Op cit. David Chalmers, (2009)

¹⁶³ Ibid.

[I]t may be that the gap between 1-possibility and 2-possibility could be closed. In particular, when a statement S has the same primary intension and secondary intension, then a world will verify S iff it satisfies S, so S will be 1-possible iff it is 2-possible. If P and Q both have primary intensions that coincide with their secondary intensions, then so will $P \& \neg Q$...¹⁶⁴

And as above if P and Q are semantically neutral, the premise is true and the argument valid. Again, if the problem is with P, then materialism has found itself positing exotic intrinsic properties and these properties may even be, as we saw above, phenomenal.

Other Arguments (Sydney Shoemaker, Ned Block)

Sydney Shoemaker taking his cue from John Locke and Ludwig Wittgenstein wrote a now famous paper on the inverted spectrum. Shoemaker draws upon the following passage in Wittgenstein:

Consider this case: someone says “I can’t understand it, I see everything red blue today and vice versa. “ We answer “it must look queer!” He says it does and, e.g., goes on to say how cold the glowing coal looks and how warm the clear (blue) sky. I think we should under these or similar circumstances be inclined to say that he saw red what we saw blue. And again we should say that we know that he means by the words ‘blue’ and ‘red’ what we do as he has always used them

¹⁶⁴ Ibid.

as we do.¹⁶⁵

Shoemaker labels this as a case of “*intra*-subjective spectrum inversion”.¹⁶⁶ Although, Shoemaker notes that Wittgenstein rejects the idea of *inter*-subjective spectrum inversion, noting that Wittgenstein’s rejection is obscure, he thinks the possibility of the former legitimizes the possibility of the latter:

[I]t seems, offhand, that if intrasubjective spectrum inversion is possible, intersubjective inversion must also be possible. For suppose that someone, call him Fred, undergoes intrasubjective inversion at *t*. Assuming that others also did not undergo intersubjective inversion at *t*, it would seem that either before *t* or after (or both) Fred’s colour experience must have been radically different from that of others.¹⁶⁷

Shoemaker considers a few objections. For example, perhaps the similarity/difference relation is well defined only for the intrasubjective case? Yet, the problem here is that we cannot say of others that they enjoy (or do not enjoy) the same colour experiences as we do. Shoemaker also rejects that the sufferer of inversion is simply suffering from aphasia or memory loss. He thinks it is possible to imagine a situation that rules these causes out

¹⁶⁵ Ludwig Wittgenstein, (1968), “Notes for Lectures on ‘Private Experience’ and ‘Sense Data’”, (1968), in (Eds.) Rush Rhees, (1968), *Philosophical Review*, 3, July LXXVII, 275-320, p.284

¹⁶⁶ Sydney Shoemaker, (1982), “The Inverted Spectrum”, in *The Journal of Philosophy*, Volume LXXIX, No. 7, July, p.358

¹⁶⁷ *Ibid.* p.359

i.e. a gradual and recorded inversion of someone's full colour quality space.¹⁶⁸ Furthermore, he rejects the conclusion that the possibility of inversion leads to scepticism about other minds. He thinks there are two senses in which colour may be said to 'look the same' to others, an "intentional sense" and a "qualitative sense". The former relates to the experience of colour *regardless* of, say, what it is like to experience it. The latter relates to the experience of colour *in terms of* what it's like to experience it. Only the second sense is undermined by the possibility of spectrum inversion.¹⁶⁹ Lastly, Shoemaker contends that though, as an empirical fact, colour inversion may not be possible *for us*, it does seem to be logically possible:

Even if our colour experience is not invertible, it seems obviously possible that there should be creatures, otherwise very much like ourselves, whose colour experience does have a structure that allows for such mapping.¹⁷⁰

Shoemaker draws two conclusions, a metaphysical conclusion and an epistemological conclusion. First, the metaphysical conclusion:

If spectrum inversion is so much as a logical possibility—whether or not it is a possibility *for us*, as we are currently constituted—then it is clear that no behaviouristic account of qualia will do...it also appears that the possibility of intersubjective spectrum inversion is incompatible with what many regard as the

¹⁶⁸ Ibid. p.363

¹⁶⁹ Ibid. p.366

¹⁷⁰ Ibid. p.367

most respectable descendent of behaviourism, namely functionalism, where this is understood as the view that mental states are definable in terms of their causal relations to sensory inputs, behavioural outputs, and other mental states.¹⁷¹

Behaviouristic accounts fail because an inverted creature may display exactly the same behaviours as a non-inverted creature.¹⁷² Functionalism fails because the quale that plays one role in the inverted creature plays a different role in the non-inverted creature.¹⁷³

The epistemological problem is this: if behaviouristic and functional accounts fail, then two creatures may have colour quality spaces with the same structure but none of the colour experiences of the other without anyone knowing to the contrary.¹⁷⁴ So it seems that inversion is a metaphysical and epistemological threat to both behaviourism and functionalism.

Ned Block also attacks functionalism, this time by way of the absent qualia argument.¹⁷⁵ Block argues that a pair of physical systems *could be* functionally equivalent, but the one enjoy qualitative states, the other lack them. Here it is in his own words:

Suppose we convert the government of China to functionalism, and we convince

¹⁷¹ Ibid. p.367

¹⁷² Ibid. p.367

¹⁷³ Ibid. p.368

¹⁷⁴ Ibid. p.369

¹⁷⁵ See N. Block, (1978), "Troubles With Functionalism", in *Minnesota Studies in the Philosophy of Science* 9, 261-325

its officials to realize a human mind for an hour. We provide each of the billion people in China (I chose China because it has a billion inhabitants) with a specially designed two-way radio that connects them in the appropriate way to other persons and to [an] artificial body...we arrange to have letters displayed on a series of satellites placed so that they can be seen from anywhere in China. Surely such a system is not physically impossible. It could be functionally equivalent to you for a short time, say an hour¹⁷⁶

Latter:

The force of the prima facie counterexample can be made clearer as follows: Machine functionalism says that each mental state is identical to a machine-table state. For example, a particular qualitative state, Q, is identical to a machine-table state, Sa. But if there is nothing it is like to be the homunculi-headed system, it cannot be in Q even when it is in Sa. Thus, if there is prima facie doubt about the homunculi-headed system's mentality, there is prima facie doubt that Q = Sa, i.e., doubt that the kind of functionalism under consideration is true. Call this argument the Absent Qualia Argument.¹⁷⁷

Ned Block does not just take this argument to be a threat to functionalism *per se*, but it is a threat to physicalism, for as he notes, functionalism is either too liberal allowing everything and anything to have mental properties, for example, the Chinese nation, or

¹⁷⁶ Ibid. p.279

¹⁷⁷ Ibid. p.281

too chauvinistic, restricting the functional realizers of the mental states in question to brain states, but since exactly the opposite was the driver of functionalism in the first place,¹⁷⁸ that seems decidedly unappetizing.

If functionalism is true, physicalists face a dilemma. Either they must abandon the attempt to propose a theory of mental universals such as pain, anger, etc., and talk instead of human pain, Martian pain, etc. (or worse, *deny* that anything has pain or anger, etc.), or they must claim that mental states are, for example, brain states and thus embrace chauvinism.¹⁷⁹

Both of these arguments are, as the vocabulary they are couched in testifies, *possibility* arguments. As Chalmers says: “Many of those arguing for the possibility of absent and inverted qualia have been arguing only for logical possibility; this is all that is required to refute a reductive form of functionalism.”¹⁸⁰ If this is true, I think, they are arguments that can be said to depend on the validity of Kripke’s modal argument. I have already introduced this argument to the reader. In the next section, I will be concerned to interpret and defend this argument, especially in its psychological mode. To do that I draw on the work of Stephen Kosslyn and Nelson Goodman, which allows me to form a pictorial theory of the imagination which makes psychological sense of Kripke’s argument. And to the extent that the argument is the foundation of the arguments mentioned above, its

¹⁷⁸ For example, see Hilary Putnam, (1975a), “The Nature of Mental States”, in Hilary Putnam, (1975), *Mind, Language and Reality*, (Cambridge: Cambridge University Press), pp.429-441

¹⁷⁹ Ibid. p.266

¹⁸⁰ Op cit. Chalmers, (1996), p.250

success or failure will be telling. Also note that in what follows. I will distinguish Kripke's argument from Chalmers. I will argue that Kripke's argument is *merely* a *modal* argument and *not* a *conceivability* or *imagination* argument at all and I will show points of discord between Kripke and Chalmers which I think licenses this move. So a couple of things will be done in the next section. Kripke's argument will be interpreted and defended; it will be distinguished from Chalmers's arguments. This takes up a third of the chapter, following this I will look at some problems for dualism, which amounts to an attempt to falsify dualism, last I draw the substantive arguments of chapter two together and make a comment or two about physicalism.

Chapter Two: Kripke's Argument for Dualism, Private

Knowledge and Other Problems

Section D: Modality, Imagination and Pain: Kripke's Argument for Dualism

In the previous chapter I explained Kripke's modal argument. It was an exposition and not a critique or defence and therefore carried little detail. The last chapter, indeed, only served as a introduction to themes and topics. From here on in, I push forward with my personal spin on matters. In the chapter that follows, I will present Kripke's argument in more depth and I will also fill in the gaps he left in his account, namely, the lack of detail around the psychological aspect of his argument. This requires me to develop what I call the picture theory account of the imagination, which draws upon the work of Stephen Kosslyn and Nelson Goodman. I think Kripke's argument is successful. And that so far as some other arguments mentioned for dualism are considered to be "possibility" arguments, they are reducible to Kripke's argument. I contrast Chalmers conceivability argument for dualism, which I argue is incautious and gratuitous, given Kripke's argument. I advocate caution, which, ultimately, does not view dualism, or even its possibility, as proved but presents its reality in a logical form which neither rules it in or out but from which it may be ruled in or out according to further fact or reason. The second section of this chapter will examine such reasons. In order to do this, I will say what I think the minimal assumptions that dualism needs are. There are three, a necessary condition, Kripke's conclusion, which is a sufficient condition, and a relation, which meets anti-sceptical, existential, ethical, phenomenological and scientific concern. It is, I believe, though not

necessary to dualism, a very important relation for dualism. Unfortunately, this requires the abandonment of the complete reduction of knowledge to physics, too. And this seems to lead to the conclusion that dualism, as defined, leads to a kind of "private knowledge". I will consider whether this is a problem for dualism and gives one reason to reject it. I also consider Davidson's and Searle's contributions to the philosophy of mind in relation to the set of conditions identified. And last, I grapple, with the attacks on dualism from causality, for example, from epiphenomenalism, on the one hand, and interactionism, on the other. This includes a general discussion on mental causation. In other words, in the second section, I'll have dualism run the gauntlet in order to see if it can be ruled out by reason or fact. The last section of this chapter will summarise the arguments presented and draw some conclusions bringing this work to an end with a note on physicalism.

First, then, I turn my attention to Kripke. The first thing to say is that there are *genuine* modal possibilities and there are *disingenuous* modal possibilities. For example, consider the following set of statements:

(1) Richard Dawkins could have been a rock star

(2) Squares could have been circles

Intuitively, the first does seem to express a genuine modal possibility, but the latter does not. There seems to be no absurdity in thinking there could have been some set of circumstances in which Richard Dawkins rocked. On the other hand, the second statement seems to be absurd. The reason why the first statement does not entail an absurdity is that

it is logically possible for Richard Dawkins to be *himself* and possess all the qualities associated with a rocker, for example, long hair, impressive guitar skills, a bottle of Jack Daniels in one hand and a loosely rolled cigarette in the other, etc. So it seems had things been different, Dawkins could have rocked. However, it is not logically possible for a square to be what it is and possess the defining qualities associated with a circle, i.e. being the set of all points equidistant from a point. There just seems to be no way a square could have been circular any more than one can depict a square that is a circle. This suggests we can connect modal possibility to logical possibility via absurdity if we think in something like the following manner: A statement of genuine modal possibility is a statement that can be true at a logically possible world. Such a statement can be true at a logically possible world because it does not entail an absurdity. And a statement that does not entail an absurdity is a statement of logical possibility. On the other hand, a statement of disingenuous modal possibility is a statement that cannot be true at a logically possible world. Such a statement cannot be true at a logically possible world because it does entail an absurdity. And a statement that entails an absurdity is a statement of logical impossibility. On this account a genuine modal possibility is, ultimately, a statement of logical possibility and a disingenuous modal possibility is a statement of logical impossibility. This is an idea of how one might understand modal possibility to be related to logical possibility. One discussion point may centre about what seems to be a logical limit on modality, what isn't absurd is something that could have been the case, and what is absurd is something that could not have been the case. That is, modality is defined in terms of what can and cannot be true at a logically possible world, rather than just what can or cannot be true in any broader sense (if there is a broader sense). I don't think this goes against Kripke, and since our discussion is based on his contributions I think the

definition of modality given above is consistent with his work. Smith, for example, writes that “Kripke has evinced no awareness of a logical/metaphysical distinction in his writings”¹⁸¹, “Kripke...emphatically identified logical and metaphysical necessity...”¹⁸², and quotes Alvin Plantinga, who is very much in the Kripkean tradition, thus, ““could have” expresses, broadly speaking, logical or metaphysical possibility.”¹⁸³ Note, I’m not identifying metaphysical possibility with logical possibility, myself, but only noting that there is an equivalence between the two. ¹⁸⁴

Sometimes we get confused about whether or not a statement of modal possibility is genuine or not i.e. a statement of logical possibility or not.

Consider the following statement:

(3) water is H₂O

This statement, if true, is necessarily true, that is, true in all possible worlds. Recall the kind of logical argument worked over in chapter one, section one. Essentialist statements

¹⁸¹ Quentin Smith, (1999), “A More Comprehensive History” in (Eds.) Paul W. Humphreys and James H. Fetzer, (1999), *The New Theory of Reference: Kripke, Marcus, and Its Origins*, (Dordrecht, The Netherlands: Kluwer Academic Publishers), p.269

¹⁸² Ibid, p.270

¹⁸³ Ibid, p.273, See, Alvin Plantinga, (1970), “World and Essence”, in *The Philosophical Review*, 79, No. 4, (Oct. 1970), 461-492

¹⁸⁴ Cf. Joseph Levine, (1998), “Conceivability and the Metaphysics of Mind”, in *Nous* 32-4, 449-480, Levine suggests that the following principle connects modality to logical consistency in the following terms, “A situation S is metaphysically possible just in case it has no accurate representation that is logically inconsistent”. p.450

like the one in question are necessarily true, if true, for the same reasons that Richard Dawkins could not have been Immanuel Kant, or that squares could not have been circles. These are absurd states of affairs and what is absurd is logically impossible. Nevertheless, even if we *know* that (3) is actually true and that therefore it is necessarily true, we may still entertain some vague notion that its negation could be true at some logically possible world. For example, we may think we could travel to a distant planet where we come across a substance that tantalizes our senses in all the various ways that actual water does, but has a molecular structure that is quite different from actual water or, perhaps, I can even think that I could have been born on such a world. So isn't it correct to say that we can think it possible that at some logically possible world water is not H₂O? No, because it leads directly to logical absurdities and logical absurdities are not genuinely possible. So we had better say that if anyone thinks that the negation of (3) is logically possible, they are making a mistake about what state of affairs they are actually thinking about. Kripke gives us a way to make sense of what is actually going on here: What someone really thinks about when they think the negation of (3) is logically possible is really just the logical possibility associated with a *qualitative analogue* of the actual substance, entity, kind, process etc. in question (hereafter, object is used as a catchall). A qualitative analogue is an object that shares identifying qualities and relations, with another (often actual) object. In more detail, say water-likeness is constituted by just those non-essential qualities associated with actual water that we use ordinarily to identify actual water, the qualities that tantalize our senses; taste, feel, sight, etc. A qualitative analogue of actual water will then be a substance that shares water-likeness with actual water, but does not share its essential molecular structure, H₂O. Thinking that the negation of the statement 'water is H₂O' is logically possible is really just thinking something about this qualitative

analogue. This may be captured by the following statement:

(4) water is not H₂O (where “water” refers to a qualitative analogue of actual water)

Such a statement does not entail any obvious absurdity, and therefore, seems to be a successful statement of logical possibility and as such a genuine modal possibility. Of course such a statement, we can know a priori, is a statement of modal *necessity* because if it weren't, again, we would face the very same logical absurdities that we just skirted. Recall, the logic behind Kripke's thinking from the first section of chapter one.

We might term this water-analogue “fools-water” as Kripke does in the following passage:

We identified water originally by its characteristic feel, appearance and perhaps taste, (though the taste may usually be due to the impurities). If there were a substance, even actually, which had a completely different atomic structure from that of water, but resembled water in these respects, would we say that some water wasn't H₂O? I think not. We would say instead that just as there is fool's gold there could be fool's water; a substance which, though having the properties by which we originally identified water, would not in fact be water. And this, I think, applies not only to the actual world but even when we talk about counterfactual situations.¹⁸⁵

¹⁸⁵ Op cit. Saul Kripke, (1980), p. 128

Here is another standard Kripkean example which makes room for a qualitative analogue. This example is included to help bring the imagination into play because it is often thought to be central to Kripke's thinking. First, a famous passage where Kripke talks about a lectern standing in front of him:

What I am saying is, given that it [the lectern] is in fact not made of ice, in fact is made of wood, one cannot imagine that under certain circumstances it could have been made of ice. So we have to say that though we cannot know a priori whether this table was made of ice, it is *necessarily* not made of ice. In other words, if *P* is the statement that the lectern is not made of ice, one knows by a priori philosophical analysis, some conditional of the form "if *P*, then necessarily *P*."¹⁸⁶

In this passage, I understand Kripke to state three things: First, given the facts (which I will take as captured by an essentialist statement of the kind introduced in the first chapter abbreviated by *P*), there are no circumstances we can imagine in which $\neg P$ is the case. Second, the facts determine modality, *regardless* of how we come to know about them i.e. even if they come to be known a posteriori. Third, we can ascertain a lot a priori because we know a priori that if *P* is true, it is necessarily true. The first claim is a psychological claim, the second a modal claim and the last a logical claim.

Kripke continues in the following manner:

¹⁸⁶ Saul Kripke, (1971), "Identity and Necessity" in (Ed.) M. K. Munitz, (1971), *Identity and Individuation*, (New York: New York University Press), p.154

If someone protests, regarding the [wooden] lectern, that it *could* after all have *turned out* to have been made of ice, and therefore could have been made of ice, I would reply that what he really means is that *a lectern* could have looked just like this one, and have been placed in the same position as this one, and yet have been made of ice. In short, I could have been in the *same epistemological situation* in relation to *a lectern made of ice* as I actually am in relation to *this lectern*.¹⁸⁷

In this passage, Kripke makes use of the notion of an epistemic situation. An epistemic situation is introduced by Kripke relative to an object and observer. This kind of situation, relative to the object and observer, is characterized or defined by the way in which the object ordinarily becomes known to the observer. So, relative to water, an epistemic situation involves all the qualities and relations associated with water that we ordinarily use to identify water by, those qualities that I said tantalize our senses. In this passage, the relevant epistemic situation is characterized by Kripke's lectern as it appears to the observers who come to know it by certain qualities and relations, for example, being in front of Kripke as he delivers his lecture, etc. Kripke says, when we say that the actual lectern could have been made of ice, which he says we cannot *really* imagine, we are really saying something about a qualitative analogue. We mistake it for the actual lectern because it could well stand in the same epistemic situation to the observer, that is, be known to the observer by the same set of set of ordinary qualities and relations that the actual lectern is ordinarily known by. So epistemic situations and the related relations and

¹⁸⁷ Ibid. p.154

qualities have the ability to confuse. These befuddling epistemic properties make the imagination unreliable. This explains why we can become confused about what it is that we are actually imagining. Qualitative analogues help to disambiguate the imagination by showing that something can be epistemically indistinct but essentially distinct.

It is worth summarizing the main points here: Facts determine both what can be imagined and what could have been so. However, sometimes the imagination *seems* to exceed what is logically possible, for example, when we think we can imagine the negation of (3). But this is not really so and can be explained as a confusion arising from the epistemic relations and qualities involved. Qualitative analogues can be used to clear the fog and separate the wheat from the chaff, distinguishing the epistemic relations and qualities shared by a possible analogue and the object itself. Thus qualitative analogues help disambiguate what is really being imagined and what we cannot really imagine but might think we can.

Is it really the case that Kripke holds such views? Does Kripke really believe that so far as the imagination is concerned, the facts constrain it; epistemic qualities and relations serve just to fog it up; and qualitative analogues can be used to clear the fog and keep the imagination within its narrow logical bounds? Well, every time the imagination seems to exceed its factual bounds Kripke reigns it in by deploying a qualitative analogue. *Not only* does he say we can't imagine $\neg P$, for example, but when he does, *he shows* that $\neg P$ isn't being imagined at all by use of a qualitative analogue. Kripke even refers to this as a "strategy".¹⁸⁸ So I think there is enough in Kripke to suggest Kripke thinks that the

¹⁸⁸ Op cit. Kripke (1980), p.150

imagination is constrained by factual bounds and enough practical evidence to suggest he wants to defend such borders. Unfortunately, Kripke says little specifically on the imagination. So, I'd like to say a little bit about the imagination myself.

Basically, I think, we can build independent support for the way I've interpreted Kripke above from psychology, specifically the psychology of mental imagery. We can think of mental imagery as either "depictive" or "descriptive", that is, pictorial or linguistic. That is, having a mental image of such and so takes the form of a visual picture before the mind's eye, a sound picture before the mind's ear, or a tactile picture before the mind's skin;¹⁸⁹ or takes a linguistic form.¹⁹⁰ If we think of mental imagery as pictorial, then we can think of mental imagery misleading us in pictorial terms. For example, a picture may refer to two essentially distinct things at one and the same time or represent something as another essentially distinct thing, thus leading to confusion. Clarity can be restored by understanding how pictures work and how, thereof, to disperse the pictorial mist clouding the understanding. In providing the detail, I will draw on Nelson Goodman.

Pictorial representation involves two forms of reference: representation *per se*, and representation-as.¹⁹¹ The former simply denotes; something like a label. The latter refers, and what it refers to refers back to it, the archetype is a swatch. This two-way pictorial

¹⁸⁹ S. M. Kosslyn, W. L. Thompson, W. L., G. Ganis, (2010), *The Case for Mental Imagery*, Oxford: Oxford University Press, p.4

¹⁹⁰ Z. W. Pylyshyn, (2003), "Return of the Mental Image: Are There Pictures in the Brain?", in *Trends in Cognitive Sciences*, 7, 113-118

¹⁹¹ Nelson Goodman, (1976), *Languages of Art*, (Indiana, Indianapolis: Hackett Publishing Company, Inc.), pp.27-31

relation involves exemplification.¹⁹² This passage sums it up quite nicely:

The difference amounts to this: for a word, say, to denote red things requires nothing more than letting it refer to them; but for my green sweater to exemplify a predicate, letting the sweater refer to that predicate is not enough. The sweater must also be denoted by that predicate; that is, I must also let the predicate refer to the sweater. The constraint upon exemplification as compared with denotation derives from the status of exemplification as a subrelation of the converse of denotation, from the fact that denotation implies reference between two elements in one direction while exemplification implies reference between the two in both directions.¹⁹³

So if P is a pictorial representation and represents x, then P denotes x, pure and simple. Moreover, if P represents x, and $x = y$, then P denotes y. Secondly, if P represents x and represents x as F, then P refers to F. It might be said that P is an F picture and that P exemplifies F, where F is generally taken to be a description or set of descriptions. Remember exemplification is a two-way function; we can think if P exemplifies F, then P refers to F-descriptions, and those F-descriptions refer to how x is represented by P. Let's try some substitutions. If P represents the Duke of Wellington, then P denotes the Duke of Wellington. Moreover, if P represents the Duke of Wellington, and the Duke of Wellington = Arthur Wellesley, then P denotes Arthur Wellesley, too. Say, now, that P

¹⁹² "Description-as and representation-as, though pertaining to labels, are matters of exemplification rather than of denotation." Ibid. p.66

¹⁹³ Ibid. p.59

represents the Duke of Wellington as a soldier. Then, P exemplifies a set of descriptions associated with soldiers and those descriptions in return describe how the Duke of Wellington is represented. Let's consider water. If P is a picture of actual water, then it denotes H₂O. If P represents actual water as actual water, then it exemplifies a set of descriptions associated with actual water. Well, what if P *doesn't* represent actual water as *actual* water at all? Consider the following case, P represents actual water as lacking H₂O yet possessing all the properties that tantalize our senses, how are we to understand P in this case? Isn't it correct to say that *water* is being pictured as lacking its essential molecular character? So, don't we have an image of *waterless water*? No! Let's clear the fog. If P is such a picture and P denotes water, because water = H₂O, P denotes H₂O. On the other hand, P *represents* its object *as* teasing our senses in the many ways that actual water does, but lacking its essential character, so simply enough P exemplifies all the descriptions associated with the way that water tantalizes our senses, which in turn refer back to what the picture represents its object as being like. However, what P *represents* water *as*, even if it exemplifies descriptions that designate water-like substances, is not actual water, because actual water = H₂O. What P *represents* water *as* just *because* it exemplifies descriptions that designate water-like substances, but not the essential characteristics of actual water, must be a qualitative analogue of water. That is something that resembles water to a high degree in respect of the way that water plays and laps against our senses, its sensible qualities and relations, less its essential characteristics. Confusion arises because the two modes of reference are not sufficiently distinguished. Two different things are referred to by two different functions of the one picture, one that pertains to what is represented and one that pertains to what it is being represented as, the denotation function and the exemplification function. The confusion is compounded by

the fact that the resemblance of the thing represented and the thing it is represented as is *so* profound.

So, I am taking mental images as pictures. Since a mental image of water is a picture, the same thinking as applied to pictures applies to mental images. I can mentally picture water *as* tantalizing my senses in all the typical ways it does but lacking its essential characteristics. But by keeping the nature of pictorial representation and the two different types of reference involved in mind, I will simply note that I am picturing water as another object which exemplifies the same descriptions I associate with water. That is, I am *picturing* water *as* a qualitative analogue of water, something that is qualitatively indistinguishable from water ordinarily, but essentially different. Confusion is understandable so far as the picture in question represents water as something resembling water to such a high degree; unfortunately (but, I think, naturalistically) to a confusing degree.

Well, if all this is acceptable, some faculty of the mind is responsible for this. But now just what would this faculty be if not the imagination? After all aren't we talking about mental picturing? I think, we can accept the picturing we have been discussing as something undertaken by the imagination for several reasons. First, just consider being asked to *picture* yourself on a sandy white beach in the South Pacific with the clear emerald water lapping at your bare feet reflecting the blue cloudless skies above, a hat tipped over your eyes, a cocktail precariously dangling from your right hand, etc. *What* are you being asked to *do* here? What *do* you *do*?¹⁹⁴ Wouldn't you answer, "Imagine" to

¹⁹⁴ What are you doing, right now?

both these questions? I don't think anyone who understands the meaning of 'imagine' would answer anything different.

The second reason is related, it just seems to be a fact that images and the imagination are and always have been closely related. Nigel Thomas writes in the following fashion:

Prima facie imagery and imagination are intimately related—certainly "imagination" is often used to name the faculty of image production (or the mental arena in which images appear)—and this fact itself demands explanation.¹⁹⁵

Furthermore, Amy Kind writes that for the essentialist, "mental images play an essential role in the imagination."¹⁹⁶ So it is quite in line with the Kripkean take on things that we should see fit to follow suit. So there is reason to believe that imagery and imagination are related, psychologically (Thomas) and philosophically (Kind). According to Kosslyn, the imagery-debate has been strengthened, due to advances in scientific-imaging of the brain.¹⁹⁷ The good news for Kosslyn is that the debate is moving in favour of the picture-based view of imagery. So if mental imagery is indeed pictorial, then something like the account above might well fit the bill.

¹⁹⁵ N. J. T Thomas, (1999), "Are Theories of Imagery Theories of Imagination? An Active Perception Approach to Conscious Mental Content" in *Cognitive Science* Vol 23 (2), 207-245, pp.207-208

¹⁹⁶ Amy Kind, (2001), "Putting the Image Back into Imagination", in *Philosophy and Phenomenological Research* 62, 85-109

¹⁹⁷ He writes, "[T]here is good evidence that the brain depicts representations literally, using space on the cortex to represent the world." Op cit. Kosslyn, (2010), p.15

Another reason for accepting picturing as imagining is explanatory power. This sense of the imagination is motivated by the possibility of using it to make sense of Kripke. So, if P is a mental picture of actual water pictured as tantalizing our senses in the typical ways actual water usually does but lacking in its essential character, we refer to two distinct things—that is, actual water, that is, H₂O, so far as the representation function of P is concerned; and a qualitative analogue of actual water, that is, not H₂O, so far as the representation-as function of P is concerned. One denoted thing is pictured as another essentially distinct but closely resembling thing, indeed, qualitatively indistinguishable thing. We are confounded because the descriptions or properties exemplified by P are the same descriptions or properties exemplified by actual water or even a picture of actual water that represents it as actual water. The confusion is dispersed once we understand our picture of actual water is being pictured as a qualitative analogue of actual water, a water-like substance, but not anything that could actually be water, given the facts. This explanation is quite simple and rather intuitive, which brings us back to the way we ordinarily respond to the term ‘imagination’.

So there are reasons to run picturing and imagining together, ordinary language reasons, psychological and philosophical reasons, and explanation, simplicity and intuitiveness. It is important to note that mental images understood as pictures went out of fashion during the last century (at least as *functional* pictures). This was due to the rise of behaviourism and the predominance of linguistic dominance in analytic and continental philosophical circles. So an account of mental imagery that flatters this iconoclasm might be asked for. A non-pictorial account of mental imagery is available to the psychologist. The exact

details are not worked out here because the pictorial understanding of mental imagery is preferred for the reasons given above.¹⁹⁸ Anderson, however, argues that a pictorial theory may be interpreted in a propositional manner.¹⁹⁹ Well, it certainly seems this way given descriptive theorists spend a considerable amount of energy reinterpreting experiments in which the pictorial theory of mental imagery is thought to be evidenced. Kosslyn, himself, shows how such experiments can be reinterpreted in a propositional manner.²⁰⁰ Further, one can always take me as talking idiomatically if you wish because what I take from Goodman is applicable to reference and representation in general. Indeed, Catherine Elgin argues that Goodman's theory of exemplification can be applied very broadly:

Exemplification, the relation of a sample to a label of which it is a sample, is widespread. Symbols function as samples, or have the semantic structure of samples, in art works, scientific experiments, and mathematical proofs, as well as in marketing campaigns. Moreover, interpreting samples is crucial to learning.²⁰¹

¹⁹⁸ Cf. Colin McGinn, (2006), *Mindsight: Image, Dream, Meaning*, Boston, (Massachusetts: Harvard University Press)

¹⁹⁹ J R. Anderson, (1978), "Arguments Concerning Representations of Mental Imagery" in *Psychological Review*, 85, 249-277

²⁰⁰ See, for example, Stephen Kosslyn, (1983), *Ghosts in the Mind's Machine*, (New York: Norton)

²⁰¹ Catherine Z. Elgin, (1983), *With Reference to Reference*, (Indianapolis, Indiana: Hackett Publishing Company), p.73

Well, for Elgin, even an equation exemplifies.²⁰² The ubiquitous nature of Goodman's theory means that the way I have used his theory may be taken as idiomatic for those disinclined to the pictorial side of my theory.

For the moment, let me review a bit, here are the claims I read into Kripke previously. First, facts constrain the imagination. In terms of what has been said, mental pictures defer to the factual necessities involved. For example, if P is a picture of water then it denotes water *and* H₂O, because water = H₂O. If P, however, is a picture of water that *represents* it *as* lacking the essential molecular character of water, then it represents water *as* a qualitative analogue of water, rather than actual water, because water = H₂O. Second, epistemic qualities and relations fog the imagination. In terms of what we have said, mental pictures may exemplify descriptions that are shared by essentially distinct objects and this can confuse us. For example, a picture of water that represents it as a qualitative analogue of water, something that is qualitatively indistinguishable from water, but essentially distinct, can befuddle us into thinking actual water could have lacked its essential molecular characteristic. Third, the fog is dispersed by noting one is picturing actual water as a qualitative analogue—that is, understanding that one thing is being represented as another thing that is qualitatively indistinguishable from the other but essentially distinct. For example, recognizing a picture that represents water is representing it *as* an essentially distinct object that exemplifies all the same descriptions as actual water but lacks the essential characteristics, disperses the confusion. Such

²⁰² Catherine Z. Elgin, (2009), "Exemplifications, Idealization, and Understanding", in (Ed.) Mauricio Suarez, (2009), *Fictions in Science: Essays on Idealization and Modelling*, (London: Routledge), pp.77-90

confusion is easy to understand given that what is represented is represented as something resembling it to such a high degree, indeed, qualitatively indistinguishable from it. A picture-based theory of the imagination, then, seems to be able to give sense to the claims associated with Kripke.

But why do we need an account that involves the imagination at all? Didn't Kripke say that the facts constrain modality and we can know a priori the ways in which they are constrained? For example, we know that if P, then necessarily P for a certain type of statement as we learned in the first section of chapter one. So if the facts tell us that P is the case, we know that necessarily P is the case, but if the facts tell us that the negation of P is possible, we know that P is false, and so on. Alex Byrne makes a similar point:

[W]hen Kripke says that we can't imagine a situation in which the Queen is the biological daughter of Mr. and Mrs. Truman, 'imagine' could easily be otiose: we can't imagine (or "think of") such a situation, not because of limits to our imagination, but because there is no such situation to imagine.²⁰³

So, as asked, why do we need an account of the imagination at all? For a simple reason, I think: We need an account of the imagination *because sometimes the imagination misleads us*. That's it!²⁰⁴ It is true that the facts determine modality and neither the

²⁰³ Alex Byrne, (2007), "Possibility and Imagination", extracted <http://web.mit.edu/abyrne/www/possandimag.pdf> 20/07/2013 p.4

²⁰⁴ Cf. George Bealer, (2002), "The Origin of Modal Error" in *Dialectica*, Vol. 58, No. 1, 11-42, who writes: "Modal intuitions are not only the primary source of modal knowledge but also the primary source of modal error", p.11

imagination nor knowledge are necessary to this process. However, it is also true that the imagination interferes with our understanding of the relevant modalities involved. That is, we can think of the imagination as a kind of mental rain or psychological noise that we need to factor in and reduce in order to stop blurring our view of the facts as they really are. Luckily the imagination is constrained by the facts, so one cannot really imagine something that is impossible, that's what Kripke says, but it can *seem* to be the case that one can. How and why is the imagination so constrained? The picture-theory of the imagination explains this referentially and factually. For example, if P is a mental picture of actual water, then P represents H₂O. If P represents x as actual water, then P represents x as H₂O. If P represents x as lacking H₂O, even if exemplifying all the descriptions associated with the way actual water tantalizes the senses, P does not *represent* x as actual water. That's because, simply enough, anything that is water *is* H₂O. That is determined by the facts. Therefore, it is not possible of P to represent or represent x as water that is not also H₂O because the facts determine anything that is water *is* H₂O.

A similar story can be told about statements and what they mean. Let's say the facts determine the meaning of a term in a wide sense. The meaning of the term 'water' is determined by its molecular structure. That is, meanings are not in the head as Hilary Putnam once put it.²⁰⁵ So even if I could imagine that water was not H₂O, I could not mean so. Consider:

(5) Water is XYZ (where XYZ ≠ H₂O)

²⁰⁵ Hilary Putnam, (1975b), "The Meaning of 'Meaning'" in Op cit. Hilary Putnam, (1975), pp.215-272

The term “water” in this statement cannot refer to actual water. (This term will be distinguished from the term that picks out actual water by scare quotes)

On the other hand, we can talk about the stereotype of a term. The stereotype of a term is determined in a narrow sense.²⁰⁶ The narrow sense of a term is determined by the typical descriptions associated with the term in question, for example, in the case of water the descriptions associated with the way water tantalizes our senses. It is also possible to assume that the stereotype of the term “water” is associated with the descriptions that describe the way that actual water plays against our senses. Yet, this still does not mean that that term refers to actual water. And we can say this is because it is not possible for “water” to refer to actual water because the facts determine anything that is water *is* H₂O.

This brings me back to the question I have just been considering, given we can tell such a story about statements like (5), and understand the factual constraints, why do we need a story about the imagination? The answer is the same: *The imagination can mislead us and confuse us to the extent that we lose track of what we are **saying** and what we **mean**.* Telling a story about the imagination, in pictorial terms, as done here, helps us to clarify just how the befuddling effects of the imagination conspire against our modal-judgments and our understanding of our linguistic expressions. And this story needs to be told because philosophers, empiricists like Hume say, “nothing we imagine is absolutely impossible”;²⁰⁷ rationalists like Descartes say, “the power of the imagination is

²⁰⁶ Ibid. pp.247-53

²⁰⁷ David Hume, (1968), *Treatise of Human Nature*, (Oxford: Clarendon Press), p.32

something which really exists and is part of my thinking”;²⁰⁸ up to modern day philosophers of consciousness (David Chalmers) have thought that one can move from what is conceivable to what is genuinely possible. This is what Chalmers says:

The most popular bridge [between the epistemic and modal domains] is the method of conceivability. One argues that some state of affairs is conceivable, and from there one concludes that this state of affairs is possible. Here, the kind of possibility at issue is metaphysical possibility, as opposed to physical possibility, natural possibility, and other sorts of possibility... And there is at least some plausibility in the idea that conceivability can act as a guide to metaphysical possibility.²⁰⁹

Well, the picture-based theory of mental imagery and the imagination given above, sits well with the story about meaning just given. The picture theory of the imagination offered here helps explain why we sometimes get confused about what we say and mean. Such picturing I understood as the imagination for the several reasons given above. Referential limits make it impossible to imagine, for example, water as lacking H₂O. If P is a picture and P represents water, then it denotes water and H₂O. If P represents water *as tantalizing the senses in all the ways actual water does but lacking H₂O*, it must be because water is being represented as a qualitative analogue, something that qualitatively resembles water to a high degree, but something that is essentially distinct from it. So

²⁰⁸ Rene Descartes, (1985), *Meditations on First Philosophy*, (Eds.) (Trans.) John Cottingham, (1985), (Cambridge: Cambridge University Press), see second meditation

²⁰⁹ Op cit. Chalmers, D. J., (2002), “Does Conceivability Entail Possibility?”

when the imagination seems to exceed its factual bounds, qualitative analogues are posited to show that no factual bounds are in fact transgressed. Although one can tell a similar story in linguistic terms by saying things like, ‘You can’t mean _____’, and, ‘What you’re really saying is _____’, one can guard against the mental noise of the imagination interfering with modal judgments by focusing on the mental process that may cause us to misinterpret our words and statements and also block moves from what is conceivable to what is possible by demonstrating that what one thinks is conceivable is not really so. I don’t attribute this thinking completely to Kripke, but I think it is a way to understand and support the way Kripke has been understood above.

At this point I would like to summarize everything that has gone before as a precursor to the following discussion. First, the facts constrain modality. That is, it is the facts that determine what a genuine modal possibility is and what a disingenuous modal possibility is. It is a fact that water is H₂O, therefore it is necessarily the case that water is H₂O. Second, we can draw modal implications a priori. That is, we know a priori for essentialist statements that, $P \rightarrow \Box P$, $\Diamond \neg P \rightarrow \Box \neg P$, and etc. Again, in terms of the essentialist statements run over in the first section of the first chapter. That is, we can know a priori what will count as a genuine possibility and what won’t. So, given the facts, $\Diamond \neg P$, we can draw the relevant conclusion, $\neg \Diamond P$. To connect this to the opening of the paper where genuine possibilities were defined as those that did not lead to absurdity and disingenuous possibilities those that did, it seems, given the reading of Kripke presented here, the facts determine what is a genuine, as opposed to disingenuous possibility. For example, if the facts tell us $\Diamond \neg P$, then it follows that $\Diamond P$ is a disingenuous possibility.²¹⁰ Third, the

²¹⁰ Again, of course, it does matter what kinds of statements we are talking about here.

imagination is constrained by the facts, too, for the same reasons as language is constrained by the facts. According to the picture theory of the imagination presented here, the imagination is pictorial. Pictures refer in two ways: representing and representing as. Representation is constrained by the facts. If P represents x and $x = y$, then P represents y. Representation-as is constrained by the facts, too. If P represents x as y and $y = z$, then P represents x as z. Fourth, unfortunately, the imagination can act as an annoyance, a kind of mental rain fogging our ability to make the correct modal judgments, even interfering with our ability to understand what it is that we are saying and what it is that we mean. Never fear, the pictorial theory of the imagination explains exactly what is going on here and explains it in terms of qualitative analogues central to the strategy that Kripke gives us for clearing the air. It is possible for a picture representing one thing to represent that one thing as something completely distinct. For example, Winston Churchill may be (and has been) represented as a bulldog. What confuses most is when one thing, x, is represented as an essentially distinct thing, not x, yet exemplifying all the typical descriptions that x does. The archetypical example is water represented as water-like, qualitatively resembling water to such a high degree that it is qualitatively indistinct from water; exemplifying all the typical descriptions associated with water, but denoting essentially something else, since it lacks the essential molecular structure of water. Understanding how psychology may conspire against us in this way helps to keep the fog of confusion away and hopefully helps us to judge modality and meaning aright. It should be noted the chances of confusion are proportional to one's ignorance; the less one knows, the more apt to be confused one becomes. At the extreme, confusion is inevitable.

So qualitative analogues can help defog epistemic confusion and a picture theory of the

imagination, I believe, can help make sense of how epistemic qualities and relations can confuse and how such confusion can be dispersed. However, sometimes trying to deploy a qualitative analogue to highlight and disperse an apparent confusion fails, which seems to suggest (not imply) that there was no confusion to start with in the first place. Here is an example. Kripke's argument against the so-called Mind-Body Identity Theory (MBIT). MBIT is a dual-aspect and materialist theory²¹¹ of consciousness that identifies pains and the like with brain processes.²¹² The following hypothesis is a statement of MBIT:

(6) Pain is identical to the firing of C-fibres

Of course, the statement in question is held widely by materialist of every ilk. So, of course, the argument below applies to the materialist in general. Unfortunately, for the materialist, it seems that the possibility of a statement that entails the negation of (6) as it stands has meaning:

(7) It is possible that a pain occur without the firing of C-fibres and vice versa

But if this is the case, then we know a priori that (6) is false. This, of course, follows the logic of the first section of chapter one.²¹³ So the physicalist in question might want the use of a qualitative analogue in order to dissolve the contingency. But Kripke says *this is*

²¹¹ Dual-aspect theory needn't be materialistic in bent.

²¹² Ullin Place (1956), "Is Consciousness a Brain Process", in David Chalmers, (2002), *Philosophy of Mind: Classical and Contemporary Readings*, (Oxford: Oxford University Press), pp.55-60

²¹³ Put succinctly, the pattern of logic this follows is thus: $A \rightarrow \Box A, \Diamond \neg A$, therefore, $\neg A$.

not possible.

To be in the same epistemic situation that would obtain if one had a pain *is* to have a pain; to be in the same epistemic situation that would obtain in the absence of a pain *is* not to have a pain. The apparent contingency of the connection between the mental state and the corresponding brain state thus cannot be explained by some sort of qualitative analogue as [for example] in the case of heat.²¹⁴

Kripke is arguing, the materialist cannot explain the logical possibility of the negation of (6) away by arguing, for example, in the following mode:

What someone thinks when they think (7) is genuinely possible is really just the logical possibility associated with a statement about a qualitative analogue of pain. The epistemic situation relative to pain and an observer is the way she ordinarily identifies pain – the way she ordinarily identifies pain in herself, that is, that oft unpleasant sensation associated with actual pain. A qualitative analogue of pain, is something that shares the same epistemic situation relative to an observer that actual pain does, but does not have its essential neurological identifier, the firing of C-fibres. What one imagines when one thinks that (7) is possible is just something about this qualitative analogue. That may be captured by the following statement:

*(8) It is possible that **pain** is not identical to the firing of C-fibres (where **pain** refers to*

²¹⁴ Ibid., p.152

fools-pain, or a qualitative analogue of pain, defined by the relevant epistemic relations and qualities)

This explanation will not do for the MBIT theorist, or materialist, because a qualitative analogue of pain is not really anything other than pain, it is rather, *identical to pain* because pain is identical to (not just identified by) its qualitative aspects! Kripke assumes what we may call the phenomenological assumption, that is, that pain is essentially the sensation of pain. Given this assumption if both (6) and (7) are correct, then pain is and is not the firing of C-fibres, which is absurd.

The assumption may be dropped, but then MBIT mis-describes pain. That is, epistemic relations and qualities associated with a pain and pain itself cannot be distinguished because the former and the latter are the same. So nothing can *share* the epistemic relations and qualities associated with pain and fail to be pain. So, if (6) and (7) are true, absurdity looms because (6) says pain is identical to the firing of C-fibres and (7) denies this. But if the phenomenological assumption is denied in order to distinguish (6) from (7) and circumvent absurdity, pain is misrepresented, and misrepresented by (6).

Using the picture theory the argument can be made out in contrast to the case of water in the following way. If P is a picture of water then it denotes water and H₂O. If P, however, is a picture of water that represents it as lacking certain qualities associated with it, for example, its essential molecular characteristic, P may still represent water but as *water-like*, that is, by representing it as tantalizing the senses in the ways that water does. P exemplifies descriptions that refer to properties associated with objects that are water-like,

that is, qualitatively indistinguishable from actual water. It is in this sense that water can be said to be represented as a *qualitative* analogue of water, but not as actual water. And it is for this reason that we may get confused about what it is that we are imagining. We are not picturing waterless water, but picturing something as water-like, qualitatively indistinguishable, though essentially distinct, from actual water.

But now contrast pain. If P is a picture of pain, then it denotes pain *and* the qualitative aspect of pain, because the essence of pain is its qualitative, or felt, aspect. And if P is a picture of pain and P represents pain as pain, then it must represent pain as bothering our senses in the troublesome way that pain ordinarily does, the ways that make you groan, grimace, cry and moan. So if P is a picture of pain that represents it as lacking certain characteristics, including a neurophysiological correlate, representing pain as *pain-like*, that is, as bothering the senses in the way it does, it *still* represents pain *as* pain. If P exemplifies descriptions that refer to the properties associated with something that is qualitatively indistinguishable from pain, P refers to descriptions that *essentially* refer to the properties associated with pain. So we aren't talking about a qualitative analogue here, at all, but pain itself. A psychological confusion is, therefore, not suggested.

I think, Kripke's argument can be put in the following form: In the case of water: if it is the case, if P is a picture that represents water as water-like, qualitatively indistinguishable from actual water, but less its essential molecular structure, then water is not represented as water, then water is being *represented as* a qualitative analogue of water, that is, as

something that qualitatively resembles water but is essentially distinct from it.²¹⁵ Thus, the confusion. The representation of water is confounded with what it is represented as, something water-like less H₂O, so not actual water. Analogously, we might argue, in the case of pain, too: if it is the case, if P is a picture that represents pain as pain-like, qualitatively indistinguishable from actual pain, less the neurophysiological correlate in question, then pain is not represented as pain, then pain is represented as a qualitative analogue of pain, that is, as something that qualitatively resembles pain but is essentially distinct from it.²¹⁶ Thus, the confusion. The representation of pain is confounded with what it is represented as, that is, something pain-like, less the firing of C-fibres, so not pain. However, so far as P is a picture that represents pain as pain-like where pain-like exemplifies all the ways in which pain bothers and troubles the senses, that is, P represents pain qualitatively, then P does not represent pain as a qualitative analogue of pain *but as identical to pain*. Even worse for the materialist, that ultimately suggests that pain can be represented as actual pain less some neurophysiological correlate.²¹⁷ That is, it suggests, that it is possible that pain may occur without the firing of C-fibres. The main point is, the materialist cannot point to a psychological confusion here, a mental picture of pain is not being confused with what the denoted object is being represented as.

Well, Kripke also indicates the argument can run in the other direction, that is, the

²¹⁵ That is, (P represents x as water-like but not H₂O → x is not represented as actual water) → x is represented as a qualitative analogue of water.

²¹⁶ That is, (P represents x as pain-like but not neurophysiological → x is not represented as actual pain) → x is represented as a qualitative analogue of pain.

²¹⁷ However, pain = pain-like. Therefore, x is not represented as a qualitative analogue of pain. Therefore, x is represented as pain-like and not neurophysiological, and it is false that x is not represented as actual pain.

materialist has no way of ruling out (7) when it is pictured as C-fibres firing in the absence of pain:

[Is it possible that] a stimulation of C-fibres should have existed without being felt as pain? If this is possible, then the stimulation of C-fibres can itself exist without pain, since for it to exist without it being *felt as pain* is for it to exist without there *being any* pain. Such a situation would be in flat contradiction with the supposed necessary identity of pain and the corresponding physical state, and the analogue holds for any physical state which might be identified with a corresponding mental state.²¹⁸

Again, one can put the point being made here in terms of the picture theory of the imagination. To understand how, I need to introduce another kind of confusion that may blur our psychological vision. This time a qualitative analogue is not to blame, but the *essential character* of the thing in question is what produces the trouble in this case. Kripke's point will then be that the materialist cannot explain (7) away by arguing something like a lack of knowledge or conceptual paucity is involved when one fails to see that (7) is false.

I said, if a picture P represents x, then it denotes x, and if $x = y$, it denotes y. A picture P may represent x *as* y, whereby, it represents x as it essentially is, x. This, however, can lead to confusion. A picture P may represent x and represent it as y, but such a picture may lead someone into thinking that x is anything but x! Consider the following case: A

²¹⁸ Ibid, Kripke, (1980), p.151

picture P may represent water as H₂O. The picture represents water and so represents H₂O. It also represents water as it essentially is, water, at its molecular level, H₂O. However, let's say this picture represents water as H₂O, as it essentially is at the molecular level, devoid of any resemblance to or indication of the way it plays upon the senses. In this case, water is represented as actual water, yet because the representation of water as H₂O is so alien to the way that water is ordinarily encountered, those ways that tantalize the senses, to someone who is ignorant of the essential characteristics of water, the picture will look as if it doesn't represent water at all, perhaps, anything but. So far as such a picture is a mental picture, one may think that what one imagines is not water at all, but something other than water, even anything but water, failing to make the requisite theoretical connection.

Here is a very famous example. Consider the following substance description and allow it to enter the imagination. It is called "hydroxyl acid". Here are some of its characteristics:

The substance is the major component of acid rain. It contributes to the "greenhouse effect". It may cause severe life-threatening burns. It will cause death if inhaled. It erodes the natural landscape it is so powerful, and causes corrosion and rusting of many metals. It may cause electrical failures. It can also cause traffic accidents by decreasing the effectiveness of automobile brakes. It has been found in excised tumors of terminal cancer patients. Despite being so dangerous, it is used as an industrial solvent and coolant. It is even used in nuclear power plants, but caused severe damage at Fukushima. It is, yet, used as

a fire-retardant. And in forms of animal research that cause suffering to animals. It is present in the distribution of pesticides, yet it is an additive in junk foods and other products.²¹⁹

What are you imagining? Take a look at the last footnote. If you were fooled, as I was, it shows how the imagination may not easily foster links between, say, a picture of water representing it as H₂O, that is, as it essentially is, but devoid of its *ordinary* qualitative characteristics,²²⁰ one may even be led astray and think one is imagining something that is anything but that thirst quenching giver of life that we call water.

So as said, a picture P may represent x and represent it as y, but such a picture may lead someone into thinking that x is anything but x. Consider the following case: A picture P representing pain as the firing of C-fibres. The materialist wants to say that such a picture represents pain and it represents pain as pain, that is, as it essentially is. One can think that such a picture represents pain as the firing of C-fibres less the felt-aspects of pain, those troublesome aspects that make us wince, groan, cry and moan. That is, this picture represents pain devoid of any resemblance to the way it torments the senses. The materialist may want to say that the way pain is presented here is so alien to the way that it is ordinarily experienced that someone who is ignorant of the essential characteristics of pain, will reject the picture in question as anything but a picture of pain. Perhaps because the thinker in question has no knowledge of the science or underlying

²¹⁹ See the last footnote for a link.

²²⁰ Or even a representation of it in slightly unfamiliar terms consistent with its true nature, as the example given suggests.

connections involved. The problem is that pain, as said above, and as captured by Kripke's phenomenological assumption, *just is felt-pain!* That is, *pain is identical to felt-pain.* Therefore, a picture that represents pain as the firing of C-fibres, less its felt nature, doesn't represent pain as pain at all. This suggests that "the stimulation of C-fibres can itself exist without pain", as Kripke says. And that what the materialist thinks is a picture of pain, or what he imagines to be pain, is not really such a picture at all. It is, perhaps, then, the materialist's thinking which ends up looking back to front and inside out! The main point is that the materialist cannot lay the blame for assenting to (7) on a lack of knowledge of pain or conceptual paucity.

Here is another way of stating this argument contrasting water and pain: If P is a picture that represents water as H₂O, essentially indistinguishable from actual water, but missing its qualitative aspects, those that tantalizes our senses in all the ways water ordinarily does, then, regardless, water is *still* represented as actual water, then water is being *represented as* it essentially is, that is, as something essentially indistinguishable from actual water, but in a way that is distinct from how it ordinarily appears to us.²²¹ Confusion may arrive in such a case if one lacks knowledge of the chemical theory of water. This, again, is natural and intuitive. Any picture, mental or otherwise of water represented as H₂O, but lacking in the ordinary qualities we associate with it, may make us think of anything but water if we lack knowledge of its molecular structure, if we are ignorant of molecular theory completely, we won't even have the concepts required to make the connections even when informed of the molecular structure in question. Nevertheless, we have a

²²¹ That is, (P represents x as H₂O but not water-like → x is still represented as actual water) → x is being represented as water actually is.

picture of water represented as water before us. Analogously, with the emphasis on the *absence* of felt-pain; if it is the case, if P is a picture that represents pain as the firing of C-fibres, essentially indistinguishable from actual pain (remember the firing of C-fibres is assumed to be identical to actual pain), but lacking its qualitative aspects, the way it feels to experience pain, then pain is *still* represented as actual pain, then pain is represented as it essentially is, that is, as something essentially indistinguishable from pain, but in a way that is distinct from how it appears.²²² The materialist may want to say that confusion arises here if one is unfamiliar with the science or neurophysiology of pain (something akin to Dennett's assertions in the first section of chapter one²²³). Such a picture of pain, therefore, may lead us astray. It may lead us to think that C-fibres was anything but pain. The problem here is that pain is *not* distinct from how it appears to us! Pain *just is* the way we experience it! Even worse, this line of reasoning suggests that if pain is represented as the firing of C-fibres, less the felt aspect, it isn't represented as pain at all!²²⁴ That is, the suggestion is that C-fibres may fire in the absence of felt-pain.

So, it seems that there can be no appeal to the befuddling effects of mental imagery in order to show that a statement like (7) is the result of blurry thinking due to the psychological rain of the imagination obscuring one's modal vision. Two possible reactions are left to the materialist: First, deny the analogy between water and pain;

²²² That is, (P represents x as neurophysiological but not pain-like → x is still represented as actual pain) → x is represented as pain actually is.

²²³ Or see Op cit. Thomas Polger, (1999).

²²⁴ However, pain-like = the ways in which pain bothers the senses → x is not represented as pain essentially is. Therefore, x is represented as neurophysiological but not pain-like *and* it is false that x is still being represented as actual pain.

second, deny that pain and its qualitative aspects are essentially related. The former denial accepts that one cannot use Kripke's strategy and the latter denial misrepresents pain. Either way there is *no reason* to give up (7) and there is no reason to give up the possibility that there may be a pain in the absence of C-fibres firing or that there may be C-fibres firing in the absence of pain.²²⁵ That is, one can have either of the intuitions captured by the following statements:

(9) Pain could occur in the absence of the firing of C-fibres

Or

(10) C-fibres could fire in the absence of pain

²²⁵ I should note that Christopher Hill thinks he presents an alternative account to Kripke's of why one intuits the possibility in question based on Thomas Nagel's distinction between phenomenological and material imagination. One has an intuition captured by a statement like (7) because the phenomenal imagination and material imagination splice together an image of the possibilities in question. But Hill's explanation is unsatisfying, because one can merely say "so what?" A Kripkean identification of modal error allows one to clearly see just where one went wrong and what is being confused, that is, what objects are being confused; or my picture based theory allows one to see what functions of the imagination are being confused. Both allow one, thereof, to recalibrate one's thoughts in order to see the world correctly. One can't do this it seems on Hill's theory, it just seems we are prone to the intuition in question. So why should I think that the intuition is mistaken? In fact, doesn't it suggest, akin to the evolution of two ways of imagining, two distinct worlds to imagine, phenomenal and material? On the picture theory advanced by me, one might say, a picture, for example, of C-fibres firing can be represented as pain phenomenologically, less the firing in question, yet still *as* pain. Two kinds of mechanism may be involved, but why does that suggest a confusion? Hill gives no convincing reason to think so. See, Christopher Hill, (1997), "Imagination, Conceivability, Possibility and the Mind-Body Problem", in *Philosophical Studies*, 87, 61-85

And Kripke's strategy, of highlighting a psychological confusion is redundant because, in effect, it ends up legitimizing the aforementioned statements rather than undermining them. Kripke has said that this "holds for any physical state which might be identified with a corresponding mental state", as I have attempted to suggest in the footnotes, too, and as will be highlighted in a little more detail in the next section.

To sum up, (7) is not ruled out by applying the essentialist strategy that Kripke says can be used to uncover confusions that arise with other statements due to the workings of the imagination. In the terms here, no confusion can be identified with a failure to distinguish (or conflate) the representative and representative-as functions of mental imagery.

So, then, there is no reason to drop (7), no psychological confusion can be located (rather the opposite). However, I want to chart a moderate course. I think, one should only conclude that the dualist cannot be *shown* to be befuddled by the psychological noise of the imagination affecting (no doubt otherwise) sound modal judgments, that is, so far as Kripke's strategy of deploying a qualitative analogue to clear the fog of confusion fails. That's all! Not being able to employ a qualitative analogue to highlight a psychological confusion, I think, isn't really sufficient to justify holding some statement as true. The following passage, in Kripke, implies as much: "That the usual moves and analogies are not available to solve the problems of the identity theorist is, of course, not proof that no moves are available."²²⁶ The strategy of showing that one is confused about what one imagines, that is, showing that the mental rain of the imagination is blurring one's otherwise sound modal judgments is not available to the materialist. *But other moves and*

²²⁶ Op cit. Kripke, (1980), p.155

analogies may well be sufficient for this purpose! This is why, I think, Kripke's argument is moderate in conclusion. There are other obvious reasons for moderation, which I'll try and list. But before I do so I want to reintroduce David Chalmers to the discussion. I want to do this because Chalmers argument is useful for focusing another reason for the moderation I think is called for.

As stated, David Chalmers talks about "negative" and "positive" conceivability. Here, again, is what Chalmers says about positive conceivability in his own words:²²⁷

Positive notions of conceivability require that one can form some sort of positive conception of a situation in which S is the case. One can place the varieties of positive conceivability under the broad rubric of imagination²²⁸

There is a sense of conceivability that is termed "negative":

Negative notions of conceivability hold that S is conceivable when S is not ruled out...More relevant notions of conceivability can be obtained by constraining the ways in which S might be ruled out.²²⁹

For example, S may not be ruled out a priori. Now, as seen, the argument from conceivability against materialism which Chalmers contends has as a key premise

²²⁷ In the following passage S can be understood as standing for a statement.

²²⁸ Op cit. David Chalmers, (2002)

²²⁹ Ibid.

something like the following:

(C) $P \ \& \ \neg Q$ is conceivable²³⁰

The strongest version of the argument takes the conceivability to be *positive* and based on ideal rational reflection.²³¹ To say that the conceivability involved here is positive is to say that what is being conceived, namely $P \ \& \ \neg Q$, can be *imagined* to be true. However, Chalmers says, the conceivability can be taken in a negative sense, too. The full force of Chalmers's 2-dimensional framework is deployed in order to get from this premise to the conclusion that materialism is false, as seen above.²³²

Looking back to the Kripkean argument, it was argued by Kripke, and reinterpreted here in terms of the picture theory of imagination that a statement like (7) cannot be shown to be false based on the psychological rain of the imagination fogging our view of the true modal nature of pain. In these terms, I said, there is no reason to give up (7). However, I called for a moderate conclusion because, as Kripke says, there may be other moves and analogies to be deployed against (7). Here is another reason why, I think, moderation is needed in relation to the way Chalmers has been presented: If we compare (7) with (C) there is one obvious difference, (7) is a statement of *possibility* and (C) is a statement of

²³⁰ One can note this is a consequence of (7), where P abbreviates 'there is something physical going on' and $\neg Q$ abbreviates 'it is false there is some associated qualitative mental state present'.

²³¹ One can review section three of chapter one if one wishes.

²³² Recall, Chalmers 2-D framework divides meaning, conceivability and possibility into two types: epistemic and metaphysical. It won't be discussed in any detail here, but one can refer to the last section of chapter one if one wishes to review a bit. Chalmers's actual conclusion to his conceivability argument is, I think, very interesting, which can also be reviewed in chapter one.

conceivability. As said, Chalmers brings the full force of his 2-dimensional framework to bear on getting from a statement like (C) to (7). As I see it, this is problematic for three reasons: First, the bridge from conceivability to possibility is, and always has been, controversial.²³³ Just as we have seen we may think we say or imagine something that is *in fact* impossible and *in fact* impossible to imagine. On the other hand, not being able to rule something out, does not imply that it is ruled in, something that Kripke seems to adhere to, too. Second, the 2-dimensional framework, which bridges the gap for Chalmers, is itself controversial.²³⁴ That's controversy upon controversy, right? Third, because Kripke's strategy fails to provide a reason to give (7) up, and (7) is sufficient to defeat (6), and the negation of (6) is sufficient to suggest that materialism is false, there is no need for (C), it's *gratuitous*. Moreover, if (7) is true, and (6) is false, then (C) is explained; which is in stark contrast to the controversial nature of getting from (C) to (7). Not only, then, do, I think, conceivability muddies the waters, but, I think, Kripke's argument is sufficient *alone* to suggest the falsity of materialism *if you want to claim materialism is false*, that is. For these reasons, then, I think, Kripke's views moderately speaking ground the view that (7) can't be ruled out, not that (7) is conceivable, not that one can take it as such and use it as a bridge to possibility, and that if wants to do that, taking (7) as it is, as a modal claim, is sufficient to refute materialism alone. It may be thought that accepting the validity of Kripke's argument, which establishes that (7) can't be ruled out, legitimizes

²³³ See Stephen Yablo, (1993), "Is Conceivability a Guide to Possibility", in *Philosophy and Phenomenological Research*, (53) 1, 1-42, Cf. Thomas Reid, (1785), *Essays on the Intellectual Powers of Man*, extracted www.earlymoderntexts.com/authors/reid.html, 01/12/2013

²³⁴ See, for example, Scott Soames, (2006), "The Philosophical Significance of The Kripkean Necessary *Aposteriori*" in *Philosophical Issues: A supplement to Nous*, Volume 16, Issue 1, September 2006, 288-309; Stephen Yablo, (2000), "Textbook Kripkeanism and the Open Texture of Concepts", in *Pacific Philosophical Quarterly*, Volume 81, Issue 1, March 2000, 98-122,

a conceivability in the negative sense, but, if so, it makes *the possibility the ground* of the conceivability and not vice versa. Well, to be sure, I want to chart a moderate course, so I don't accept Kripke's argument as a refutation of (6), nor a refutation of materialism, I do however accept that (7) is not ruled out and that this is significant. I accept (7), that is, as a legitimate hypothesis about mind.

Here are more reasons for further moderation: First, so far as Kripke's argument bears on conceivability, which it surely does, the conceivability in question may just be a psychological affair connected to human psychology. For example, it doesn't seem possible to picture something, x, as a square circle, even if x is a square represented as a circle. Even if there were square circles, it just seems impossible to picture square circles or picture something *as* a square circle. Likewise, it doesn't seem possible to picture something, x, as living and dead, even Schrodinger's cat defies our pictorial mental abilities. Even if, apparently, the cat that Schrodinger envisaged could be in such a state, it doesn't seem possible to picture a *single* cat as *both* dead *and* alive. This suggests that the limits on what we can imagine may not just be factual, but involve other psychological criteria, perhaps, as McGinn argues. Perhaps, the same thing is going on when we fail to picture pain as physical.²³⁵ So, there is reason to be a little cautious. However, of course, just a little caution is called for in this sense because such a warning may extend to a conclusion to everything we can't seem to picture. So we might come to argue the same in relation to squares and circles and, then, what else, but scepticism. Second, as said, I don't think there is anything in Kripke that suggests he is giving us a refutation of materialism. Rather, Kripke says a statement like (7) is intuitive, and the intuition cannot

²³⁵ Op cit., Colin McGinn, (1993)

be ruled out by deploying the usual moves and analogies. He seems, for sure, sceptical about the possibility of deploying other strategies, and even talks about some arguments for materialism as having an “ideological” base, but he doesn’t take his argument as a cut and dry refutation of materialism, as far as I can tell. This is in accord with Yablo’s writings, who writes in relation to jumping from not ruling a statement out to its possibility, as “odd” thus: “the poorer my evidence for p’s truth, the better evidence for its possibility”.²³⁶ That, indeed, would be a strange pattern of reasoning. I will give one more reason for caution, something that may be apparent from the discussion of Kripke above. According to Kripke the facts determine modality, but logic can tell us a priori how the facts will determine modality. For example, say we don’t know what water is, but we think either water is H₂O or XYZ. Logic tells us (abbreviating water = XYZ by P and water = H₂O by Q) that: $P \rightarrow \Box P$; $Q \rightarrow \Box Q$; $\neg(\Diamond P \ \& \ \Diamond Q)$; $\neg\Diamond P \vee \neg\Diamond Q$; $\Diamond P \rightarrow \neg\Diamond Q$; etc. as it should be following the Kripkean logic consistent with the review in section one of chapter one. These formulas reflect my *ignorance* in the *best* possible way.²³⁷ The wrong way to reflect this ignorance is like this $\Diamond P$, or colloquially, ‘P is possible’. Why? Because if *in fact* Q, then $\Box Q$, *regardless of whether I know or not*, such that $\Box\neg P$, that is, $\neg\Diamond P$. On uttering ‘P is possible’, even if ignorant of the fact that Q, one will not even be talking about what one thinks one is talking about, but rather talking about a qualitative analogue, and one will have lost touch with one’s words, meanings and thoughts because of an *incautious* way of talking or defining the context of one’s discussion. In such a case, not only is one *ignorant* about the fact but one is talking *confusedly* about the fact, and if one acts wilfully in pursuing this course, given one knows that one is ignorant of the facts

²³⁶ Op cit. Stephen Yablo, (2004), section 3

²³⁷ Everyone is ignorant to some degree because no one is omniscient, except for God.

and may be speaking confusedly, one will be stalking truths on irrational grounds. Thus if, *so far as I know*, P is a hypothesis and Q an alternative hypothesis (or set of hypotheses) and there is no fact of the matter or reason to dispense with the one over the other, then I will not affirm, for example, $\Diamond P$, but will affirm something like $\neg(\Diamond P \ \& \ \Diamond Q)$, or $\Diamond P \rightarrow \neg\Diamond Q$, etc. It is only permissible to state $\Diamond P$ in conditional form, that is, only if it is stated as an assumption in an argument or hypothesis in order, for example, to show its consequences *if true* or in order to scrutinize it in an attempt to falsify it. This leads me to the conclusion that, so far as there is no reason to put dualism aside, its possibility can only be stated in the aforementioned ways, in a logically clear way and as an assumption or hypothesis to be falsified either logically or empirically. That is, Kripke's argument doesn't refute materialism and establish dualism. But it keeps alive a mere hypothesis.

Let me conclude by summarizing: I think that Kripke's argument has to do with modality and not conceivability. Conceivability muddies the waters and indeed the strategy deployed by Kripke was an exercise in turning the psychological noise of the imagination down not employing the imagination as grounds for modal conclusions. Furthermore, the imagination may just be psychologically tuned to picturing pain and its physical nature distinctly, as it seems to be tuned to picturing squares/circles, living cats/dead cats, etc. distinctly. Next, there isn't anything in Kripke to take him as positing an argument from conceivability against the materialist, anyway. Moderate conclusions are called for; to jump from something not being ruled out to its possibility, as suggested by Yablo would be bizarre. Last, ignorance and the possibility of confusion speaks against blind acceptance of (7), (9) and (10); however, they may be assumed in conditional form or hypothesized in order, for example, to falsify them or work out their consequences.

One last point, if the other arguments for dualism mentioned in section three of chapter one are possibility arguments, then, I think, they're acceptable. If they are conceivability arguments, I think, one needs to wonder why they need to be taken as such, in the spirits of the argument just advanced. For my part, for caution sake, I will consider them as possibility arguments.

So, I conclude that, Kripke's argument is not a refutation of materialism, *but it is significant*. Kripke writes, "Materialism, I think, must hold that a physical description of the world is a *complete* description of it, that any mental facts are 'ontologically dependent' on physical facts in the straight-forward sense of following from them by necessity. No identity theorist seems to me to have made a convincing argument against the intuitive view that this is not the case."²³⁸ Ultimately, then, just so far as (7), (9) and (10), too, haven't been shown to be false by deploying Kripke's strategy, there is *no reason* to give (7), (9) and (10) up. I'll go on to consider the conditions needed to establish dualism in conjunction with Kripke's argument, in quite an austere form, and subject that form to the other moves and analogies talked of by Kripke in order to see if dualism can't, after all, be falsified.

²³⁸ Op cit. Kripke, (1980), p.155

Section E: Private Knowledge and Other Problems

Above it was said that there was no reason to reject, by way of establishing a psychological confusion, the following kind of statement, it's possible that pain could occur in the absence of the firing of C-fibres and vice versa. Following the last branch, this means the firing of C-fibres could have occurred in the absence of the associated feelings of pain. More, generally, this can be stated in the following terms, a neurophysiological process could occur in the absence of the associated qualitative mental state. And even more generally, a physical process could occur in the absence of any associated mental state. And this can be stated, also, in the following manner, it is false that necessarily a physical process entails an associated qualitative mental state. These kinds of statement cannot be shown to be false according to Kripke by appealing to the psychological confusion associated with the mental rain of the imagination, which so easily can obscure our modal judgments and the meaning of what we actually say and mean. For example, the picture theory of the imagination advanced above explains how one may come to psychologically confuse one thing for an essentially distinct thing, water was the archetype here: If P is a picture of water, then it denotes water and H₂O. If P, however, is a picture of water that represents it as lacking certain qualities associated with it, for example, its essential molecular characteristic, P represents water as *water-like*, that is, by representing it as tantalizing the senses in the ways that water does. But not as actual water. The resemblance between the two objects is so strong that one may get confused about what it is that one is imagining and even saying when one talks about water. But the same argument does not work with pain. No psychological confusion can be identified following Kripke's strategy, therefore, there is no reason to abandon a statement that

negates the essentiality of the relationship between pain and some neurophysiological correlate. If P is a picture of pain that represents it as lacking certain characteristics, including a neurophysiological correlate, but, nevertheless, represents pain as *pain-like*, that is, as bothering the senses in the way it does, it *still* represents pain *as* pain. If P is such a picture, P refers to descriptions (including labels) that *essentially* refer to the properties associated with pain. So we aren't talking about a qualitative analogue here, at all, but pain itself. That is, no qualitative analogue is (or can be) confounded with actual pain and no psychological confusion can be associated with any related statement or what is said, meant or imagined. And, in the other direction, if P is a picture of pain that represents it as some purely physical correlate, less the troublesome aspects of its feel, P is a picture that fails to represent pain *as* pain. And there is no misrecognition of pain *as* pain due to a lack of knowledge about or conceptual paucity in relation to the essential characteristics of pain. If there are no other moves or analogies that can convince, then it is reasonable to accept that it is false that a physical process and an associated qualitative mental state are identical, whereby, there is no reason to accept that necessarily a physical process entails an associated qualitative mental state. That is, there is no essential kinship between the two. So far, then, these kinds of proposition seem to be *genuine* modal possibilities, that is, they seem to be *logically possible*. Recall, it was said that if something was a genuine modal possibility, it was logically possible and if something was a disingenuous possibility, it was logically impossible. For example, the notion that Richard Dawkins could have been a rocker is genuinely possible because the associated possibility is not logically impossible. But the notion that squares could be circles is not a genuine modal possibility because the associated possibility is logically impossible. Absurdity is not involved in the former case, but it is in the latter case. Likewise, there

seems to be no logical inconsistency in thinking that it is false that a physical process entails an associated qualitative mental state, for example, so far as it can't be shown that the mental rain of the imagination is obscuring thought here. It follows that it is *reasonable* to entertain the conclusion that it is false that necessarily a physical process entails an associated qualitative mental state. So far, then, as it is assumed it is false that necessarily a physical process entails an associated qualitative mental state, the following proposition can be hypothesized:

(K) $\diamond(P \ \& \ \neg Q)$ ²³⁹

This is true for any proposition about physical truth that P stands for and any proposition about the presence of an associated qualitative mental state that Q stands for, for example, if we replace P with 'there are C-fibres firing' and Q with 'there is an associated feeling of pain', or if we replace P with 'the whole of China is functionally organized at time t in exactly the same way as a human brain is functionally organized at time t' and Q with 'there is an associated smell of vetiver', or if we replace P with 'there is scratching behaviour' and Q with 'there is an associated feeling of itchiness', etc. That is, whether physicalism is understood to be neurophysiological, functional or behavioural, (K) is true. In fact, given these specifications of physicalist positions and the application of Kripke's argument to them, different versions of (K) can be rationally hypothesized:

(K)^N $\diamond(N \ \& \ \neg Q)$

²³⁹ Note, this means that the following are, of course, also false, $\Box(P \leftrightarrow Q)$, and $\Box(P \rightarrow Q)$, etc.

$$(K)^F \diamond (F \& \neg Q)$$

$$(K)^B \diamond (B \& \neg Q)$$

Here N stands for a proposition capturing a neurophysiological characterization of the mind, F a proposition capturing a functional characterization of the mind, B a proposition capturing a behavioural characterization of the mind. So, the first denies it is necessary that a neurophysiological characterization of the mind entails an associated qualitative mental state; the second denies that it is necessary that a functional characterization of the mind entails an associated qualitative mental state; the third denies that it is necessary that a behavioural characterization of the mind entails an associated qualitative mental state. I take it, that neurophysiological, functional and behavioural characterizations of the mind are physical. So (K) entails (K)^N, (K)^F and (K)^B. That is, for example, with respect to some assumed functional characteristics of the mind, to hypothesize that it is false that necessarily a physical characterization of the mind entails a qualitative mental state, implies that it is false that a functional characterization of the mind entails a qualitative mental state. Another way to state this is, for example, this time with respect to a behavioural characterization, if necessarily a behavioural characterization of the mind entails an associated qualitative mental state, then necessarily a physical characterization of the mind entails an associated mental state. But given the latter is false (assuming (K)), the former is false. The result is (K)^B. This can be stated in the following terms:

$$(K)^+ (K) \rightarrow (K)^B$$

More fully,

$(K)^+ (K) \rightarrow (K)^N, (K)^F, (K)^B$ abbreviated to $(K)^{NFB}$

Finally, I don't think anything can be gained by saying that neurophysiological, functional or behavioural characterizations of the mind are not physical characterizations of the mind *except, of course, the defeat of physicalism itself!* Moreover, Kripke's argument can be applied to neurophysiology, function or behaviours *individually* in order to derive $(K)^N$, $(K)^F$, and $(K)^B$. For example, consider the following proposition:

(B) Pain is B (where B stands for a proposition listing an associated cluster of pain behaviours, groans, grimaces, flinches, verbalizations, etc.)

It is not possible to show that $\diamond\neg(B)$ is based on a psychological confusion. That is, if P is a picture of pain that represents it as lacking certain characteristics, including behavioural correlates, but, nevertheless, represents pain as *pain-like*, that is, as bothering the senses in the way it does, it *still* represents pain *as* pain. If P is such a picture, P exemplifies descriptions (and labels) that *essentially* refer to the properties associated with pain. So we aren't talking about a qualitative analogue here, at all, but pain itself.²⁴⁰ So no aspersions of psychological confusion can be cast on any associated hypothesis denying any essential relation between pain and pain behaviour.²⁴¹ This argument is

²⁴⁰ The argument can be presented thus: (P represents x as pain-like but not behavioural \rightarrow x is not represented as actual pain) \rightarrow x is represented as a qualitative analogue of pain. But, pain-like = pain. Therefore, x is not represented as a qualitative analogue of pain. Therefore, x is represented as pain like and it is false that x is not represented as actual pain.

²⁴¹ Cf. Hilary Putnam, (1975a)

sufficient to undermine the identity of the relationship between pain and pain behaviour, but here is its mirror image. This time one can show no confusion based on a lack of knowledge or conceptual paucity is involved. This time, taking P as a picture that represents pain as pain behaviour, devoid of the qualitative aspects of pain, should allow the behaviourist to argue that pain is *still* being represented as pain, since pain is being represented as pain-behaviours and *that's all pain is*, right? Yet, this is even more obviously false than the previous example. If P is such a picture, P exemplifies descriptions that *do not essentially* refer to the properties associated with pain, since P does not exemplify any aspects of the qualitative aspects of pain. So pain isn't be being represented as pain at all here.²⁴² And no aspersions of ignorance or conceptual paucity can be cast on anyone denying any essential relation between pain and pain behaviour, here.

Now, *if* (K) is true, it follows that I could have been identical to my present physical self in all physical respects less the qualitative aspects of subjective experience that I enjoy and suffer presently because (K) says that it is logically false that necessarily any physical process entails any associated qualitative state. No contradiction is entailed, therefore, by the possibility of being physically identical to myself but qualitatively distinct since no logical absurdity seems to be entailed by accepting the modality associated with (K). So, if (K) is true, I could have been physically identical to my present self, but lacked the

²⁴² Again, an argument can be presented in the following manner: (P represents x as behavioural but devoid of the qualitative aspects of pain → x is still represented as actual pain) → x is represented as pain actually is. But, pain = the qualitative aspects of pain. Therefore, x is not represented as pain essentially is. Therefore, x is represented as behavioural, but not pain-like *and* it is false that x is still being represented as actual pain.

qualitative aspects that accompany my subjective experiences, for example, I could have lacked the qualitative aspects associated with the way that water tantalizes my senses, say, the cool refreshing feel of it against my skin on a hot summers day. And this is analogous to the possibility associated with thinking that Richard Dawkins could have been a rocker. Likewise, for other beings that enjoy and suffer the qualitative aspects of their subjective experiences, Thomas Nagel, for example. If (K) is true, then, he could have enjoyed the fruits of his physical existence less the qualitative aspects of subjective experience that he enjoys and suffers presently, for example, the troublesome aspects of the pain he feels on stubbing his toe. Let's say that if Thomas Nagel had been born to such a world he would have been a physical equivalent of himself but not an experiential equivalent of himself. Of course, the reasoning applies to neurophysiological, functional and behavioural equivalents, too, given (K)⁺. I'll baptize this hypothetically possible physical duplicate that lacks the qualitative aspects associated with subjective experience 'Zagel', it will be important in drawing some conclusions below.

So just what conclusions will be drawn with the help of Zagel? I will argue, first, that there are certain logical properties inherent in being able to know what it's like to be another, and Thomas Nagel and Zagel can be distinguished by these properties, as Thomas Nagel can be distinguished from inanimate objects, fictional characters and creatures like bats. Second, I'll argue that if what Thomas Nagel and Zagel *know* cannot be distinguished by their physical attributes, an absurdity can be derived. Third, I'll argue that the dualist must, thereof, give up the notion that all knowledge is determined physically, because of the aforementioned absurdity and because of the limited options open to her. This, I argue, leads to the conclusion that dualism is committed to *a form of*

private knowledge, which is suitably defined below. And also it initiates an ontological and epistemological dualism, a full-bodied and rounded kind of dualism. I explore whether this outcome is tenable against several lines of possible attack and address some of the issues raised in chapter one section one, for example, in Dennett's term, I explore if in fact dualism is "ludicrous". To this end, I try to derive a number of absurdities in order to falsify dualism, of which the argument to private knowledge is just one. I will, thereof, explore some of the other moves and analogies Kripke has been quoted as referring to. This section, then, can ultimately be seen as asking and answering the question, are there any other reasons, other than the psychological reasons presented and rejected by Kripke, to give up (K)? I answer this question in this second section.

Thomas Nagel has already been mentioned above and he is important to what comes next. Nagel famously argued that you and I cannot know what it is like to be a bat, though I can know what it's like to be you and vice versa.²⁴³ This raises the question, what is it for one entity to be able to know what it's like to be another entity? Here is an answer to this question:

(R) x can know what it's like to be y iff:

- (a) x and y are conscious,
- (b) x and y experience the world in the same way with respect to the qualitative aspects associated with the set of experiences that x and y have, and
- (c) x has the cognitive abilities associated with knowledge.

²⁴³ Thomas Nagel, (1974), "What is it like to be a bat?", in *Philosophical Review*, 83, 435-50

The first condition requires that there is something it is like to be both x and y , for example, that there is something it is like to smell vetiver for each individual. The second condition requires that the experiences that x and y enjoy and suffer are qualitatively the same, for example, that what it's like to smell vetiver is the same for both x and y . The third condition requires cognitive competence of x so far as the ability to know is concerned. That is, that x has the capacity to *know* what it is like to smell vetiver, be in pain, see red, etc.

So defined, this relation has certain logical properties and limits. The relation seems to be reflexive. That is, if x can know what it is like to be y , xRy , then x is conscious and x obviously experiences the world in the same way as itself with respect to the qualitative aspects associated with the set of experiences x has, and x has the cognitive abilities associated with knowledge. It follows that x can know what it's like to be x , xRx . However, the relation does not seem to have many other logical properties often associated with relations. For example, the relation is not symmetrical, if xRy , then even though x and y are conscious and x and y 's experience are qualitatively the same, it does not follow that y has the requisite cognitive abilities. Therefore, it would be false to conclude that yRx .

What else can be understood by this relation? Let's take two individuals: Thomas Nagel and David Chalmers. Thomas Nagel can know what it's like to be David Chalmers, because Thomas Nagel and David Chalmers are both conscious, their experiences are qualitatively the same, and Thomas Nagel has the associated cognitive abilities. However, Thomas Nagel cannot know what it's like to be a rock, because a rock is not conscious

and it is unable to experience the world in any way. Neither can he actually know what it's like to be Sherlock Holmes, for Sherlock Holmes isn't conscious and there isn't anything it is like to be Sherlock Holmes.²⁴⁴ Consider, next, the physical (neurophysiological, functional or behavioural) duplicate of Thomas Nagel I mentioned above and baptized Zigel, Thomas Nagel cannot know what it is like to be Zigel, because, by definition, Zigel isn't conscious i.e. there isn't anything it is like to be Zigel. Last Thomas Nagel cannot know what it's like to be a bat because, although a bat is arguably conscious, Thomas Nagel and a bat do not experience the world in the same way with respect to the associated qualitative aspects of experience, that is, their experiences are qualitatively distinct.

Another way to think about knowing what it's like to be another invites a sense of reciprocity. A pertinent question is, what is it for *two entities* to be able to know what it's like to be *one another*? It is probably this kind of relationship that Nagel has most in mind in his paper, for he seems to talk about a species-wide relation.²⁴⁵ A similar answer as that provided above can be given here, too, I think:

(R)⁺ x can know what it's like to be y iff:

- (a) x and y are conscious,
- (b) x and y experience the world in the same way with respect to the qualitative aspects

²⁴⁴ I think, perhaps, one can think one imagines one knows, but one can't actually know because Sherlock Holmes is not a conscious being.

²⁴⁵ Op cit. Nagel, (1974)

associated with the set of experiences that x and y have, and
(c) x and y have the cognitive abilities associated with knowledge.

Again, the first condition requires that there is something it is like to be x and something it is like to be y . The second condition requires that the experiences of x and y are qualitatively the same. The third condition, this time, slightly differently, requires that *both parties* are cognitively competent, that is, have the capacity to know. This relation has a larger number of logical properties and limits. For example, once again, the relation seems to be reflexive. If xRy , then x is conscious, and, of course, x experiences the world in the same way as itself with respect to the associated qualitative aspects of experience it has and x has the relevant cognitive abilities. Since, then, x is conscious and x experiences the world in the same way as itself and x has the requisite cognitive capacity, x can know what it's like to be itself, that is, xRx . Moreover, $(R)^+$ is symmetric because if xRy , then x is conscious and y is conscious, and x experiences the world in the same way as y with respect to the qualitative aspects of experience they have, and x and y have the relevant cognitive capacity and, therefore, y and x are conscious, and y and x 's experiences are qualitatively the same, and y and x have the relevant cognitive competency, giving us yRx . $(R)^+$ is also transitive because if xRy and yRz , then x , y and z are conscious, x and y 's experiences are qualitatively the same, and y and z 's experiences are qualitatively the same, which means x , y and z 's experiences are the same, and x , y and z are all cognitively competent. Since, then, x and z are conscious and their experiences are qualitatively the same and x and z are cognitively competent, x can know what it is like to be z , that is, xRz . The relation is also shift-reflexive, that is if xRy , then y is conscious, and experiences the world in the same way as itself with respect to the

qualitative aspects associated with experience it has and is cognitively competent, whereby, y can know what it's like to be y , so, yRy . The difference between (R) and $(R)^+$ is that the cognitive capacity of y in the xRy relation is not necessary to the former as it is to the latter.

Again, these properties can help to clarify the quality relations between entities and sets of entities. Thomas Nagel can get to know what it's like to be David Chalmers, because each is conscious and each experiences the world in qualitatively the same way and they both have the relevant cognitive capacities. We can conclude, that David Chalmers can know what it's like to be David Chalmers (by shift-reflexivity) and Thomas Nagel (by symmetry), and any third individual that Thomas Nagel can come to know (by transitivity). A rock cannot know what it's like to be a rock because a rock is not conscious and it is unable to experience the world in any way. It follows that Thomas Nagel cannot know what it is like to be a rock (by shift-reflexivity). Take two more entities, Thomas Nagel and Sherlock Holmes. Again, Sherlock Holmes cannot know what it's like to be Sherlock Holmes, for there is nothing it is like, I am assuming, to be Sherlock Holmes. It follows that Thomas Nagel cannot know what it's like to be Sherlock Holmes (by shift-reflexivity).²⁴⁶ Introducing Zagel again, Zagel cannot know what it's like to be Zagel, because, by definition, there is nothing it is like to be Zagel. This is true even if we assume that Zagel has the requisite cognitive capacity. And so it follows, Thomas Nagel cannot know what it's like to be Zagel (by shift-reflexivity). Last, Thomas Nagel and a bat.

²⁴⁶ As said, perhaps, one can think one can imagine one knows, but one can't actually know because Sherlock Holmes is not a conscious being. Perhaps, too, one can think one imagine Holmes has outstanding cognitive prowess, however, he doesn't actually have such powers.

According to Nagel there is something it is like to be a bat. However, $(R)^+$ fails (specifically with respect to a bat's capacity to circumnavigate the world by sonar). That is, Thomas Nagel cannot know what it's like to be a bat, so a bat cannot know what it's like to be Thomas Nagel (by symmetry). And, likewise, anyone that can know what it's like to be Thomas Nagel, say David Chalmers, will not be able to know what it is like to be a bat either. The relation breaks down in this case because the experiences are not all of the same kind with respect to the qualitative aspects associated with the experiences that Thomas Nagel enjoys and suffers and those that a bat enjoys and suffers. This would be true even if bats had the requisite cognitive capacity.

Here is another version of the kind of relation under discussion. This time the relation considers knowing what it's like to be another with respect to *a specific quality*, for example, pain. I think, what it is for one being to be able to know what it's like to be another entity with respect to pain can be given the following condition:

$(R)^{\text{pain}}$ x can know what it's like to be y iff:

- (a) x and y are conscious,
- (b) x and y experience the world in the same way with respect to the qualitative aspect associated with the pain experiences that x and y have, and
- (c) x has the cognitive abilities associated with knowledge.

Again, the first condition requires that there is something it is like to be both x and y. The second condition requires that the experiences that x and y enjoy and suffer, i.e. pain, are

qualitatively the same. The third condition requires the requisite cognitive competency of x . Here are the logical properties and limits. Again, the relation seems to be reflexive. That is, if x knows what it is like to be y with respect to pain, then x is conscious and x obviously experiences the world with respect to pain in the same way that x experiences the world with respect to pain and x has the cognitive abilities associated with knowledge, that is, xRx . However, again, given (c), the relation does not seem to have many other logical properties sometimes associated with other relations. The relation is not symmetrical, if xRy , then it follows that y is conscious and y experiences the world in the same way as x with respect to pain, but it does not follow that y has the requisite cognitive abilities. Therefore, it would be false to conclude that yRx . Neither is the relation transitive or shift-reflexive. A version of $(R)^{\text{pain}}$ could be stated that allowed for the aforementioned, that is, by adding a condition that required x and y both to be cognitively competent, one would introduce the idea of reciprocity captured by $(R)^+$ to the relation, call this relation $(R)^{\text{Pain}+}$.

Here are some conclusions that can be drawn from $(R)^{\text{Pain}}$, Thomas Nagel can know what it's like to be himself with respect to pain; Thomas Nagel can know what it's like to be David Chalmers with respect to pain because both are conscious and experience the world in the same way with respect to pain and Thomas Nagel has the required competency; Thomas Nagel cannot know what it's like to be a rock in pain because there is nothing it is like to be a rock; Thomas Nagel cannot know what it's like to be Sherlock Holmes in pain because there is nothing it is like to be Sherlock Holmes; Thomas Nagel cannot know what it's like to be Zagel in pain because, again, by definition, there is nothing it is like to be Zagel; however, Thomas Nagel *can* know what it's like to be a bat in pain because bats

are conscious and bats experience the world in the same way with respect to pain, keep in mind that pain is *essentially* how it feels regardless of where, when and by whom it is felt.²⁴⁷ The difference between a rock, Sherlock Holmes and Zagel, on the one hand, and a bat, on the other, is that for the former three there is nothing that it is like to experience the world, let alone, experience pain, but for the latter, there is something it is like to be conscious and in pain.

Let's summarize a bit. Five types of entity have been mentioned:

- A. Types like Thomas Nagel: David Chalmers, etc. i.e. most humans, and similar.
- B. Types like a rock: a slab, stones, trees etc. i.e. inanimate objects.
- C. Types like Sherlock Holmes: Madame Bovary, Doraemon, etc. i.e. fictional characters.
- D. Types like Zagel: Ned Block's China head, David Chalmers's philosophical zombies, etc. i.e. hypothetically possible physical entities that lack the qualitative aspects associated with experience.
- E. Types like bats: octopuses, fish, crows, etc. i.e. non-humans or dissimilar.²⁴⁸

Type A's are most humans or similar creatures that draw from the same pool of

²⁴⁷ Compare what I said was Kripke's phenomenological assumption above, for example.

²⁴⁸ C types and D types may seem to be similar. However, C types are less well defined than D types because we don't know enough about their physical and mental status to classify them as such. Indeed, if more were said about such characters, then they might be reclassified, for example, if Sherlock Holmes was taken to be physically identical to Basil Rathbone. Anyway, the list of five types is not exhaustive and each type has sub-types to consider, too. These are not relevant to the discussion here so have not been mentioned.

experiences; Type B's, inanimate objects; Type C's, fictions; Type D's, if (K) is true, genuinely possible physical entities that lack the qualitative aspects associated with experience, Type E's, non-human or dissimilar creatures which do not draw from the same pool of human-like experiences. It seems:

(a) So far as (R) is concerned: Type A's can know what it's like to be type A's; type A's cannot know what it's like to be type B's; type A's cannot know what it's like to be type C's; type A's cannot know what it's like to be type D's, type A's cannot know what it's like to be type E's.

(b) So far as (R)⁺ is concerned: Type A's can know what it's like to be type A's; type A's cannot know what it's like to be type B's; type A's cannot know what it's like to be type C's; type A's cannot know what it's like to be type D's, type A's cannot know what it's like to be type E's.²⁴⁹

(c) So far as (R)^{Pain} is concerned: Type A's can know what it's like to be type A's; type A's cannot know what it's like to be type B's; type A's cannot know what it's like to be type C's; type A's cannot know what it's like to be type D's, type A's can know what it's like to be type E's.²⁵⁰

²⁴⁹ Strictly speaking, in lieu of (c), one should probably say that *generally* type A's cannot know what it's like to be E's.

²⁵⁰ It is interesting to note that congenital analgesics who do not feel pain mean that so far as (R)^{Pain} applies to type A members it does not mean that every type A member can get to know every other type A member.

It seems that however we construe the relation that Thomas Nagel introduced (and there are more ways than mentioned just now, for example, $(R)^{\text{pain}^+}$), a type A, like Thomas Nagel, will never be able to know what it's like to be a rock, Sherlock Holmes, or, most importantly, here, Zagel in *any* sense (though he might be able to know what it's like to be a bat in *some* sense). What is common to all of these entities is that they lack consciousness. That means, the relationship in question cannot be construed in a way that makes it sensible to say that Thomas Nagel can know what it's like to be any one of them. *It is sufficient, indeed, to say, that x cannot know what it's like to be y in any experiential sense if there is nothing it is like to be y.* This can be contrasted with denying knowledge competency of one or the other party, which is *not* always sufficient to deny the R-type relation. Also, contrast this with the sameness of experience that is required to instantiate an R-type relation. Denying consciousness is more fundamental than this sameness condition because consciousness is necessary for this condition to be realized to whatever degree it is realized. It follows that for all x and all y if either x or y do not enjoy or suffer consciousness, then x cannot know what it's like to be y and y cannot know what it is like to be x. So far as (K) allows us to say that it is possible that Thomas Nagel is physically (neurophysiologically, functionally, behaviourally) identical to an entity like Zagel and Zagel lacks consciousness, then even if Thomas Nagel is physically identical to Zagel, Thomas Nagel cannot know what it is like to be Zagel because Zagel lacks consciousness. Likewise, it follows that even if Thomas Nagel is physically identical to Zagel, Zagel cannot know what it is like to be Thomas Nagel because, again, Zagel lacks consciousness.

It is obvious that, as I have done throughout, Nagel assumed that there is a certain qualitative feel to consciousness. It is *obviously* also assumed in Kripke's argument

leading to (K). It is captured by the following assumption:

(O) There is something it is like to be conscious.

This assumption is widely held by the community of philosophers and scientists who try to explain consciousness. In fact, most philosophers and scientists think it is necessary to explain just *this* feature of the mind if the mind is to be explained at all. Failing to explain this feature fails to provide an adequate theory of the mind.²⁵¹ I'll say more about this below.

To summarize at this point, there seems to be three assumptions that are consistent with the dualist outlook outlined here. They are (O), (R), and (K). Here is another widely held assumption, which will be important to the discussion that follows:

(P) Knowledge is physically (neurophysiologically, functionally, behaviourally) determined.

So, for example, if knowledge is neurophysiological, then it may be determined in the way that John Searle says:

The basic form of consciousness and intentionality are caused by the behaviour of neurons and are realized in the brain system, that is itself composed of neurons.

What goes for thirst goes for hunger and fear and perception and desire and all

²⁵¹ See the first section of chapter one and latter in this section.

the rest.²⁵²

Or if knowledge is determined by functionality, then it might get something like the following definition:

[I]f the concepts of the various sorts of mental state are concepts of that which is, in various sorts of ways, apt for causing certain effects and apt for being the effect of certain causes, then it would be a quite unpuzzling thing if mental states should turn out to be physical states of the brain.

The concept of a mental state is the concept of something that is, characteristically, the cause of certain effects and the effect of certain causes.²⁵³

Then, again, if knowledge is determined behavioural, it might be determined in the following kind of manner:

Modern physicalism...unlike the materialism of the seventeenth and eighteenth centuries, is behaviouristic....In the case of cognitive concepts like “knowing”, “believing”, “understanding”, “remembering”, and volitional concepts like “wanting” and “intending” there can be little doubt, I think, that an analysis in terms of dispositions and dispositions to behave is fundamentally sound.²⁵⁴

²⁵² Op cit. John Searle, (2004), p.115

²⁵³ D. M Armstrong, (1981), “The Causal Theory of the Mind”, in Op cit. Chalmers, (2002), p. 82

²⁵⁴ Ullin Place, (1956), p.55

(P) can be written in the following form, where P, standing for physical truth related to one of the determinants above as what the physical identity of knowledge consists in, entails K, that is, something that is known:

(P) $P \rightarrow K$ ²⁵⁵

This allows one to draw the following conclusion:

(P)⁺ If x is physically identical to y and x knows p, then y knows p.

So, for example, if one is allied to a neurophysiological account of knowledge, then so far as a neurophysiological state determines what it is for x to know p, and x and y are neurophysiologically identical, then y knows p. Or, if we take a functional approach to the matter, so far as x's inputs and outputs lead to the conclusion that x knows p and x and y are functionally identical, then y knows p. Last, if we choose to accept a behaviouristic approach, then inasmuch as a set of behaviours and dispositions to behaviour, linguistic or otherwise, determines what it is for x to know p and x and y are behaviourally identical, then y knows p.

So, so far, I have associated a set of assumptions with the dualist, (O), (R), and (K), and lately noted a certain physicalist notion of knowledge, (P). *Unfortunately, this set of propositions is inconsistent* because we can derive a contradiction. This means, as will be

²⁵⁵ Of course, (K) and K are to be distinguished.

shown, that the dualist has to reject (P), whereof a form of *private knowledge* threatening absurdity in a variety of different ways emerges. Does this, then, falsify dualism? Have we discovered one of Kripke's "other moves and analogies"? Is dualism, as Dennett says, "ludicrous"? Let me, first, try and show how this problem comes about.

From (P) one can draw the conclusion that if x and y are physically (neurophysiologically, functionally, behaviourally) identical and x knows p, then y knows p. If (K), it is possible that Thomas Nagel is physically (neurophysiologically, functionally, behaviourally) identical to an entity I labelled Zigel. It is also, of course, possible for Thomas Nagel to be physically identical to Zigel and for Thomas Nagel to know p, where p abbreviates everything that Thomas Nagel knows. It follows that it is possible for Zigel to know p. That is, it is possible for Zigel to know everything that Thomas Nagel knows. On the other hand, from (R), it follows that for all x and all y if x or y are not conscious, x cannot know what it's like to be y and y cannot know what it's like to be x. If (K), it is possible that Thomas Nagel is physically (neurophysiologically, functionally, behaviourally) identical to an entity like Zigel which lacks consciousness. It follows that even if Zigel is physically identical to Thomas Nagel, Zigel cannot know what it is like to be Thomas Nagel. That is, Zigel cannot know something that Thomas Nagel knows. This leads to the conjunction, Zigel can know p, that is, *everything* that Thomas Nagel knows, *and* Zigel cannot know what it is like to be Thomas Nagel, that is, Zigel cannot know *something* that Thomas Nagel obviously knows. This is manifestly absurd. This is the first absurdity facing dualism.

According to the usual philosophical practice, one of the assumptions must be given up,

but which one? I think the dualist has but one choice and this is what I will attempt to show next. The first assumption, (0), cannot be given up. Dualists cannot drop the assumption without giving up a necessary condition on dualism. Even physicalists cannot give up the assumption without giving up the essential aspects of mental reality. I have already mentioned that most philosophers and scientists assume (0). So to try and explain consciousness, whilst denying (0), is aberrant. We've already seen according to Chalmers *it* is what defines the "hard" problem of mind. Well, John Searle, along with a variety of others, thinks that materialism most often fails, in its many guises, due to the fact that in trying to explain mental reality it deconstructs itself and leaves out the very thing that it is trying to explain, i.e. mental reality or the subjective aspects of experience. Here is what he says:

[Materialism] is a rather easy view to refute, because it denies the existence of things we all know to exist. It asserts that there are no ontologically subjective phenomena, and we know this is false because we experience them all the time. As philosophers we find this sort of refutation unsatisfying because it is too simple, so we invent more complex arguments to make the same point, about bats and colours and inverted spectra and qualia and Chinese Rooms and so on. But this is the point they, in their different ways, are making.²⁵⁶

And, again:

What we find in the history of materialism is a recurring tension between the

²⁵⁶ John Searle, (2004), p.91

urge to give an account of reality that leaves out any reference to the special features of the mental, such as consciousness and subjectivity, and at the same time account for our “intuitions” about the mind.²⁵⁷

Well, what makes the problem of consciousness so “hard” according to David Chalmers is precisely explaining what it is like to be conscious as stated in the first chapter. Leading scientists in the research field also take this aspect of experience very seriously. Ramachandran writes:

Qualia are the ‘raw feels’ of conscious experience: the painfulness of pain, the redness of red. Qualia give human conscious experience the particular character that it has. For instance, imagine a red square; that conscious experience has (at least) two qualia: a colour quale, responsible for your sensation of redness, and a shape quale, responsible for the square appearance of the imagined object.²⁵⁸

Even Daniel Dennett writes: “I agree wholeheartedly, there seems to be qualia”.²⁵⁹

It’s also possible to argue against the eliminativist that it’s not possible for her to demonstrate that (0) is false, for what possibly could *show* the dualist that she was mistaken in holding that there was something it was like to be conscious, given that she

²⁵⁷ John Searle, (1994), p.52

²⁵⁸ V. S. Ramachandran and William Hirstein, (1997), “Three Laws of Qualia: What Neurology Tells Us about the Biological Functions of Consciousness, Qualia and the Self”, in *Journal of Consciousness Studies*, 4, No. 5-6, 429-458, p. 430

²⁵⁹ Daniel Dennett, (1991b), p.372

has personal experience to the contrary, that is, *that there is something it is like to be conscious*, which the majority of us have complete sympathy with. Yet, it must be admitted that there are some philosophers who are eliminativists about mental phenomena.²⁶⁰ Needless to say that for the dualist this position is untenable, and, I think, quite correctly. If eliminativists argue that there is nothing that it feels like to be conscious, then they are plainly wrong because simply enough there is something it is like to be conscious, for example, I know that there is something it is like to smell a skunk and I don't think I much like it. If eliminativists argue that there is something it feels like to be conscious and what it feels like to be conscious is physical, then that is just another version of the identity theory (I think Rorty is advocating something like this)²⁶¹. But that the dualist, and many physicalists, reject on the basis of an argument like Kripke's. Last, the dualist can argue that to reject (O) tends to scepticism since denying (O) denies (R), the conditions that state what makes it possible for one conscious entity to know what it's like to be another conscious entity (and oneself, I'd suggest). I think, then, it is not just impossible to reject (O) for the dualist, but very difficult on a non-dualist basis, too. It doesn't, therefore, look odd for the dualist to hold (O); rather to the contrary.

So, what about (R)? A reason for retaining (R) is that if (R) is denied, it suggests that one cannot come to know what it is like to be another, on the one hand; and one cannot come to know what it is like to be a self, on the other hand. That is, the denial of (R) suggests scepticism about other minds and, worse still, scepticism about one's own mind, for example, through time. That is, to say that it is false that x can come to know what it's

²⁶⁰ For example, see the third section of the first chapter.

²⁶¹ Cf. Richard Rorty, (1965), and Richard Rorty, (1970)

like to be y , where $x \neq y$, is to be sceptical about other minds, and to say that x cannot come to know what it's like to be y , where $x = y$, is to be sceptical about one's own ability to come to know what it's like to be oneself, say, about one's past self and one's present self. I think both of these outcomes are clearly *very limiting*. It seems to me that most philosophers would want to steer clear of such scepticism. I think, this is especially the case for the dualist, who is often faced with accusations that the perspective tends in the direction of scepticism.²⁶² Another way to make the same point is this, if scepticism is false and (0) is true, then (R) follows: if there is something it's like to be conscious and others are conscious, too, then surely it is at least *possible* to know what it's like to be another given the set of conditions stated for (R) are satisfied. If so, this means that if (0) is, indeed, true for the reasons suggested above (it is a necessary condition on dualism) and (R) is false, then scepticism is true.²⁶³

A second reason is existential, (R) is manifest in our everyday reality. I assume certain things which it *feels* necessary for me to assume. For example, my behaviours and responses are often regulated by how I perceive, rightly or wrongly, how others *experience* the world when I interact with them. For example, if I see in someone's behaviours indications of pain, say, groans and grimaces, my behaviour seems to adapt accordingly and quite naturally. I may ask the person if they need my help, I may call an ambulance, emotions may be stirred, I may intervene, feel sickened, and so on. Crucially, I don't think this is just a matter of a functional response, but this is a matter of responding

²⁶² Op cit. Sydney Shoemaker, (1982)

²⁶³ Let S stand for scepticism, then the form of the argument is this: $((0) \& \neg S) \rightarrow (R)$, (0) and $\neg(R)$, therefore, S.

because I believe that I understand that the being in question is conscious and shares the same kinds of experiences as me, at least in the relevant respects, here, pain. So, if I see someone in pain, the psychological impact is the greater if I have a sense of the pain felt, for example, a sense of the severity of the pain in question. The same is true of other examples and associated feelings and emotions, itchiness, happiness, disgust, etc. I believe all these scenarios apply as much to others as to me. Others regulate their behaviours in response to others as I do, but not only in response to behaviours but in response to what they believe to be indications of others conscious mental states. In fact, I think, *we* expect as much of each other. For example, when I tell the doctor about a sharp pain, rather than a dull pain, I expect her to understand the qualitative difference and she expects me to be able to comprehend what feeling I am experiencing in kind, severity and type. As Piccinini says, “In our everyday life, we ask people how they feel or what they think, then use the answers as evidence of their mental states and events. Many scientists use sophisticated variations of the same method”.²⁶⁴ The point I’m trying to make is this, (R) is assumed by us as we go out about our daily business and interact with others.

A particularly interesting existential choice has to do with ethical decisions made with respect to animals. Many people (including me) regulate their behaviour with respect to living creatures because they assume that there is something it is like to feel pain and suffer and they think it is unethical to inflict pain or suffering on another sentient being (human or non-human). Such behaviours obviously assumes that other beings are conscious and share some of the same experiences that we do, that is, even where *non-*

²⁶⁴ Gualtiero Piccinini, (2009), “First Person Data, Publicity, and Self Measurement”, in *Philosophers Imprint*, Volume 9, No. 9, October 2009, p.1

humans are concerned. This is hardly a ludicrous idea, but a *scientific* one. Peter Singer for example cites Lord Brain, thus:

Every particle of factual evidence supports the contention that the higher mammalian vertebrates experience pain sensations at least as acute as our own. To say that they feel less because they are lower animals is an absurdity; it can easily be shown that many of their senses are far more acute than ours—visual acuity in certain birds, hearing in most wild animals, and touch in others; these animals depend more than we do today on the sharpest possible awareness of a hostile environment. Apart from the complexity of the cerebral cortex (which does not directly perceive pain) their nervous systems are almost identical to ours and their reactions to pain remarkably similar, though lacking (so far as we know) the philosophical and moral overtones. The emotional element is all too evident, mainly in the form of fear and anger.²⁶⁵

Singer's philosophical idea is that the behavioural, physiological and evolutionary similarities between living beings (human or non-human) suggest that they are conscious and conscious in some of the ways that we are conscious, especially with respect to pain and suffering and that this is not species-specific. Nagel stresses a similar point when he writes: "Some extremists have been prepared to deny [consciousness] even of mammals other than man."²⁶⁶ John Searle damns the denial of both consciousness and

²⁶⁵ Peter Singer, (1990), *Do Animals Feel Pain?*, Excerpted from Peter Singer, (1990), *Animal Liberation*, (New York: Avon Books), pp.1-2 extracted <http://www.animal-rights-library.com/texts-m/singer03.htm> 16/09/2013

²⁶⁶ Thomas Nagel, (1974), p.1

intentionality. Searle writes of his dog, Ludwig, thus: “I am certain that (a) he is conscious; (b) he is aware of my presence (intentionality); and (c) that awareness produces in him a state of pleasure (thought process).²⁶⁷ He goes on to anathematize the denial in the higher mammals as “breathtakingly irresponsible”.²⁶⁸ The point here is not to defend this view (although highly inclined to accept and live by it as I am) but to remark that (R) is implicated in the ethics of how we treat others (human and non-human) individually and as a species and regulates our thoughts and behaviours accordingly.

A vast number of peoples, individuals and groups of people religious or otherwise, live their lives according to such ethical principles. Another large portion of people are *legally bound* to live their lives according to such principles (whether they believe in them or not) because (R) manifests itself at the *legal* level in the structures of society. For example, consider the following excerpt from the British Animals (Scientific Procedures) Act 1986, where scientific procedures on animals are regulated and defined according to the law as legal or illegal based on the assumption and legal recognition of consciousness and the associated ability to feel pain and suffer:

Subject to the provision of this section, “a regulated procedure” for the purposes of this Act means any experimental or other scientific procedure applied to a protected animal which may have the effect of causing that *animal pain*,

²⁶⁷ John Searle, (2002a), “Animal Minds” in *Consciousness and Language*, (Cambridge: Cambridge University Press), p.62

²⁶⁸ Ibid. p.64

suffering, distress or lasting harm.²⁶⁹

So it seems clear that (R) is not only *assumed* in our everyday interactions with each other, it is *enshrined* in many of our ethical and legal contexts in the social structures that surround us. To deny (R) is to deny a principle with such ubiquitous reach and power over our lives and leaves the philosopher sounding unnatural and out of touch; even possibly criminal.

Well, there is evidence for (R) from the cognitive sciences and neuroscience, too! In recent times cognitive science has come under the influence of Edmund Husserl and phenomenology. Evan Thompson presents an overview of this coming-together in his article “Empathy and Consciousness”.²⁷⁰ He cites both Husserl and Edith Stein (Husserl’s student) in this respect.²⁷¹ Thompson identifies an all pervasive convergence between the fields of phenomenology, as represented in the Husserlian tradition, and cognitive science:

Indeed, there is a remarkable convergence between these two traditions, not simply on the topic of intersubjectivity, but on virtually every area of research within cognitive science, as a growing number of scientists and philosophers

²⁶⁹ Section 2, Animals (Scientific Procedures) Act 1986, <http://www.legislation.gov.uk/ukpga/1986/14/section/2> 16/09/2013 My Italics.

²⁷⁰ Evan Thompson, (2001), “Empathy and Consciousness”, in *Journal of Consciousness Studies* 8, No. 5-7, 1-32, extracted www.imprint.co.uk/pdf/Thompson.pdf, 21/11/2013

²⁷¹ Ibid. p.16

have discussed.²⁷²

Francisco Varela, who is famous for his *neurophenomenological* approach, is an obvious example. He writes, in relation to neurophenomenology:

[I]t is the re-discovery of the primacy of human experience and its direct, lived quality that is phenomenology's foundational project. This is the sense in which Edmund Husserl inaugurated this thinking in the West, and established a long tradition that is well and alive today not only in Europe but world-wide.²⁷³

Even Dan Dennett considers his hetrophenomenology somewhat Husserlian.²⁷⁴ Dennett, for example, describes his hetrophenomenology as:

[T]he third-person parallel to Husserl's notion of bracketing or epoche, in which the normal presuppositions and inferences of one's own subjective experience are put on hold, as best one can manage, in order to get at the core experience,

²⁷² Ibid. p.2

²⁷³ Francisco Varela, (1996), "Neurophenomenology: A Methodological Remedy for the Hard Problem" in *Journal of Consciousness Studies*, 3, No. 4, 330-349, p.335

²⁷⁴ Indeed, Dennett spends a portion of one essay considering "whether hetrophenomenology is a trivial redescription of familiar practices, or a restatement of Husserl with nothing original in it, or a betrayal of Husserl, or a revolutionary proposal on how to study consciousness, or a thinly disguised attempt to turn back the clock and make us all behaviourists, or an outrageous assault on common sense, or something else." In this paper, Dennett describes his approach to Husserl as a "buffet" approach, "take what I like and leave the rest untouched". Daniel C. Dennett, (2006), "Hetrophenomenology Reconsidered", ase.tufts.edu/cogstud/papers/hetroreconsidered.pdf, p.1

as the theory-neutral and unencumbered as possible.²⁷⁵

This phenomenological turn has elevated the status of “empathy” for many under the influence of Husserl and his phenomenology. For example, Thompson identifies empathy as one of the key components of this influence.²⁷⁶ And Gallese, whose team is famous for discovering “mirror neurons”, writes, for example, that “Phenomenology has further developed the notion of *Einfühlung* [empathy]”.²⁷⁷ He accords this development to Husserl, Edith Stein and Merleau-Ponty.²⁷⁸ In a very large part, the phenomenological turn has been driven by the actual neuroscience and reflects a move away from classical cognitive science.

Thompson writes:

Classical cognitive science was cognocentric: it conceived of cognition as the manipulation of affectless representations. New developments, especially in affective neuroscience, have shown that affect and emotion lie at the basis of the

²⁷⁵ Daniel C. Dennett, (2003), “Who’s On First? Hetrophenomenology Explained”, in *Journal of Consciousness Studies*, Special Issue: Trusting the Subject? (Part 1), 10, No. 9-10, October 2003, 19-30, p.22

²⁷⁶ Op cit. Evan Thompson, (2001), especially p.13 onwards.

²⁷⁷ Vittorio Gallese, (2009a), “Mirror Neurons, Embodied Simulation, and the Neural Basis of Social Identification”, in *Psychoanalytic Dialogues*, 19, 519-536, p.525

²⁷⁸ Ibid. pp.525-526, see also, for example, Vittorio Gallese, (2001), “The ‘Shared Manifold’ Hypothesis: From Mirror Neurons to Empathy”, in *The Journal of Consciousness Studies*, 8, No. 5-7, 2001, 33-50, especially pp.43-44

mind.²⁷⁹

Since affect and emotion, feelings of pain or disgust, for example, are primarily phenomenological and their understanding is understanding of what is primarily what it is like to experience them, it should be no surprise that phenomenology and empathy have come back to the fore of psychology.

The importance of empathy to psychology, in general, can be noted:

Human empathy is a psychological construct regulated by both cognitive and affective components, producing emotional understanding. Impaired empathy is a central characteristic of several neurological and psychiatric conditions such as frontotemporal lobar degeneration, autism, and schizophrenia.²⁸⁰

Empathy, then, is an important psychological construct. But how exactly is it defined? It is generally accepted that it has two components, a cognitive component and an affective component. As said while, generally, the cognitive side of the cognitive sciences used to predominate, with its bias for propositional analysis, the affective side has come into prominence. Tracy Cassels et al. contrasts the two thus:

²⁷⁹ Op cit., Evan Thompson, (2001), p.4

²⁸⁰ Simone G, Shamay-Tsoory, Judith Aharon-Peretz and Daniella Perry, (2009), “Two Systems for Empathy: A Double Dissociation Between Emotional and Cognitive Empathy in Inferior Frontal Gyrus Versus Ventromedial Prefrontal Lesions”, in *Brain: A Journal of Neurology*, 2009: 132; 617–627, p.617

Cognitive empathy refers to one's ability to recognize and identify another person's feelings. This is distinct from the affective component in that it focuses exclusively on the cognitive processes and ignores the emotional reactions to others' feelings.

Affective empathy...refers to one's emotional responses to another person's emotion or situation....Typically, this involves experiencing emotions that are similar to those of the other person, but at times, affective empathy can manifest in different emotions.²⁸¹

It is, as noted above by Thompson, the affective side of psychology that is at the fore today. And Gallese, connecting with what Thompson says, gives us a somewhat more psychologically expansive idea of affective empathy thus:

When we empathize with others, we understand what others are feeling, be it a particular *emotion* or *sensory state*.²⁸²

Tania Singer et al. drawing on Gallese define empathy in the following way:

²⁸¹ Tracy Cassels, Sherilynn Chan, Winnie Chung and Susan A. J. Birch, (2010), "The Role of Affective Empathy: Cultural and Bicultural Differences", in *Journal of Cognition and Culture*, 10 (2010), 309–326, p. 310

²⁸² Vittorio Gallese, (2003), "The Manifold Nature of Interpersonal Relations: The Quest for a Common Mechanism", in *Philosophical Transactions of the Royal Society*, (2003) 358, 517-528, p.517

Empathy...broadly refers to be being able to understand what others feel, be it an emotion or sensory state. Accordingly, empathic experience enables us to understand what it feels like when someone else experiences sadness or happiness, and also pain, touch, or tickling.²⁸³

This is the sense of empathy that is important for present purposes. It allows one to think of empathy in the following terms, empathy can be thought of as *knowing what it's like to be in an emotional or sensory state experienced by another*. It should be clear this is relevant to the kind of relation I discussed above, since (R), and related (R) type relations, extrapolate conditions that must be met if one conscious entity is to be able to know what it's like to be another conscious entity in terms of sameness of experience. Empathy, so far as it explains how one entity can understand what another feels or senses, explains this as part of the neurophysiological structure of a conscious entity. That is, neuroscience tells us that empathy, indeed, a relation like (R), is a primitive process that is embedded or embodied in the neurophysiological structures of the brain.

Specific studies can be cited in relation to this claim. For example, so far as the emotion of disgust is concerned, Wicker et al. establish the following:

Our core finding is that the anterior insula is activated both during the observation of disgusted facial expressions and during the emotion of disgust

²⁸³ Tania Singer, Ben Seymour, John O'Doherty, Holger Kaube, Raymond J. Dolan, Chris D. Frith, (2004), "Empathy for Pain Involves the Affective but Not Sensory Components of Pain", in *Science* Vol. 303 20 February, p.1157

evoked by unpleasant odorants. This result indicates that, for disgust, there is a common substrate for *feeling* an emotion and *perceiving* the same emotion in others.²⁸⁴

It has also been shown that where the part of the brain that feels disgust is damaged in the individual that that individual has difficulty in recognizing the emotion in others.²⁸⁵ Both studies suggest a core capacity to understand the disgust in others in the actual capacity to feel disgust in oneself.

Again, so far as a sensation is concerned, here pain, similar findings have been reported: “The finding of empathy-related activation suggests an automatic engagement of empathic processes when perceiving pain in others.”²⁸⁶ This report concludes that some of the same areas of the “pain matrix”, where we feel pain, overlap in our reaction to pain and our recognizing the feeling of pain in others:

Rostral ACC and AI appear to reflect the emotional experience that evokes reactions to pain and constitutes the neural basis for our understanding of the feelings of others and ourselves.²⁸⁷

²⁸⁴ Bruno Wicker, Chrisitan Keysers, Jane Plailly, Jean-Pierre Royet, Vittorio Gallese, and Giacomo Rizzolatti, (2003), “Both of Us Disgusted in *My* Insula: The Common Neural Basis of Seeing and Feeling Disgust”, in *Neuron*, Vol. 40, 655-664, October 30, 2003, p.655, my italics.

²⁸⁵ A. J. Calder, J. Keane, F. Manes, N. Antoun, and A.W. Young, (2000), “Impaired Recognition and Experience of Disgust Following Brain Injury”, in *Nature Neurosci.* 3, 1077-1078

²⁸⁶ Op cit. Singer et al, (2004), p.1160

²⁸⁷ Ibid. p.1160

Gallese cites another study that draws similar results:

[Hutchinson et al.²⁸⁸] reported that neurons responded not only to noxious mechanical stimulation applied to the patient's hand, but also when the patient *watched* pinpricks being applied to the examiner's fingers. Both applied and observed painful stimuli elicited the same response in the same neurons.²⁸⁹

These findings, thus, suggest that we represent the *experiences* of others literally in our bodies, that is, through the same neural processes that are responsible for representing the same kind of experience in oneself. As Thompson says,

[E]mpathy is not simply the grasping of another person's particular experiences (sadness, joy, and so on), but on a more fundamental level the *experience* of another as an embodied subject of experience like oneself.²⁹⁰

Or as Gallese puts it,

We do not just “see” an action, an emotion, or a sensation...internal representations of the body states associated with these actions, emotions and

²⁸⁸ W. D. Hutchinson, K. D. Davis, A. M. Lazano, R. R. Taskev, and J. O. Dostrovsky, (1999), “Pain Related Neurons in the Human Cingulate Cortex, in *Nature Neurosci.* 2, 403-405

²⁸⁹ Op cit. Vittorio Gallese, (2003), p.524

²⁹⁰ Op cit. Evan Thompson, (2001), p. 16, my italics

sensations are evoked in the observer, “as if” he or she was doing a similar action or experiencing a similar emotion or sensation.²⁹¹

So, then, empathetic processes are embodied in the individual, that is, what it takes to know what it’s like to be another is embodied in the one self. This suggests that the capacity to satisfy (R) is embodied in the self.

Furthermore, Gallese takes the presence of mirror matching mechanisms, constituted by mirror neurons or like neurons²⁹², to be the ground of empathy: “My proposal is that also sensations, pains and emotions displayed by others can be empathized, and therefore understood, through a mirror matching mechanism.”²⁹³ In fact, Gallese identifies all the following areas in which they may be implicated: simple movements, complex skills, action intentions, processing action related words, language semantics, emotions like disgust, the perception of pain, emotional and sensory experiences, in general, and Autistic Spectrum Disorder.²⁹⁴

Such mechanisms and, therefore, empathy are understood to be *implicit, automatic, unconscious* and *non-linguistic*:

This *implicit, automatic, and unconscious* process of motor stimulation enables

²⁹¹ Op cit. Vittorio Gallese, (2009), p.527

²⁹² Op cit. Gallese, (2001), p.44

²⁹³ Ibid. p.44

²⁹⁴ Ibid. pp.522-523

the observer to use his/her own resources to penetrate the world of the other without the need for *theorizing* about it, without the need to necessarily use *propositional attitudes*.²⁹⁵

It seems to be the case that we are just programmed *by nature* and her evolutionary dance to employ empathetic or mirroring processes. Again, Gallese: “I think that the primary way of understanding others is *direct in nature*.”²⁹⁶ That is, we are programmed, by evolution, to affectively empathize and mirror the emotions and sensations that others have through the same processes, in part, that instantiate the neurophysiological awareness of the same emotion or sensation in oneself. Again, this doesn’t solve the logical problem of other minds, but it is evidentially suggestive. What it suggests is that the (R) relation, and similar, are satisfied by natural evolutionary processes.

A lot of time has been spent on making this last point out. This is because the discussion below will make reference to what has just come again. For now, let me summarize in relation to the present point: there are anti-sceptical reasons to accept (R), everyday existential, ethical, and cultural contexts in which it is assumed, and naturalistic scientific reasons that provide evidence for its *actual* existence. It seems to me the dualist had better not reject (R) without opening up the position to lines of attack on all these fronts.

The two assumptions which are left are (K) and (P). Well, the dualist *must* reject (P),

²⁹⁵ Op cit. Gallese, (2001), p.41

²⁹⁶ Vittorio Gallese, (2009b), “We-ness, Embodied Simulation, and Psychoanalysis: Reply to Commentaries” in *Psychoanalytic Dialogues*, 19, 580-584, p. 583

because rejecting (K) leaves the dualist no sufficient way of establishing that dualism is true. That is because every sufficient reason to hold dualism entails some form of (K). That can be put in terms of a denial of Seagers' third condition for physicalism, that is, the denial of resolution, the reducibility of phenomenology to physics. There are arguments, apart from the modal argument considered above, that aim to establish dualism, but these also lead to the contradiction noted because they all lead to (K). For example, Block argues that it is false that a functional organization of physical parts entails any associated mental state,²⁹⁷ giving one (K)^F; Shoemaker imagines similarly that functional organization does not entail any given mental state,²⁹⁸ giving one (K)^F; Jackson argues that full knowledge of physical reality does not entail knowledge of mental states and, thereby, physical reality does not entail mental reality,²⁹⁹ giving one (K); Chalmers argues from (epistemic) conceivability to (metaphysical) possibility,³⁰⁰ giving one (K). In all these cases, some variety of (K) can be inferred, that is, it is possible that a functional or physical process occurs without the associated qualitative feel. This will lead, again, to the absurdity noted. Here is a quick example.

Assume, with Block, that China can be physically organized in a way that corresponds to the functional organization of Thomas Nagel at time t ³⁰¹, call this the functional Chinese state (FCS). Then this FCS is functionally identical to Thomas Nagel. Also assume, with

²⁹⁷ Op cit. Ned Block, (1978)

²⁹⁸ Op cit. Sydney Shoemaker, (1982)

²⁹⁹ Op cit. Frank Jackson, (1982)

³⁰⁰ Op cit. David Chalmers, (2009)

³⁰¹ Op cit. Ned Block, (1978)

Block, that the FCS lacks qualitative mental states all the same. This give us $(K)^F$. From the functional version of (P) one can draw the conclusion that if x and y are functionally identical and x knows p, then y knows p. If $(K)^F$ it is possible that Thomas Nagel is functionally identical to FCS and, of course, for Thomas Nagel to know p, where p abbreviates everything that he knows. It follows that it is possible for the FCS to know p. That is, it is possible for the FCS to know everything that Thomas Nagel knows. On the other hand, from (R), it follows that for all x and all y if x or y do not enjoy or suffer consciousness, x cannot know what it's like to be y and y cannot know what it's like to be x. But if $(K)^F$ it is possible that the FCS is functionally identical to Thomas Nagel but lacks consciousness. It follows that the FCS cannot know what it is like to be Thomas Nagel. That is, the FCS cannot know something that Thomas Nagel knows. This leads to the conjunction, the FCS can know p, that is, *everything* that Thomas Nagel knows, *and* the FCS cannot know what it is like to be Thomas Nagel, that is, the FCS cannot know *something* that Thomas Nagel obviously knows. And this is absurd. The same argument can be manufactured for the other examples given. The key point is this, however one defines knowledge, (neurophysiologically, functionally, or behaviourally) because a *truly* sufficient argument for dualism entails $(K)^{NFB}$, a contradiction will always be derivable, by using either $(K)^N$, $(K)^F$, $(K)^B$. That a *truly* sufficient condition argument entails or assumes $(K)^{NFB}$ should be clear by considering what it would mean to hold merely $(K)^F$; one would *not* be in a position to herald dualism because although one would have rejected the functional entailment from P (standing for some proposition about a functional truth) to Q (standing for some proposition about a qualitative truth), one would not have rejected the neurophysiological (and behavioural entailments). So, it follows, so far as Block, or someone accepting his argument for $(K)^F$ is concerned, they are

committed to the acceptance of (K)^N and (K)^B, if they advance an anti-physicalist, pro-dualist stance. Of course, we have seen already, in the third section of the first chapter, Block takes his argument to be an argument against physicalism in general.³⁰²

Dennett says that dualism is “ludicrous”. If dualism has staved off one accusation of absurdity with the ejection of (P), however, perhaps, the physicalist can still press the case against her. That is, the possibility of forcing an absurdity out of dualism is still alive. That is, the physicalist may be in a position to argue that the assumptions now held (O), (R), (K), and $\neg(P)$, in their own terms lead to another absurdity, that is, the absurdity of *private knowledge*.³⁰³ Here is the argument: Let’s say that knowledge of p is private when, for some x, x knows p, where p is an arbitrary proposition, and necessarily if any y knows p, then y = x.³⁰⁴ Denying this defines when to say knowledge of p is public, that is, when it is false that for some x, x knows p and necessarily if any y knows p, then y = x.³⁰⁵ This tells us that, for any x, if x knows p, then it is possible that there is some y such that y knows p and y \neq x.³⁰⁶ But, now, let’s consider a specific proposition, q^x, which abbreviates ‘what it’s like to be in pain *for x*’. In this case, it seems it is not possible that that there is some y such that y knows q^x and y \neq x. Why? Well, how could it be so? There is no logical or epistemological physical (neurophysiological, functional or behavioural) entailment of knowledge of any others’ qualitative mental states (note, neither of one’s

³⁰² Ibid.

³⁰³ This suggest non-scientific knowledge, which really offends the spirit of anti-dualists like Dennett.

³⁰⁴ More formally, $(\exists x)[xKp \ \& \ \Box(y)(yKp \rightarrow y = x)]$, where K stands for ‘knows that’.

³⁰⁵ Stating this formally, $\neg(\exists x)[xKp \ \& \ \Box(y)(yKp \rightarrow y = x)]$

³⁰⁶ That is, $(x)[xKp \rightarrow \Diamond(\exists y)(yKp \ \& \ y \neq x)]$

own). For, first, by (K), it is possible that P and $\neg Q$, where P stands for any statement about a physical truth and Q stands for a statement about the presence of an associated qualitative mental state, to repeat, from the above, if we replace P with ‘there are C-fibres firing’ and Q with ‘there is an associated feeling of pain’, or if we replace P with ‘the whole of China is functionally organized at time t in exactly the same way as a human brain is functionally organized at time t’ and Q with ‘there is an associated smell of vetiver’, or if we replace P with ‘there is scratching behaviour’ and Q with ‘there is an associated feeling of itchiness’, etc. Second, compounding the matter is the negation of (P), keep in mind we are assuming (O), (R), (K), and $\neg(P)$, that is, the denial that *knowledge* of Q is physically (neurophysiologically, functionally, behaviourally) determined. This negation implies that physical evidence does not provide evidence for knowledge of Q. One might sum up by saying that the logical and evidential entailment between P and Q has been denied. Given q^x is a Q-type statement, it follows that q^x cannot be logically or epistemologically inferred from P. The result, if there is something it is like for x to be in pain no one can infer what it is like for x to be in pain logically or evidentially from P. So, then, if there is something it is like for x to be in pain, q^x , and x knows what it is like for x to be in pain, where that implies there is some x such that x knows q^x , and necessarily if any y knows q^x , then $y = x$, it follows knowledge of q^x is a matter of private knowledge.³⁰⁷ But, this outcome is absurd, for surely there is *no* private knowledge.

³⁰⁷ Formally, $(\exists x)[xKq^x \ \& \ \Box(y)(yKq^x \rightarrow y = x)]$

But is private knowledge such a counter-intuitive conclusion? Consider the following remarks by the classical identity theorist Ullin Place:

[T]here would seem to be an intractable residue of concepts clustering around the notions of consciousness, experience, sensation, and mental imagery, where some sort of inner process story is unavoidable. It is possible, of course, that a satisfactory behaviouristic account of this conceptual residuum will ultimately be found...I shall assume this cannot be done and that statements about pains and twinges, about how things look, sound, and feel, about things dreamed of or pictured in the mind's eye, are statements referring to events and processes which are in some sense private or internal to the individual of whom they are predicted.³⁰⁸

Then there is Dennett:

On the face of it, the study of human consciousness involves phenomena that seem to occupy something rather like another dimension: the private, subjective, 'first-person' dimension. Everybody agrees that this is where we start.³⁰⁹

This is Valera and Shear:

[F]irst-person methods are available and can be fruitfully brought to bear on a

³⁰⁸ Op cit. Ullin Place, (1956), p.58

³⁰⁹ Op cit. Daniel Dennett, (2003), p.19, the stress is on the "seem".

science of consciousness.³¹⁰

This is Piccinini:

In psychology and neuroscience, the subjects issuing first-person reports and other sources of first-person data play the epistemic role of (self-)measuring instruments..³¹¹

He adds to this thus: “mental states and events are private in the sense that we each undergo all and only our own.”³¹² So is the idea of epistemologically private knowledge really unreasonable or counter-intuitive? To be honest, these thinkers take different views on the nature of “first-person reports”, which they are mostly discussing here. For sure, Dennett and Piccinini *do not* think that the kind of knowledge in question is private in the sense that such knowledge is not available to a third party. Dennett’s following remarks make this clear: “‘first-person’ investigations fall happily into place in ‘third-person’ hetrophenomenology.”³¹³ Piccinini writes: “Privatism runs directly contrary to one of the most basic principles of scientific methodology: scientific methods must be public.”³¹⁴ Even Valera and Shear, who accept the irreducibility of phenomenology, write in the following way: “[I]t would be futile to stay with first-person descriptions in isolation. We

³¹⁰ Francisco Valera and Jonathon Shear, (1999), “First-Person Methodologies: What, Why, How?” in *Journal of Consciousness Studies*, 6, No.2-3, 1999, 1-14, p.2

³¹¹ Op cit. Gualtiero Piccinini, (2009), p.2

³¹² Ibid. p.4

³¹³ Op cit., D. Dennett, (2003), p.23

³¹⁴ Op cit., G. Piccinini, (2009), p.4

need to harmonize and constrain them by building the appropriate links with third-person studies".³¹⁵ So all this rather counts against the conclusion that the knowledge involved here is private in the sense defined. And the outcome of holding (O), (R), (K) and $\neg(P)$, indeed, does threaten absurdity: These assumptions seem to entail private knowledge, but the received view seems to be that there is no such thing.

Furthermore, the Wittgenstein brigade might also take umbrage to the possibility of private knowledge. For example, if private knowledge is knowledge of p where p is a private proposition or the expression of a sentence of a private language, the kind of knowledge in question is out of the question because there can be no private propositions or expressions because there can be no private languages.³¹⁶

And now to make things even worse, doesn't the notion of a private language contradict (R)? It seems it is not possible to know what it's like to be another because there are no physical (neurophysiological, functional, behavioural) criteria to draw upon here, right? But (R) aims to say just what it takes for one conscious entity to be able to know another. The statement of the last criticism seems slightly off key. All that (R) does is give conditions that if satisfied render it *possible* for one conscious entity to know what it's like to be another. So it would be better to state this criticism in the following manner: (R) cannot be actualized for there is no logical or evidential manner in which it could be

³¹⁵ Op cit., Valera and Shear, (1999), p.2

³¹⁶ To be sure, Wittgenstein's original argument is obscure, as testified by the industry that grew up around its interpretation. The relevant passages in the original Wittgensteinian text are to be found in Ludwig Wittgenstein, (1963), *Philosophical Investigations*, (Tans.), G. E. M Anscombe, (Oxford: Blackwell), pp.243-351

satisfied. The relation is empty. But this runs directly against the neuroscience that I have cited above, makes the assumptions made at the existential, ethical and legal levels look unwarranted, too, and threatens outright scepticism. Indeed, this, too, lends weight to Dennett's charge of absurdity against dualism. So the threat is real. Is, then, dualism falsified?

The dualist seems to be in a pickle. However, I hold that one can accept (O), (R), (K) and $\neg(P)$, and take the view that, indeed, *there is no private knowledge*. The private knowledge defined above is only private *in the physical* sense in which it is defined. In another sense, such knowledge can be defined in a way that renders it *public*, but just *not* according to physicalist (neurophysiological, functional, behavioural) entailments, rather according to phenomenological criteria. Recall, p was defined as a matter of public knowledge by the following criteria: p is a matter of public knowledge when it is false that for some x , x knows p and necessarily if any y knows p , then $y = x$. This can be stated thus, p is a matter of public knowledge when for any x , if x knows p , then it is *possible* that for some y , y knows p and x and y are not identical. But it was said that it was *not possible* in the case of q^x , where this was read as 'what it feels like to be in pain *for* x '. The reason was twofold: First, a logical entailment was denied from the physical to the qualitative, by (K); second, an evidential entailment was denied from the physical to knowledge of the qualitative, by $\neg(P)$. However, taking the notion of empathy, which was discussed above, into account one can now expand on the conditions in question in order to allow q^x to be a matter of public knowledge, the definition reads, q^x is a matter of public knowledge when for any x , if x knows q^x , then it is possible that for some y , y knows q^x only if y knows q^y , where q^y is read as 'what it is like to be in pain *for* y ', even though x

and y are not identical.³¹⁷ That is, y can know what it is like for x to be in pain, even if it is not possible to know so physically (neurophysiologically, functionally, behaviourally), *through her own experiences of pain!* This gives one *phenomenological* criteria, distributed publically, for knowledge of other's qualitative states. It might even be said for some x such that x *has* pain but cannot propositionally know what it's like to be in pain, for example, if x lacks the cognitive component required for such knowledge, y can know what it's like for x to be in pain by the experience of the relevant type in herself.³¹⁸ The empathic or mirroring mechanisms of the body Gallese talks about could be used to understand such knowledge in a primarily non-inferential, non-propositional, naturalistic, automatic, implicit manner yet determined by natural evolutionary processes.³¹⁹ At least the conclusion drawn here needn't run counter to that line of thought. In another different sense, the conclusion even seems to fit the idea that a third-person may know more about the particular phenomenological state x is in than x does, as emphasized, for example, by Dennett.

So, it seems that the conjunction of (O), (R), (K) and $\neg(P)$ does not lead to private knowledge and no absurdity is to be had on *that* basis. Of course, this kind of knowledge is not public in the right way for all theorists. But those theorists obviously have a physicalist agenda to meet. There is no need for the dualist to acquiesce to that agenda if she has an argument that allows her to argue that phenomenological knowledge *is* public

³¹⁷ Restating, $(x)[xKq^x \rightarrow \diamond(\exists y)((yKq^x \rightarrow yKq^y) \ \& \ x \neq y)]$

³¹⁸ This might be true of our phenomenological relationship to animals, babies, etc.

³¹⁹ This last bit is purely speculative, so I won't pick it up and run with it. But it is an idea that might be explored at a later time.

in the right way: it meets intuitions, can be more familiar to a third-person, doesn't entail scepticism, or rub against the existential grain, and doesn't on the face of it entail a private language or representation system.

That brings us back to the original absurdity and the choice between (K) and (P). But, what if one could derive an absurdity from the conjunction of (O), (R), (P) and $\neg(K)$, then surely there would be even more justification to prefer (K) over (P). I think this is a possibility for this dualist. Here is the argument.

I start with an introduction to modal logic in order to define an epistemic logic. A generic modal logic is a proposition logic, PL, to which the following is added: the modal operator, \Box , a syntactic rule governing its syntactic relations, for example, 'if A is a sentence, then $\Box A$ is a sentence'. Further a set of axioms and rules are added to PL:

(Distribution) If $\Box(A \rightarrow B)$ is a theorem, then $\Box A \rightarrow \Box B$

(Necessitation) If A is theorem, then $\Box A$

And the rule of modus ponens.

This system is called *K*, after Saul Kripke, and is the basis of the different systems of modal logic. Epistemic logic, EL, is an extension of *K*. In such a logic, the modal operator is replaced by the knowledge operator, *K*. The operator is tagged to a knower using subscripts, for example, K_t , abbreviates "Tomoko knows that..." Validity in epistemic

logic is defined by providing a model $\langle W, R, V \rangle$, which is $\langle W, R \rangle$, known as a frame, and an evaluation V . W is a non-empty set of possible worlds ($w, v, u, w^1, v^1, u^1, w^2, \dots$). R is an accessibility relation, it connects possible worlds compatible with each other. V is a function that evaluates A at w (denoted $V(A, w) = 1$ or 0) for each sentence of EL in the following way:

$$(\neg) V(\neg A, w) = 1 \text{ iff } V(A, w) = 0$$

$$(\rightarrow) V(A \rightarrow B, w) = 1 \text{ iff } V(A, w) = 0 \text{ or } V(A, w) = 1$$

$$(K) V(KA, w) = 1 \text{ iff for each } v \text{ such that } wRv, V(A, v) = 1$$

An argument is valid just in case any model whose valuation assigns the premises 1 at a world also assigns 1 at the same world. A sentence is valid where an argument with no premises is valid. All such sentences are theorems. All theorems are valid. EL based on K is valid.³²⁰ This defines an epistemic modal logic.

The following is a valid process of reasoning in EL (‘/’ separates premise from conclusion):

1. A/B which says from A conclude B
2. $A \rightarrow B$ which is a theorem of EL given 1.

³²⁰ See M. J. Cresswell and G. E. Hughes, (1996), *A New Introduction to Modal Logic*, (London: Routledge)

3. $K_t(A \rightarrow B)$ by necessitation on 2.
4. $K_tA \rightarrow K_tB$ by distribution on 3.
5. K_tA by assumption
6. K_tB by modus ponens on 4. and 5.

This is usually thought to be problematic for EL because it attributes knowledge to someone, Tomoko in the above example, that she doesn't possess. It attributes knowledge of every consequent of an antecedent she knows. So, for example, Tomoko knows that $C/d = \pi$, but she doesn't know that $C/d = 22/7$. But the former entails the latter. So in EL Tomoko knows both. This is known as the problem of "logical omniscience".³²¹ However, if the person in question is an *ideal knower*, no such problem arises. The sense in which someone is an ideal knower can be defined and restricted. For example, C.D. Broad's mathematical archangel is an ideal knower relative to chemistry since he knows all the truths about chemistry,³²² Jackson's heroine Mary is an ideal knower with respect to physics since she knows all the physical truths, and God is an ideal knower with respect to everything. Turning back to Mary, Frank Jackson's argument can be restated in EL, given P abbreviates a list of all the physical truths including the a posteriori truths and nothing else but what is entailed by the physics in this world³²³, and Q abbreviates what it is like to see red, then the following reasoning is valid:

³²¹ See J. J. Meyer, (2001), "Epistemic Logic" in (Eds.) L. Goble, (2001), *The Blackwell Guide to Philosophical Logic*, (Oxford: Blackwell), pp.183-203

³²² C. D. Broad, (1925), *The Mind and Its Place in Nature*, (London: Routledge and Kegan Paul)

³²³ This follows more closely than above. See Op cit. Chalmers, (2004),

1. P/Q by $\neg(K)$
2. $P \rightarrow Q$ is a theorem of EL given 1.
3. $K_m(P \rightarrow Q)$ by necessitation on 2.
4. $K_mP \rightarrow K_mQ$ by distribution on 3
5. K_mP by assumption
6. K_mQ by modus ponens on 4. and 5.

It follows that so far as Mary knows all the physical truths, Mary knows what qualitative truths are entailed. That is, so far as Mary knows all the physical truths, Mary knows what it is like to see red. This is as (P) says. However, given Mary's situation is as described by Jackson³²⁴ it really doesn't seem that Mary could possibly know what it is like to see red. She wouldn't know, for example, what I or most others know about the qualitative aspect of redness, its phenomenological character. It, therefore, seems absurd to attribute such knowledge to Mary. That's just what the original conclusion is, right? The important question is what's the problem *here*? Steps 2, 3, 4, 6 are purely logical. There is no logical problem here, since Mary is an *ideal knower* with respect to all the physical truths including the a posteriori truths and nothing else but what is entailed by the physics in this world. The only assumptions are the materialist assumption at 1 and the assumption at 5. This latter assumption is valid because although it is probably not *practically possible*, it is *logically possible*. In fact the assumption is called for since Mary has been designated an *ideal knower*. The only assumption left to blame for the absurd conclusion is 1, the materialist assumption. The dualist can, now, argue that if (K) were dropped and

³²⁴ Op cit. Jackson, (1982)

(P) accepted, she would be committed not only to giving up dualism, but a philosophical position that led to absurdity. So the dualist, at least, can post a universal reason for not wanting to exchange (K) for (P).

Another reason for putting the argument in this way is to make it easy to see that if the materialist assumption is accepted, that is, if $\neg(K)$ is accepted, as it is at line 1, (P) is *logically* entailed, that is, as it is at line 4, without any fault of logic as Mary is an ideal knower. But this, in turn, leads to the absurd conclusion derived. Denying the conclusion and accepting the validity of 5 (which has to be assumed if Mary is taken to be an ideal knower), results in the denial of (P), which leads to the denial of $\neg(K)$. It follows that (K) is true and (P) is false. I think, then, that there are reasons to dismiss (O), (R), (P) and $\neg(K)$. I conclude that the dualist, therefore, has strong *logical* reasons to reject the aforementioned set of assumptions.

As with Chalmers and Jackson, I am assuming that if all the physical truths including the a posteriori truths and nothing else but what is entailed by the physics in this world are known, then all the facts about mental life should be deducible. A simple analogy is this: Logic tells us, for example, that if water is ____, then it is necessarily _____. Once we know the physical fact, that is, water is *in fact* H₂O, we can deduce that necessarily water is H₂O. Given other notions about the logical and empirical relations between X and Y, once we know the facts, analogously we can deduce further facts and relations, for example, that necessarily water is not XYZ. This kind of reasoning follows Kripke and parallels Jackson closely. However, as stated above, some might argue that $K_m(P \rightarrow Q)$ fails, that is because the knowledge in question is a posteriori and cannot be entailed by the physical

facts a priori despite all the physical facts available being available.

That's odd; however, let's assume so. There is *still* a whole lot of trouble for the physicalist! First, if there is a metaphysical relation that is of the type $\Box(P \rightarrow K)$, then there is no relation, no determination of mental reality by physical determinants, which is not *wholly* mysterious. It is not good enough to say "denying $K_m(P \rightarrow Q)$ just means that the *epistemological* entailment fails" because the knower in question can be considered to be an *ideal* knower with *all* the physical facts, emergent facts, and the relations between the two at her disposal. If she can't deduce mental reality from all those facts, *which are just all the facts for a physicalist*, then, neither can God! That is, the relation between mental reality and the physical facts remains *absolutely* mysterious. So if one still thinks that an epistemological gap is possible, even given that Mary is an ideal knower, yet maintain that a metaphysical gap is not possible, the relation that allows such a metaphysical reduction between the mental and the physical, involves a physically unknowable and, therefore, mysterious, reduction: Why do mental facts reduce to P? How do mental facts reduce to P? Why should I believe that mental facts reduce to P? What about the conservation laws, P gives us all the physical information, but mental facts cannot be deduced? What are mental facts? Is a physics that does not allow ideal Mary to deduce mental reality from the physical facts *really* a *complete* physics? Is there a mysterious *otherness* that relates the mental to the physical? God? Second, if $P \rightarrow Q$ does fail epistemologically, all the physical information does not allow a reconstruction of reality; no actual *physicist* can accept this, for all the physical information must allow one to reconstruct the full physical picture.³²⁵ Third, another problem is this, given a

³²⁵ Leonard Susskind, (2003), "Black Holes and the Information Paradox", in *Scientific American*,

somewhat mysterious reduction is allowed, non-reductive materialism is implied at least at the epistemic level. This threatens epiphenomenalism for only $P \rightarrow P^*$ relations are entailed by ideal knowledge of physics, that is, if ideal Mary cannot deduce mental reality from the physical facts, how will she explain how physical reality causes mental reality and, in turn, if the only causal relations are between physical properties (or events), what room will there be for mental causation at all? Last, scepticism ensues, for if *ideal Mary* can't even deduce the mental facts from all she knows, then how is anyone like me and you in a more prosaic state of epistemic being supposed to, that is, how are we supposed to know when others are, for example, *feeling* a pain? Mystery on mystery on mystery seems to be the outcome of a rejection of the argument posited. I, therefore, suggest that the dualist can stand her ground here and has little or no *reason* to acquiesce to the physicalist.³²⁶

So far, then, a number of absurdities and an alternative (which ended up looking absurd) have been avoided by the dualist. I'll turn next to other alternative views, perhaps, these are more attractive than dualism. I want to consider, next, Donald Davidson's and John Searle's positions. Davidson is an ontological monist and an epistemological dualist, that is, Davidson, given he advocates an anomalous monism, advocates anomalous epistemology stopping any conceptual reduction of psychology to physics, but

(Scientific American Inc.), Leonard Susskind, (2008), *The Black Hole War: My Battle with Stephen Hawking to Make the World safe for Quantum Mechanics*, (New York: Little, Brown), "Peter Bokulich, (2005), "Does Black Hole Complementarity Answer Hawking's Information Loss Paradox?" in *Philosophy of Science*, 72 (December 2005), 1336–1349

³²⁶ There may be other common objections raised. See *op cit.* Chalmers, (2004) for these and how they might be dealt with.

ontological monism, a token-token identity theory of mind. John Searle, I think, can be called an ontological dualist, and an epistemological monist. Apart from the passages already cited, here is another:

The mode of existence of conscious states is indeed ontologically subjective, but *ontological subjectivity of the subject matter does not preclude an epistemically objective science of that very subject matter.*³²⁷

What exactly does “ontological subjectivity” refer to? It refers to a subjective as opposed to objective ontological realm. What does epistemic objectivity entail? It implies something like (P), a neurophysiological determination of the mind:

Indeed, the whole science of neurology requires that we try to seek an epistemically objective scientific account of pains, anxieties, and other afflictions that patients suffer in order that we can treat them with medical techniques.³²⁸

So, the conclusion is that Searle is an ontological dualist and an epistemological monist. Do either of these positions threaten the kind of dualism argued for here? Perhaps. Davidson’s position amount to rejections of both (K) *and* (P) at different levels, and there acceptance at distinct levels. Searle’s position similarly rejects and accepts (K) and (P) at varying levels.

³²⁷ Op cit., John Searle, (2004), p.95

³²⁸ Ibid. p. 95

Let's take Davidson's position first. He can be understood as rejecting (K) and (P) at different levels, the former at the ontological level, the latter, at the epistemological level. His position rejects (K) because it is ontologically monist in character, so token-token identities are accepted and this seems to make (K) false. Further, the position rejects (P) at the epistemological level because it is anomalous in epistemological terms so far as the mind is concerned, leading to no conceptual reduction, as we have previously seen. However, given (K) does extend to token-token identities, Davidson's ontological monism can be said to be false. Of course, the original absurdity derived, that of the set of propositions (O), (R), (K) and (P), might be reason to accept Davidson's position. However, since, the dualist can claim that it is quite acceptable to give only (P) up, Davidson's position doesn't seem like the necessary outcome of the absurdity identified. On the other hand, Davidson's rejection of (P) doesn't extend to the ontological level, for there (K) is rejected. But this ends up making (P) looking mysterious. Furthermore, so far as Davidson rejects (P), epistemologically affirming anomalism, Davidson surely makes (R) redundant, and, in fact, so far as Davidson rejects (P) epistemologically and (R) is redundant, scepticism starts to loom. This, as charged above, runs directly against the neuroscience cited, makes the assumptions made at the everyday existential, ethical and legal levels look absurd, and as said threatens outright scepticism about other minds. There is, then, good reason to set Davidson's position aside for it isn't necessitated by the absurdity I pointed out above, and tends to mystery and scepticism (if not absurdity, as we shall see below).

Let me turn my attention to John Searle now. John Searle seems to have no truck with

conceptual dualism, he feels this lies at the heart of both property dualism and materialism.

For Searle,

[Conceptual dualism] consists in taking the dualistic concepts very seriously, that is, it consists in the view that in some important sense “physical” implies “nonmental” and “mental” implies “nonphysical”. Both traditional dualism and materialism presuppose dualism, so defined. I introduce this definition to make it clear why it seems to me best to think of materialism as really a form of dualism.³²⁹

However, Searle does say that the qualitative aspects of experience are ontologically reducible to the physics. He seems to suggest here that the conceptual difference would be lost if such a reduction took place:

*[I]n the case of consciousness we can make a causal reduction but we cannot make an ontological reduction without losing the point of having the concept.*³³⁰

Given what seems to be his antipathy to conceptual dualism, however, it is best to interpret this as his saying that the *pragmatics* of the qualitative aspects of experience would be lost if an ontological physical reduction was to take place. He stresses this when he stresses that our *definitional practices* make this the case:

³²⁹ Op cit. John Searle, (1994), p.26

³³⁰ Op cit. John Searle, (2004), p. 83

[P]retheoretically our notion of heat has something to do with perceived temperatures: Other things being equal, hot is what feels hot to us, cold is what feels cold. Similarly with colours: Red is what looks red to normal observers under normal conditions. But when we have a theory of what causes these and other phenomena, we discover that it is molecular movements causing sensations of heat and cold (as well as other phenomena such as increases in pressure), and light reflectances causing visual experience of sorts (as well as other phenomena such as movements of light meters). We then *redefine* heat and colour in terms of the underlying causes of both the subjective experiences and the other surface phenomena. And in the redefinition we eliminate any reference to the subjective appearances and other surface effects of the underlying causes.³³¹

He asks later,

Couldn't we say the same thing about consciousness? In the case of consciousness, we do have the distinction between the "physical" processes and the subjective "mental" experiences, so why can't consciousness be redefined in terms of the neurophysiological processes in the way that we redefined heat in terms of underlying physical processes?

His reply:

Well, of course, if we insisted on making the redefinition, we could....But of

³³¹ Op cit., John Searle, (1994), p. 119

course, the reduction of [for example] pain to its physical reality still leaves the subjective experience of pain unreduced, just as the reduction of heat left the subjective experience of heat unreduced....But where the phenomena that interest us most are the subjective experience themselves, there is no way to carve anything off.³³²

He summarizes the point thus:

[W]e can summarize this point by saying that consciousness is not reducible in the way that other phenomena are reducible, not because the pattern of facts in the real world involves anything special, but because the reduction of other phenomena depended in part on distinguishing between “objective physical reality”, on the one hand, and mere “subjective appearance”, on the other; and eliminating the appearance from the phenomena that have been reduced. But in the case of consciousness, its reality is the appearance; hence, the point of the reduction would be lost if we tried to carve off the appearance and simply defined consciousness in terms of the underlying physical reality.³³³

He concludes, that “this shows that the irreducibility of consciousness is a trivial consequence of the pragmatics of our definitional practices” and that this trivial result “has no deep lying metaphysical consequences.”³³⁴ Ultimately, then, *metaphysically*, he

³³² Ibid. p.121

³³³ Ibid. p.122

³³⁴ Ibid. p.122

holds that (K) is false. This is based on his belief that the qualitative aspects of experience can be *causally* reduced, and, therefrom, one *metaphysical* and *epistemological* picture of the world can be told. Here is what he says:

The fact that the causal powers of consciousness and the causal powers of its neuronal base are exactly the same shows that we are not talking about two independent things, consciousness and neuronal processes. If two things in the real empirical world have an independent existence they must have different causal powers. But the causal powers of consciousness are exactly the same as those of the neuronal substrate...Consciousness is thus an aspect of the brain, the aspect that consists of ontologically subjective experiences. But there are not two different metaphysical realms in your skull, one “physical”, and one “mental”.³³⁵

The causal reduction does not initiate an ontological reduction, for example, as with “solidity”, “liquidity”, etc. because “Consciousness has a first-person ontology, neuronal processes have a third-person ontology”. So a conceptual or rather definitional distinction would be lost.³³⁶ This is made all the more understandable since he accepts Kripke’s argument against the materialist, that is, he accepts (K)! For example, Searle talks about the point of Kripke’s argument as being “ludicrously simple and quite decisive”.³³⁷ He even, as I do, accepts the power of Kripke’s argument against the token-token identity

³³⁵ Op cit., John Searle, (2004), p.89

³³⁶ Ibid. p.89

³³⁷ Op cit. John Searle, (1994), p.118

theorist, i.e. Davidson.³³⁸

So Searle accepts a definitional distinction and an ontological distinction, but he does not accept a causal, epistemological and metaphysical distinction. This can be put like this, Searle accepts (K) at the definitional and ontological level, but rejects (K) at the causal, epistemological and metaphysical levels. Leading to the acceptance of (P) at the epistemological level, supported by the causal reduction.

First, so far as the dualist accepts Kripke's argument and thus (K), I think, she needs to accept a *metaphysical difference* between, for example, C-fibres firing and pain, because *logical* truths, including the modal truths that logic discloses, are *metaphysical* truths for Kripke! I've already quoted as much before, but I'll recite one of those quotes above to save the reader turning back to the first section: Alvin Plantinga, writing in the Kripkean tradition, says "“could have” express, broadly speaking, logical or metaphysical possibility.”³³⁹ I think, Searle would not accept this. Searle thinks that it is the description of a state of affairs that makes it logically possible or not, “Whether or not a state of affairs is logically possible depends on how it is described”.³⁴⁰ But, of course, an *essentialist* based dualism, based on Kripke's argument, is going to have no truck with that kind of idea. Indeed, (K) has its roots in an essentialist understanding of logical possibility, where what is necessarily possible, a genuine logical possibility, as I put it, is determined by the facts, regardless of knowledge or language (description)! Second, if the argument stated

³³⁸ Op cit. John Searle, (2004), p.69

³³⁹ Op cit. Quentin Smith, (1999), p.273

³⁴⁰ Op cit. John Searle, (2004), p.90

in terms of epistemic knowledge is correct and one takes the metaphysical distinction to be false, then $\neg(K)$ entails (P), but this leads to the absurd conclusion I derived above when restating Jackson's Mary argument, and all the problems for the physicalists in accepting it. Is that a conclusion that Searle would accept? He accepts Jackson's argument, so it doesn't seem like it would be a conclusion he can accept!³⁴¹ Third, if the meanings, or definitions, of the two kinds of terms in question are irreducibly different, then how come we aren't talking about two different kinds of properties or substances? Recall, the meanings are *so* different that causal reduction doesn't lead to an ontological reduction as it does in every other case, water, solidity, etc. Last, the meaning of "ontological", i.e., talk about what there is and the relations between what there is, suggests if the subjective and objective are ontologically distinct, then there are two different kinds of what there is. Dennett writes as most philosophers think: "The ontology of a theory is the catalogue of things and types of things that the theory *deems to exist*."³⁴² So if the qualitative aspects of experience are really *ontologically* irreducible to the physical aspects of reality, then Searle seems to be as committed to the distinct *existence* of the former as the latter. Searle's *deliberate* use of the "old vocabulary" in his "new way" does not help explicate his point.³⁴³ Well, that's plain enough from Searle's idea that materialism is a kind of dualism, right? For these reasons, I don't think that Searle gives us reason to reject the dualism posited on the conditions set out here. Thus far, then, dualism doesn't seem to have been falsified by absurdity nor stand in an inferior position to alternatives. Perhaps,

³⁴¹ Op cit., John Searle, (1994), p.117 and Op cit., John Searle, (2004), pp.66-68

³⁴² Op cit. Daniel Dennett, (1991b), p.33, my italics.

³⁴³ I quoted Searle at such great length for this very reason, to make sure the reader is aware that Searle's use of the relevant vocabulary is his and not necessarily mine.

there is one more way to derive the requisite absurdity that would falsify dualism, perhaps, focusing on mental causation will bring the dualist house of cards crashing in on itself.

So, lastly, I will explore the problems of mental causation and see if they pose a threat to dualism. As seen, causality, specifically causal reduction, has an important role to play in Searle's account of mind, indeed, causality is at the centre of his focus: "I believe...that understanding the nature of consciousness crucially requires understanding how brain processes cause and realize consciousness".³⁴⁴ In fact, Searle believes it is the chief problem faced by contemporary dualism. So, perhaps, pressure can be applied to the dualist in this way. Searle states: "The chief problem for the property dualists...is how can consciousness ever function causally?" He answers,

First, let us assume, as seems reasonable, that the physical universe is causally closed. It is closed in the sense that nothing outside it, nothing non-physical, could ever have causal effects inside the physical universe. If that is so, and consciousness is not a part of the physical universe, then it seems that it must be epiphenomenal. All of our conscious life plays no role whatever in any of our behaviour.³⁴⁵

This problem of causality in fact is important to many of the philosophers discussed above. For example, Davidson's conclusion that tokens of the mental must be identical to tokens

³⁴⁴ John Searle, (2002b), *Consciousness and Language*, (Cambridge: Cambridge University Press), p.58

³⁴⁵ John Searle (2002c), "Why I Am Not a Property Dualist", *The Journal of Consciousness Studies*, 9, No. 12, pp.57-64, p.59

of the physical starts from a belief in the causal relevance of the mental. Thus we get this in relation to Davidson first principle of his theory:³⁴⁶

The first principle asserts that at least some mental events interact causally with physical events. (We could call this the Principle of Causal Interaction.) Thus for example if someone sank the *Bismarck*, then various mental events such as perceivings, notings, calculations, judgements, decisions, intentional actions, and changes of belief played a causal role in the sinking of the *Bismarck*. In particular, I would urge that the fact that someone sank the *Bismarck* entails that he moved his body in a way that was caused by mental events of certain sorts, and that this bodily movement in turn caused the *Bismarck* to sink.³⁴⁷

Most philosophers want to avoid the “problem” of causal inefficacy. Dennett sums up the problem for dualism in physicalist manner thus:

Since we don’t have the faintest idea (yet) what properties minds stuff has, we can’t even guess (yet) how it might be affected by physical processes emanating somehow from the brain, so let’s ignore those upbound signals for the time being, and concentrate on the return signals, the directives from mind to brain. These, *ex hypothesis*, are not physical; they are not light waves or sound waves or cosmic rays or streams of subatomic particles. No physical energy or mass is

³⁴⁶ Of course, Davidson has been accused of epiphenomenalism in turn, as we shall see. For example, Ted Honderich, (1984), “Donald Davidson’s Anomalous Monism and the Champion of Mauve”, in *Analysis* 44, extracted <http://www.ucl.ac.uk/~uctyho/AnMon.htm>, 29/10/2013

³⁴⁷ Op cit. Donald Davidson, (2001, p.207

associated with them. How, then, do they get to make a difference to what happens in the brain cells they must affect, if the mind is to have any influence over the body?³⁴⁸

David Chalmers draws the conclusion that consciousness has little explanatory power. He calls this conclusion a paradox. And he seems to think that the dualist is committed to it. His argument is set in terms of phenomenal judgments. Here is what he wonders: “[H]ow can knowledge of consciousness be reconciled with the fact that consciousness is explanatorily irrelevant to phenomenal judgments.”³⁴⁹ The reason that Chalmers thinks consciousness is explanatorily irrelevant, which he deems a paradox, is this:

The paradox [of explanatory irrelevance] is a consequence of the facts that (1) the physical domain is causally closed; (2) judgments about consciousness are logically supervenient on the physical; (3) consciousness is not logically supervenient on the physical; and (4) we know we are conscious. From premises (1) and (2) it follows that judgments about consciousness can be reductively explained. In combination with premise (3), this implies that consciousness is explanatorily irrelevant to our judgments, which lies in tension with (4). Thus, we have the paradox.³⁵⁰

Well, the threat to causal efficacy is clearly suggested: if consciousness has some casual

³⁴⁸ Op cit. Daniel Dennett, (1991b), pp.34-35

³⁴⁹ Op cit. David Chalmers, (1997), p.183

³⁵⁰ Ibid. p.183

role to play in forming our phenomenal judgements, as we might intuitively think, then it can't be totally irrelevant to an explanation of phenomenal judgement. However, it seems to be that consciousness is totally irrelevant to an explanation of phenomenal judgement, which suggests consciousness has no causal role to play in the formation of phenomenological judgments. That's intolerable for Elitzur who plumps for an interactionist dualism against all his scientific inclinations.³⁵¹

However, Searle refers to moving from epiphenomenalism to interactionism as jumping out of the frying pan and into the fire:

[W]e may assume that the physical universe is not causally closed, that consciousness can function causally in the production of physical behaviour. But this seems to lead us out of the frying pan and into the fire, because we know, for example, that when I raise my arm, there is a story to be told at the level of neuron firings, neurotransmitters and muscle contractions that is entirely sufficient to account for the movement of my arm. So if we are to suppose that consciousness also functions in the movement of my arm, then it looks like we have two distinct causal stories, neither reducible to the other; and to put the matter very briefly, my bodily movements have too many causes. We have causal overdetermination.

Dennett perceives the absurdity of the interactionist position from his standard

³⁵¹ Chalmers has indicated that he thinks Elitzur's position is not compelling. See Chalmers, (1997), p.183 and p.163

perspective:

A fundamental principle of physics is that any change in the trajectory of any physical entity is an acceleration requiring the expenditure of energy, and where is this energy to come from? It is this principle of the conservation of energy that accounts for the physical impossibility of “perpetual motion machines”, and the same principle is apparently violated by [interactionist] dualism.³⁵²

Elitzur admits that if the causal efficacy of qualia are systematic, “then qualia must be using energy in order to interfere with the brain’s random processes in a nonrandom manner, again violating the first and/or the second laws of thermodynamics”.³⁵³ Nevertheless, he accepts interactionism.

So there is a choice between inefficacy, on the one hand, or efficacy, on the other hand, but both are highly problematic, perhaps, even ludicrous. I will argue that the kind of dualism invested in here may be able to defend itself against accusations of absurdity from mental causation on the following grounds: if A or B is a choice and A is false and B is false, then the choice is a false choice. *What we’ve seen above already suggests as much.* However, I will consider, first, two specific dualist perspectives, one that threatens causal epiphenomenalism (Chalmers) and one that is interactionist in nature (Elitzur) in order to show that they both fail to convince. I will take Chalmers first and show why, I think, he is wrong in holding the premises he does. I will also suggest a Kripkean way to

³⁵² Op cit. Daniel Dennett, (1991b), p.35

³⁵³ Op cit. Avshalom Elitzur, (2009), p.12

show that a psychological confusion is involved. His conclusion seems to be false, then, and there is reason to think that a Kripkean style psychological confusion is involved. I will consider Elitzur next and show why, I think, he is wrong in thinking dualism, follows from his discussions. Last, I'll suggest, given the problem of mental causation and the false choices it presents to philosophers can be generalized to all positions in the philosophy of mind, dualism has nothing to fear from accusations of absurdity from mental causality and, I think, can even dismiss the problem outright.

Let me turn to Chalmers first: Chalmers first premise is causal closure of the physical realm. As a dualist one can either accept this assumption or reject it. Acceptance tends to epiphenomenalism and rejection tends to interactionism. Chalmers second premise implies (P) as stated *for judgments*, as he says, "We have *defined* judgments so that they are functional states".³⁵⁴ So, I think, something like the following can be formulated, where F abbreviates some set of functional truths and J stands for judgments about consciousness:

$(P)_J F \rightarrow J$

$(P)_J$ is very much like (P) above. F is taken to entail J. Given this, I think, the following can be derived:

$(P)_J^+$ If x and y are functionally identical and x judges p to be the case, then y judges p to be the case.

³⁵⁴ Op cit. David Chalmers, (1997), p.183, see also p.174

It's clear that Chalmers thinks this is a consequence of his definition given, as witnessed by the following remark: "As I am using the term, I think, it is natural to say that my zombie twin *judges* that he has conscious experience, and that his judgments in this vicinity correspond one-to-one with mine".³⁵⁵ And that is precisely because Chalmers thinks he and his *zombie* twin are functionally equivalent and specifically, here, with respect to judgement.

Chalmers' third premise says that P does not logically entail Q, that is, $\neg\Box(P \rightarrow Q)$, where P abbreviates physical truths, and Q a qualitative truth, whereof, it is easy to derive (K), that is, $\Diamond(P \ \& \ \neg Q)$; here, (K)^F is most relevant. And from this the possibility of a Nagel/Zagel pair follows as stated at the beginning of this long section.³⁵⁶ Chalmers obviously accepts this else he wouldn't reference his *zombie twin* so much.

Chalmers last holds we know what it's like to be conscious. In the terms I have been using, this would be, I think, reflected in the acceptance of the assumption (O) *and* (R), for (R) gives conditions which are as necessary to self-knowledge of ones phenomenal status (xRx) as they are to knowledge of another's phenomenal status (xRy).

The paradox can, then, with reference to my template be brought out in the following manner: (P)_J allows one to draw the conclusion that if x and y are functionally identical

³⁵⁵ Ibid. p. 174

³⁵⁶ Formally, $\Diamond(\exists x)(\exists y)(x \text{ }_{p=} \text{ } y \ \& \ \neg x \text{ }_{q=} \text{ } y)$, where $_{p=}$ stands for 'physically identical' and $_{q=}$ stands for 'experientially identical'.

and x judges p, then y judges p. If (K)^F, it is possible that Thomas Nagel is functionally identical to an entity I have labelled Zagel. It is also possible for Thomas Nagel to be functionally identical to Zagel and make a judgement about p, where p stands for any truth about the qualitative aspects of the experiences he enjoys and suffers, for example about the qualitative aspects of tasting marmite. It follows that it is possible for Zagel to make a judgement about the qualitative aspects of tasting marmite. But this is in conflict with Nagel *knowing* what it's like to taste marmite and Zagel *not being able to know* what it's like to taste marmite. That is, Zagel can be understood as making the same *phenomenal judgments* as Nagel does, but without the possibility of knowing anything about the phenomenal aspects of the judgements he makes. That is, Zagel can *judge* what it's like to taste marmite, but Zagel cannot *know* what it's like to taste marmite.

Chalmers wants to stress the explanatory irrelevance of consciousness to judgements and this is the sense of the paradox he wants to bring out. That, too, can be given sense in the terms I have set out his argument. The physical world is causally closed, meaning all physical explanandum have a physical explanans. Judgements, phenomenal or otherwise, are determined physically, hence (P)_J. So any judgment that is in need of explanation gets a physical explanans. On the other hand, consciousness is not entailed by physics, hence (K). And, by the causal closure of the physical world, consciousness has no role to play as a physical explanation. So a judgement, even a phenomenal judgement, for example, regarding the exquisite taste of marmite, gets a physical explanation. If (K), an entity like Zagel is possible. Such an entity is a purely physical entity. So that everything about Zagel can be explained physically, including his judgements, including his phenomenal judgements, and including those, for example, about the distinctive qualitative aspects

associated with the taste of marmite. Nothing wrong so far, right? The problem is that by (K) Zagel is physically identified with Thomas Nagel. That means, whatever explains the judgements that Zagel makes, explain the judgements that Thomas Nagel makes. The difference is that for Thomas Nagel *there is something it is like to taste marmite* and *he can know what it's like to taste it* (simply enough by tasting it). But none of that seems to matter to explaining the judgements that he makes about what it's like to taste said substance. And, it'll seem to many, if consciousness and knowing what it's like to be conscious are irrelevant to the explanation of the judgements one makes about conscious matters, perhaps, consciousness and knowing what it's like to be conscious are causally inefficacious, too. This is Chalmers' Paradox.

Paradoxes, of course, are not fatal like out and out absurdities, but they are troubling. For my part, I contend that Chalmers' conclusion entails absurdity, that is, because (R) can be reformulated relative to judgment, that is, if the question of what it takes for one entity to be able to *judge what it's like to be another entity* is taken into consideration, absurdity looms. Before the reformulation I want to remind the reader that the expression 'what it's like to be' is synonymous with 'what it *feels* like to be' in all its *qualitative* glory. With this in mind, the reformulation of (R) for judgement, I think, brings out the absurdity underlying Chalmers paradox. Let's ask what does it take for one entity to be able to *judge* what it's like to be another entity? The answer is this, I think:

(R)_J x can judge what it's like to be y iff:

(a) x and y are conscious,

- (b) x and y experience the world in the same way with respect to the qualitative aspects associated with the set of experiences that x and y have, and
- (c) x has the cognitive abilities associated with judgement.

The first condition requires that there is something it is like to be both x and y; the second condition requires that the experiences that x and y enjoy and suffer are qualitatively the same; the third requires cognitive competence of x so far as the ability to judge is concerned. The relation is reflexive. That is, x can judge what it's like to be x. One can conclude, Thomas Nagel can judge what it's like to be Thomas Nagel, David Chalmers, etc. However, Thomas Nagel cannot judge what it's like to be a rock, what it's like to be Sherlock Holmes, what it is like to be Zagel, assuming that (K) is true, since each is not conscious. Likewise none of this motely-crew can judge what it's like to be Thomas Nagel, for the same reason. Pointedly, Zagel cannot make judgements about what it's like to be Thomas Nagel. The absurdity should already be apparent.

Now, from (P)_J it follows that if x and y are functionally identical and x judges p, then y judges p and from (K)^F, the possibility of Thomas Nagel being functionally identical to Zagel and making a judgement about p, where p stands for any truth about the qualitative aspects of the experiences he enjoys and suffers, for example, about what it's like to be Thomas Nagel, one can conclude that it is possible for Zagel to make a judgement about what it's like to be Thomas Nagel . On the other hand, as concluded in the last paragraph, by (R)_J, assuming (K)^F, Zagel cannot make judgements about what it's like to be Thomas Nagel. It follows that it is possible for Zagel to make judgements about what it's like to

be Thomas Nagel and it is not possible for Zagel to make judgements about what it's like to be Thomas Nagel. The *absurdity* is clear here, there is no mere paradox! What to do? Again, for the same reasons as above, I think, dualists should retain (O), as should any serious philosopher of mind. So, too, (R)_J, for exactly the same reasons as above relative to the original (R) relation, to abate anti-scepticism, to stay in step with our existential, ethical and legal assumptions, and so as not to be undermined by the neurosciences. That leaves (K)^F and (P)_J. And, again, given a dualist stand, it seems obvious that (K)^F should be retained at the expense of (P)_J. Again, if a private sense of some set of judgments follows, that sense of judgment can be explained in the same way as the sense of private knowledge was explained above, that is, such judgments will be private in the physical sense but public in the phenomenal sense. For example, for any x, if x judges what it's like to be in pain for x, then it's possible for some y to *judge* what it's like to be in pain for x only if y judges what it's like to be in pain for y, even when y is not identical to x. In relation to Chalmers argument above what this tells us is that Chalmers second assumption is false. If Chalmers is committed to the second premise, so be it, but I think so much the worse for him as retaining (P)_J (along with (O) and (K), the former is necessary to dualism and the latter is sufficient) means giving up (R)_J and opening up the sluice gate to scepticism, counter-intuitiveness (existentially, ethically, and legally) and runs contra to the sciences and phenomenology that tells us that we can know and judge another's phenomenological states by having the capacity to know, and judge our own. Plus, of course, the imminent threat to mental causation that explanatory epiphenomenalism presents. I think, dropping (P)_J, in the way that I have allowed for (or in some other way) avoids all this. Well, the main point is, at the least, one isn't committed to Chalmers' hypothesis as a dualist. Adopting the argument presented here allows one to

side-step Chalmers argument for the lack of explanatory powers of consciousness, which seems to bear so heavily on the causal efficacy of the mental.

Well, Chalmers has written that if the second premise is denied in terms of judgements, one can talk about *claims* instead:

Some might be tempted to deny premise (2), but recall that we have *defined* judgments so that they are functional states, logically supervenient on the physical. Now some might argue that there is no such functional state that remotely resembles what we think of as a judgment; but even so, we can simply retreat to *claims* about consciousness.³⁵⁷

I won't talk about "claims", instead I'll work with the term "talk about" or use "say". This shouldn't matter to the point that Chalmers is trying to make for "talk about" and "say" suffice to bring out the paradoxical nature that Chalmers is trying to bring out and refer to behaviours as he thinks the word "claim" does: "*claims* about consciousness...are behavioural acts".³⁵⁸ An outcome of the paradox in the linguistic context is that: "consciousness is explanatorily irrelevant to the sounds we produce, and to the marks we write, all of which can be systematically interpreted as concerning consciousness"³⁵⁹. I'll reconstruct the paradox in the following form. I'll define a (P) like relation and then make the paradoxical contrast in pretty much the same way as I did above. Here goes:

³⁵⁷ Ibid. p.183

³⁵⁸ Ibid. pp.184-85

³⁵⁹ Ibid. p.184

(P)_B P → L

This is where P stands for some set of physical truths and L stands for a linguistic behaviour. P is taken to entail L. Given this, I think, the following can be derived following the reasoning given earlier:

(P)_B⁺ If x and y are physically identical and x talks about p, then y talks about p.

Alternatively,

(P)_B⁺ If x and y are physically identical and x says p, then y says p.

Here comes the paradox: (P)_B allows one to draw the conclusion that if x and y are physically identical and x talks about p, then y talks about p. If (K), it is possible that Thomas Nagel is physically identical to an entity, Zagel, and possible for Thomas Nagel to be physically identical to Zagel and talk about p, where p stands for any truth about the qualitative aspects of the experiences Thomas Nagel enjoys and suffers, for example, the ticklish feeling one gets in one's nose just before one is about to sneeze. It follows that it is possible for Zagel to talk about that ticklish sensation, too. But this is in conflict with Nagel being able to *know and judge* what it's like to feel it and Zagel not being able to know what it's like to feel that kind of tickle. That is, Zagel can be understood as talking about the same thing as Nagel does, but without the possibility of knowing or judging anything about it. That is, Zagel can *talk about* (even describe) what it's like to feel a

tickle before he sneezes, but Zagel cannot *know or judge* what it's like to feel such a tickle. Or as Chalmers wants to bring out the point, the phenomenal aspects of experience, the ticklish feeling one gets in one's nose just before one is about to sneeze, is irrelevant to the explanation of Zagel's linguistic behaviour when he talks about such tickles because Zagel doesn't enjoy or suffer these elements of experience. *Ex hypothesis*, since Thomas Nagel and Zagel are physically identical, they are just as irrelevant to an explanation of Thomas Nagel's linguistic behaviour because the same *physical* explanation given for Zagel's behaviour will suffice to explain Nagel's linguistic behaviour. Threatening, also, as I am interpreting things, the causal efficacy of conscious phenomenal states to linguistic behaviours or language. However, again, I think, if it is possible for one to formulate an R type relation with respect to talking about or saying, then an absurdity will ensue and one will be done with both explanatory inefficacy and the threat of causal inefficacy. Here is a possible relation:

(R)_B x can talk about (or say) what it's like to be y iff:

- (a) x and y are conscious,
- (b) x and y experience the world in the same way with respect to the qualitative aspects associated with the set of experiences that x and y have, and
- (c) x has the abilities associated with language.

From (K) and (P)_B one will be able to show that if it's possible for Thomas Nagel to talk about what it's like to be Thomas Nagel and to be physically identical to an entity we are calling Zagel, then it is possible for Zagel to talk about what it's like to be Thomas Nagel.

But from (K) and (R)_B, it follows that it's not possible for Zagel to talk about what it's like to be Thomas Nagel, because Zagel is not conscious in the same way since he is not conscious at all. And herein lies the absurdity. The important question, here, is can (R)_B be given up at the expense of (P)_B? The analogy I wish to pursue is that of talk about, for example, water that isn't H₂O. Talk about this stuff, as Kripke says, would not be talk about water at all, but talk about 'fool's water', or a qualitative analogue of water (and however much such talk didn't diverge from talk about actual water). Similarly, talk about *what it's like to sneeze*, the qualitative aspect, devoid of any reference to *the qualitative aspect associated with the experience of sneezing* is not talk about *what it's like to sneeze* but talk about something essentially distinct, a 'fools experience', if you will, or some kind of mere *physical analogue*, an amalgam of the physical aspects associated with the act of sneezing, and why not? After all aren't we talking of a physical analogue of Thomas Nagel?

Let me say a bit more about this: Recall, according to Kripke the facts constrain modality and the imagination. Prior to discovering the facts, though, we can know a priori that if A, then necessarily A (where A is statement like $a = b$). On discovering that A is *in fact* true, we will have discovered that A is necessarily true. But even if we don't ever discover the fact, if *in fact* A is true, then necessarily A is true *regardless*. The fact one doesn't know that A or necessarily A is *just a mark of ignorance* or confusion. In this case, if A is true, but we don't come to discover the fact *and* even come to the erroneous conclusion that $\neg A$, when we make such statements or think we imagine such a state of affairs, in the best case scenario, we will really just be talking about a qualitative analogue or imagining a qualitative analogue of a (or b) in $a = b$, something that is *essentially* distinct

from *a* (or *b*), but resembles it in all its qualitative glory, as I have discussed above. In such a case, one isn't just ignorant, one is confused. This thinking applies to the case where we let A abbreviate a statement like the following: 'what it's like to have a migraine is what it feels like to have such a headache'. We know *a priori* that if there are such things as migraines, then necessarily what it's like to have a migraine is what it's like to feel a migraine in all its qualitative menace. That is true, even if one has never had a migraine, or if no one has ever had a migraine. Again, we might come to an erroneous conclusion that what it feels like to have a migraine is identical to all the physical aspects associated with having a migraine. (This isn't so difficult to imagine because we are all behaviourists when it comes to the third-person.) We might even assert so or imagine this to be at least possible, that is, as many a materialist does, associate what it feels like to have a migraine with a list of behaviours and functions less the meaning and debilitating sensations associated with the violence of this pain. On the Kripkean account I want to press, since this is available to the dualist who accepts his arguments for dualism, we are not really talking about or imagining what it's like to have a migraine at all, rather we are talking about and imagining merely the physical phenomena associated with the experience of having a migraine. That's quite consistent with talking about or imagining something entirely distinct. We might call this a "quantitative analogue". I think, Zagel, at best, can only talk about a quantitative analogue when he talks about "what it's like for Thomas Nagel to have a migraine" because the term as used by him cannot make reference to the qualitative aspects associated with a migraine since, *by definition*, he has no access to the referent in question. At worst, Zagel use of the term refers to nothing. Zagel, therefore, can't talk about an actual migraine, but can only talk about a quantitative analogue of an actual migraine, which is essentially distinct. Therefore, Zagel can't really

talk about what it is like to be Thomas Nagel, for example, with a migraine. Although, he can *seem* to talk about the matter. I think, (R)_B can be defended in this manner.

The thoughts of the last paragraph dove-tail with another argument that I think holds good against Chalmers' Paradox. This time I try to explain, using the Kripkean template, how one might fall foul of the mental rain of the imagination, this time with respect to a quantitative analogue, and come to think that explanatory inefficacy follows. The picture theory posted earlier explains how such a confusion may arise. According to the picture theory, the imagination is pictorial. Pictures refer by representing and representing-as. Representation is constrained by the facts. If P represents x and $x = y$, then P represents y. Representation-as is constrained by the facts, too. If P represents x as y and $y = z$, then P represents x as z. However, it is possible for a picture representing one thing to represent that one thing as something completely distinct. For example, Charles Darwin may be represented as a monkey. What confuses most is when one thing, x, is represented as an essentially distinct thing, not as x, yet exemplifying all the typical descriptions that go along with x, so as to appear extremely x-like. The archetypical example is water, and the example was worked through above. When one imagines 'waterless water' what one has in mind is a picture that represents water in all the ways it tantalizes the senses, less its essential characteristic, H₂O, that is, one *represents* water *as* water-like, exemplifying all the descriptions that describe the way it invades our senses, but as something essentially distinct. Basically, representing water as a qualitative analogue of actual water. Now, couldn't the same thing be happening here? When one thinks, even convinces oneself, one can imagine what it's like to have a migraine, less the menace of the felt headache and associated symptoms (nausea, smells, etc.), and perhaps, even with a set of other feels,

one is really just *representing* a migraine *as* something that exemplifies all the descriptions associated with the physical attributes ordinarily associated with the menace of this debilitating condition, but *as* something which is essentially distinct. What one has in mind is a picture that represents a migraine in all the ways its menace and discomfort is physically manifest, less its essential characteristic (*its feel*) that is, one *represents* the migraine *as* quantitatively migraine-like, exemplifying all the physical descriptions of the ways it presents its nuisances, but as essentially distinct. It can be argued that something similar is going on here. Zagel can't know, judge or talk about conscious states because what he talks about are not conscious states but their physical analogues. What makes Chalmers think he can is just the corresponding set of quantitative analogues. It is no surprise to be confused by the resemblance between Zagel and Thomas Nagel and their intentional states such as judgements, and behaviours, etc. because they resemble each other to such a high degree *being physically identical*, but they are nevertheless experientially distinct. Also, it is how we usually identify such knowledge, judgement and talk about phenomenology *in the third-person*. But, I think, we shouldn't let these superficial qualities fool us into thinking that Zagel and Nagel share knowledge, judgement and talk any more than we let, for example, the qualitative aspects of heat impair our knowledge, judgement and talk about heat. I think, then, there are reasons to cast doubt on Chalmers' conclusions and the dualist can appropriate this argumentation if she wants to avoid explanatory epiphenomenalism and its imminent threat to mental causation.

Here is a brief worry about the analysis presented. Chalmers says that the judgements about consciousness are logically supervenient on the physical, but consciousness, is not

logically supervenient on the physical. I took this to exemplify a (P) relation $P \rightarrow J$ and a (K) relationship $\diamond(P \ \& \ \neg Q)$.³⁶⁰ Is this a legitimate interpretation of Chalmers? To answer this as briefly as possible, I'll say a bit about supervenience. Supervenience is a dependency relation: so if B supervenes on A, B depends on A. But it is not just a dependency relation: "First and foremost, the supervenience relation is to be a relation of dependence or *determination*."³⁶¹ Terence Horgan traces supervenience back to G. E. Moore's notion of "intrinsic value".³⁶² That dependency bound tight with determination is clearly expressed in the precursor of supervenience: "It is not *possible* that of two exactly similar things one should possess [intrinsic value] and the other not, or that one should possess it in one degree, and the other in a different one."³⁶³ Davidson, who introduced the relation into the philosophy of mind expresses its strength like this:

[S]upervenience might be taken to mean that there cannot be two events exactly alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect.³⁶⁴

³⁶⁰ The usual interpretations are legitimate, here, too.

³⁶¹ Lynne Rudder Baker, (1993), "Metaphysics and Mental Causation", in (Eds.) John Heil and Alfred Mele, (1993), *Mental Causation*, (Oxford: Clarendon), p.79, my italics.

³⁶² See Terence Horgan, (1993), "From Supervenience to Superdupervenience" in Op cit., Chalmers (2002), pp.150-162, and G. E. Moore, (1922), "The Conception of Intrinsic Value", <http://www.ditext.com/moore/intrinsic.html> extracted 01/11/2013

³⁶³ Ibid. p.5

³⁶⁴ Op cit. Donald Davidson,(1970), p.213

Today the often cited kinds of supervenience are “global” and “strong”. Global supervenience is defined by Terence Horgan in the following way:

There are no two physically possible worlds that are exactly alike in all physical respects but different in some other respects.³⁶⁵

Lynne Rudder Baker defines it like this:

Two possible worlds that are indiscernible in all micro-physical respects are indiscernible in all respects.³⁶⁶

The kind of supervenience that Chalmers is discussing is logical, his definition goes like this:

If B-properties are logically supervenient on A-properties, then there is a sense in which once the A facts are given, the B-facts are a free lunch...The B-facts merely redescribe what is described by the A-facts. They may be different facts (a fact about elephants is not a microphysical fact), but they are not *further* facts.³⁶⁷

Later:

³⁶⁵ Op cit. Terrence Horgan, (1993), p.155

³⁶⁶ Op cit. Lynne Rudder Baker, (1993), p.79

³⁶⁷ Op cit. David Chalmers, (1997), p.41

[M]aterialism is true if for *any* logically possible world *W* that is physically indiscernible from our world, all the positive facts true of our world are true of *W*.³⁶⁸

Lynne Rudder Baker says, quoted below, that strong supervenience is required for a materialist determination of mind.³⁶⁹ Strong supervenience includes necessity. We have already seen this kind of formulation in Stoljar and Nagasawa.³⁷⁰ I think that (P) and (P)_J, as defined above, capture the essential ideas here that judgement is logically and strongly supervenient on the physical, for if judgement supervenes on the physical, then the physical determines judgement, (P)_J $P \rightarrow J$, and if judgement supervenes on the physical in the a strong logical sense, where that means that *P entails J*, then it follows that (P)_J can be understood as $\Box(P \rightarrow J)$. That is, in *any* possible world, if *P*, then *J*.³⁷¹ In fact, all the (P) relations can be taken as supervenience relations in this way if one wishes to. And they can be taken in the strongest supervenient terms such that materialism about knowledge, judgement, sayings, etc. are determined physically and necessarily, for example, (P) can be read as $\Box(P \rightarrow K)$. It should be clear that (P) relations can be so formulated because those relations were talked of as *entailment* relations, i.e. *P* was said to entail *K*, *J*, *L*, etc. with respect to the particular (P) relation under discussion. Furthermore, that (K), $\Diamond(P \ \& \ \neg Q)$, clearly implies the lack of supervenience so far as it is

³⁶⁸ Ibid. p.42, my italics

³⁶⁹ Op cit. Lynne Rudder Baker, (1993), p.80

³⁷⁰ See Section C above.

³⁷¹ Op cit. David Chalmers, (1997), p.41

expressed in the strongest terms thus, $\Box(P \rightarrow Q)$. I conclude that stating things in the way I have does not misrepresent Chalmers claim that consciousness is not supervenient on the physical, but that judgement is.

I wanted to avoid using the term “supervenience” because it is so metaphysically loaded. Terence Horgan highlights that it has sometimes looked even more mysterious than the mystery it is meant to explain.³⁷² Furthermore, Lynne Rudder Baker writes that strong supervenience “is a metaphysical, not an epistemological thesis; we may never discover the relevant micro-physical properties on which higher-level properties supervene. So, [it] does not imply reducibility in any epistemological sense”.³⁷³ Of course, as seen in chapter one, Chalmers would seem to disagree with this. My point here, though, is to suggest that talk of supervenience is just mysterious, and that this is why I have eschewed talk of it thus far.

To return to Chalmers Paradox, whichever way the argument is stated, my way or with an explicit focus on supervenience, Chalmers paradox will fail due to the clash between the (P) relation (or related supervenience relation), and the Nagelian (R) relation, I think, the dualist can and should accept. But am I not saying, then, that phenomenon cause phenomenal knowledge, judgement (and other intentionality)? Not necessarily, an argument has only been made that halts Chalmers’ epiphenomenal conclusions, an argument hasn’t been made to establish interactionism.

³⁷² Op cit. Terence Horgan, (1993), p. 154, Cf. Stephen Schiffer, (1987), *Remnants of Meaning*, (Cambridge, Massachusetts: The MIT Press)

³⁷³ Op cit. Lynne Rudder Baker, (1993), p.81

I'm going to turn to Elitzur next, who does present an argument for interactionist dualism. An introduction was already provided in the first section of chapter one, but I'll go over the argument again here. Elitzur seems to think that physics *can* tell a pair like Thomas Nagel and Zagel apart. For example, Thomas Nagel is baffled about the relation between physics and the irreducible aspects of subjective experience, likewise Zagel. But what baffles Thomas Nagel are those very irreducible aspects of subjective experience. On the other hand, what baffles Zagel cannot be these aspects of mental life. If it isn't these (he has none) it must be some other aspect. This other aspect will be physical. Thus, Nagel is baffled by the qualitative aspects associated with experience and Zagel is baffled by some physical aspects associated with his experience. And, therefore, Nagel's phenomenology acts on him and causes his bafflement and Zagel's physics acts on him and causes his bafflement, each investigating the cause of their bafflement will give different answers to questions asked about the cause of their bafflement. For example, Thomas Nagel will be able to imagine a being physically identical to him but lacking the qualitative aspects of consciousness that baffle him so, but Zagel will not be able to imagine a being physically identical to him but lacking the physically explained percepts that explain Zagel's bafflement, because this is a physicalist explanation of the cause of such bafflement. However, I think, in effect, this denies the possibility of a Zagel like character, because the difference between Thomas Nagel and Zagel will be manifest physically, through the knowledge, judgement and language used to discuss their bafflement, which means Thomas Nagel and Zagel will *not* be physically identical. They physically diverge because the Zagel character has a physical difference explaining his bafflement that

Thomas Nagel doesn't have³⁷⁴. This leads to the denial of (K). And if (K) is denied, then what sufficient reason does one have to hold to dualism. It seems assumed rather than argued for. Here is why, I think, (K) fails. To hold (K) is to hold $\diamond(P \ \& \ \neg Q)$, where P abbreviates a statement of physical truth and Q a statement about the presence of an associated qualitative fact. To hold $\diamond(P \ \& \ \neg Q)$ is to hold $\diamond(\exists x)(\exists y)(x \text{ } _p= \text{ } y \ \& \ \neg x \text{ } _q= \text{ } y)$, where $_p=$ abbreviates 'is physically identical to' and $_q=$ abbreviates 'is experientially identical to' But according to Elitzur a Zagel type character would be physically distinct from Thomas Nagel. That is, if truly Zagel and Thomas Nagel were physically identical, they would both enjoy and suffer the qualitative aspects of experience. But that gives us $\neg\diamond(\exists x)(\exists y)(x \text{ } _p= \text{ } y \ \& \ \neg x \text{ } _q= \text{ } y)$ or $\Box(x)(y)(x \text{ } _p= \text{ } y \ \rightarrow \ x \text{ } _q= \text{ } y)$. This implies, $\Box(P \ \rightarrow \ Q)$. That is, the denial of (K)!

So the only way to accept dualism is to assume it. Well, Elitzur argues that physics cannot reduce qualia, but that, of course, is not an argument for dualism, since many materialists argue much the same. Another way to say all this is that Elitzur is left with (O) and (R) (this latter I assume on the principle of charity since it can do him no harm to hold this and, I think, actually does him some good), and probably (P), given the conclusion $\Box(P \ \rightarrow \ Q)$, but, unfortunately, $\neg(K)$. That *isn't* sufficient for dualism; though, it is sufficient for some kinds of materialism, non-reductive materialism, for one. It also, unfortunately, may imply absurdity.

³⁷⁴ Note the difference with my argument, I argue that the presence and absence of consciousness distinguish the parties, their knowledge, judgements, etc. and so bafflement, too. Elitzur seems to be arguing physics will distinguish them.

To sum up, the arguments presented here that tend towards epiphenomenalism or interactionism are flawed. Therefore, so far as we have looked in to the matter, there are no good reasons to choose either position. And in fact there are good reasons to reject the choice all together, for if one is faced with a choice A or B, and A is false and B is false, then the disjunction is itself false. This is the stance I wish to generalize below. There are reasons to believe that mental causation in general is a false problem and the choices it offers false choices. Whereas, Chalmers writes: “[T]he question of whether consciousness is *causally irrelevant* in the production of behaviour is a complex metaphysical issue that is best left open”,³⁷⁵ I think, there is reason to take a sceptical stance towards the issue. Indeed, there are ways to avoid the metaphysical grip of mental causation altogether. For example, one way to escape the problem of mental causality is via Hume. His ideas on causation are well known, summed up in the following way by Yablo:

Causes are not strictly necessary for their effects, because the latter are conceivable as uncaused; nor are they sufficient since it is always conceivable that the effect should not ensue.³⁷⁶

Causes and effects are, therefore, not related logically. But neither are they related empirically: For nothing can be found internal to the cause that would explain it. That is, nothing “may point out that circumstance in the cause, which gives it a connexion with its effect.”³⁷⁷ Furthermore, the whole idea of a cause seems to be nonsensical: “We have

³⁷⁵ Op cit. David Chalmers, (1997), p.177

³⁷⁶ Op cit. Stephen Yablo, (2002), p.2

³⁷⁷ Ibid.

no idea of this connexion, nor even any distant notion what it is we desire to know, when we endeavour at a conception of it.”³⁷⁸

Yet if there is no logical connection between cause and effect, no evidential reason to accept it, and what we mean by a “cause” is extremely vague at best, it is still somewhat ubiquitous:

Our thoughts and enquiries are, therefore, every moment, employed about this relation: yet so imperfect are the ideas which we form concerning it, that it is impossible to give any just definition of cause, except what is drawn from something extraneous and foreign to it. Similar objects are always conjoined with similar.³⁷⁹

How are we to explain this paradoxical state of affairs? Hume does so by identifying the locus of the idea of cause and effect in human psychology:

[W]hen many uniform instances appear, and the same object is always followed by the same event; we then begin to entertain the notion of cause and connexion. We then feel a new sentiment or impression, to wit, a customary connexion in the thought or imagination between one object and its usual attendant; and this

³⁷⁸ Ibid.

³⁷⁹ David Hume, (1910), *An Inquiry Concerning Human Understanding*, (Harvard Classics Vol 37) Section VII, part 2

sentiment is the original of that idea which we seek for.³⁸⁰

We can thus understand causation, and mental causation, too, as nothing more than a mere configuration of like events conjoined in the imagination or human mind, but not grounded upon logic, evidence, or strict definition. Thus, the problem of mental causation can be reduced to a psychological need to connect events together, a need which doesn't have an actual basis in logic, evidence or meaning. Then, the problem of mental causation is merely based on this human, though illogical, non-evidential, vague, and all round psychological urge to satisfy some need for connection in the human psyche. It may serve an evolutionary purpose, it may be advantageous to learning, but one needs to keep it in perspective if one is to advance on certain perplexing questions. This fits nicely with the psychological story that has been told about modal confusions above.

For sure, some will think that this kind of conclusion is extreme. Perhaps, a more palatable approach is offered by Lynne Rudder Baker. She is also exercised by the metaphysical knots of mental causation. However, she argues that we should reverse the priority of order between causation and explanation, that is, we should prioritize explanation:

How, then, are we to understand causation? My suggestion is to take as our philosophical starting-point, not a metaphysical doctrine about the nature of causation or of reality, but a range of explanations that have been found worthy of acceptance. These include, pre-eminently but not exclusively, scientific explanations...[M]y proposal is to begin with explanations that earn their keep,

³⁸⁰ Ibid. Section VII, part 2

rather than with the metaphysics, which seems to me a freeloader that just interferes with real work.³⁸¹

An “explanation that earns its keep” is simply to be taken as an ordinary answer to why someone did something, for example, “why didn’t you bring more beer?”³⁸² An ordinary explanation might, then, be, “I just don’t want to drink too much today”. She continues:

If we reverse the priority of explanation and causation that is favoured by the metaphysicians, the problem of mental causation just melts away...Does what we think ever affect what we do? With the reversal of priority of cause and explanation, the metaphysical version of the question does not arise, and the original question has an easy answer.³⁸³

This offers another way to escape the conundrum of mental causation. I have already argued, in relation to Chalmers, that the dualist is not compelled to explanatory epiphenomenalism, roughly because knowledge, judgements, sayings and other behaviours come in two flavours, one kind is necessarily related to conscious phenomena, the other type is not. This latter type may appear to resemble the former type, as a water-like substance may appear to resemble water, but the two are essentially distinct. Knowledge etc. with and knowledge etc. without consciousness may be considered quantitative analogues, but they are essentially distinct. This surely plays a part in the

³⁸¹ Op cit. Lynne Rudder Baker, (1993), pp.92-93

³⁸² Ibid. p.92

³⁸³ Ibid. p.93

explanation of why an entity with the former kind thinks, does or says something and a being with the latter kind thinks, does or says something. However, each kind remain indistinguishable at the physical level. This dissects the two kinds in question and doesn't allow one to move from the physics alone to the knowledge, etc. in question, i.e. Chalmers second premise is defeated, or in my terms some kind of (P) relation is denied, and the epiphenomenal paradox for explanation resolved. So Lynne Rudder Baker's suggestion suits the conclusion drawn here fairly well.

Well, Kim responds to her in the following manner:

But I doubt that very many of us who have “worried” about mental causation have actually been concerned about the possibility that our thoughts and desires might turn out to have no powers to move our limbs. Our worries are not *evidential* or *epistemological* worries...But what all this shows is that the problem of mental causation is primarily a metaphysical problem. It is the problem of showing *how* mental causation is possible, not *whether* it is possible.³⁸⁴

But hasn't Kim spectacularly just missed the point here? How is this response supposed to help him? The whole point is *forget the metaphysical knots and there is no problem!*

He continues, however:

³⁸⁴ Jaegwon Kim, (2000), *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, (Cambridge, Massachusetts: The MIT Press), p.61

[T]he how-question of mental causation arises because there are certain other commitments, whether metaphysical or of other sorts, that demand our presumptive respect but that make mental causation *prima facie* problematic.³⁸⁵

What are these “other commitments”? They aren’t, he says to do with choosing between “explanatory practice” or “epistemology”.³⁸⁶ To me this, again, suggests that metaphysics is the problem *not explanation*. He does attempt to argue that “mental explanations—explanations that invoke mental states in their explanata—are often causal explanations”³⁸⁷ But the question arises whether they need to be given Chalmers sets his paradox out *entirely* in *explanatory* terms, shunning the causal terms because they are so loaded.³⁸⁸ Well, even if causation needs to be taken into account, the point that Baker is making is that explanations of the type that answer “why” questions in an ordinary fashion should be given *priority*, rather than the metaphysical problems associated with the distortions of these simple answers. Those knots can just be put to one side once this is done. I think the dualist can view her stand as more compelling than Kim’s.

Another reason that compounds the case against Kim and the metaphysical notion of mental causation is the fact many, if not all, materialist explanations of mind suffer at its

³⁸⁵ Ibid. p.62

³⁸⁶ Ibid. p.62

³⁸⁷ Ibid., p.63

³⁸⁸ Op cit. Chalmers, (1997)

hands! In fact, since these positions tend to make the claim *for* mental causation, *but* threaten epiphenomenalism, they tend to absurdity! For example, Ted Honderich argues that Davidson's Anomalous Monism is epiphenomenal. Of course, if that is so, it makes Davidson's position look rather absurd for his first principle accepts mental causation. Honderich offers the following analogy:

Certainly it is true that when I put some pears on the scale, something green and French did cause the pointer to move to the two-pound mark, but there in fact is no entailed law connecting greenness and Frenchness with the pointer's so moving. There is in fact no law at all connecting the event in virtue of its being of something green and French with the pointer's moving to the two-pound mark. There is no lawlike connection connecting the first event in virtue of greenness and Frenchness with the second event in virtue of its being the pointer's moving to the two-pound mark.³⁸⁹

Honderich's argument is that mental properties are related to their physical properties in the same way as the Frenchness and greenness in question are related to the movement of the pointer, that is, mental properties may be identified with some physical event (just as the Frenchness and greenness are connected to the pears), but the laws that connect that physical event as cause to another physical event as effect, will do so in respect of some of the properties alone, the relevant properties (just as the law-like relations between the pears and the movement of the scales connected the weight of the pears to the movement of the pointer in respect of some of the properties of the pears and *not* their Frenchness or

³⁸⁹ Op cit. Ted Honderich, (1982)

greenness). The relevant physical properties that connect a physical event with another physical event will be the *physical properties alone* since it is only the physical properties that are connected in law-like fashion, according to Davidson; if one recalls the mental is anomalous unpossessed of law-like relations. That renders mental properties identifiable with physical events but causally irrelevant. And, if one is honest, *absurd!*

I think, it is easy to see that this is an argument that will threaten other forms of non-reductive materialism, too. Of course, the inability to explain mental causation has been fired a broadside at this position many times. Kim has been one of the main architects of such attacks. As Lynne Rudder Baker notes: “Jaegwon Kim has mounted a sustained attack on various versions of (NRM) [Non-reductive materialism] in numerous articles. Unless mental properties are reducible to physical properties, he argues, they are causally inert or else there is massive (and implausible) overdetermination.”³⁹⁰ Here is one argument for causal irrelevancy based on Kim and aimed at non-reductive materialism: Every event has one and only one complete cause. So, every physical event has one and only one complete cause. But every physical event has one and only one complete physical cause because the physical world is causally closed, for example, to preserve the conservation laws. That’s what Dennett tells us, right? No mental event is reducible to a physical event in virtue of its mental properties, assuming that “mental states (events) are...mental properties instantiated at a time”,³⁹¹ following Lynne Rudder Baker.

³⁹⁰ Lynne Rudder Baker, (2008), “Non-Reductive Materialism”, in Brian McLaughlin and Ansgar Beckermann (Eds.), *The Oxford Handbook for the Philosophy of Mind*, (Oxford: Oxford University Press), p.6

³⁹¹ Ibid. p.1

Therefore, no mental event causes a physical event.³⁹²

Kim develops a form of the argument that puts the blame on supervenience, the determination relation that was discussed briefly above. He states his case in the following manner:

Under the assumption of mind-body supervenience, M^* occurs because its supervenience base P^* occurs, and as long as P^* occurs, M^* must occur no matter what other events preceded this instance of M^* —in particular, regardless of whether or not an instance of M preceded it. This puts the claim of M to be a cause of M^* in jeopardy: P^* alone seems fully responsible for, and capable of accounting for, the occurrence of M^* .³⁹³

This eventually leads to the dire conclusion:

If mind-body supervenience fails, mental causation is unintelligible, if it holds, mental causation is again unintelligible. Hence mental causation is unintelligible.”³⁹⁴

Perhaps, Kim should stop here, take note of the false choice, and forget about mental causation altogether, but no. This dilemma leads Kim to take his “reductionist way

³⁹² Cf. Stephen Yablo, (1992), “Mental Causation”, in *Philosophical Review*, 101, 245-80

³⁹³ Op cit. Jaegwon Kim, (2000), p.42

³⁹⁴ Ibid. p.46

out”,³⁹⁵ which of course entails the rejection of non-reductionist materialism. But, now, reductionism is not immune to epiphenomenalism either! This means it, *too*, leads to absurdity, so far as it claims to explain mental causation by the reduction in question! Lynne Rudder Baker makes an argument against the reductionist tendency. I’ll reconstruct her argument in the following way: First, distinguish two aspects of behaviour, “physical properties” and “content properties”, the latter are intentional or mental properties, “those properties in virtue of which beliefs are themselves true or false...properties determined by what she believes, what she desires, and so on”.³⁹⁶ So, for example, “Why did Jill raise her eyebrows?”³⁹⁷ gets the ordinary answer “Because she *wanted* to make a bid for the Rubens”. Second, assume that viable forms of materialism require the causal closure principle and strong supervenience: “[A]ll significant versions of materialism (I think) are committed to CCP [Causal Closure Principle] and SS [Strong Supervenience]”.³⁹⁸ That is, only the physics affects the physics and if P determines M, then M supervenes on P, or as Baker observes of Kim: “If P *realizes* M, then M supervenes on P”,³⁹⁹ where P is a base physical property, and M a supervenient, dependent or determined, property. The problem is that content properties are left causally irrelevant to behaviour on this picture because the physical reduction of neurophysiological processes alone suffices for causal relevancy. Another way to state this is to say that beliefs, desires etc. don’t *need* to be

³⁹⁵ Ibid. p.47

³⁹⁶ Op cit. Lynne Rudder Baker, (1993), p.76

³⁹⁷ Ibid. p.76

³⁹⁸ Ibid. p.86

³⁹⁹ Op cit. Lynne Rudder Baker, (2008), p.11, my italics, see also Jaegwon Kim, (1993), “The Nonreductivist’s Troubles with Mental Causation”, in *Supervenience and Mind*: 309-335

included in stating the causal processes that underlay a behaviour.⁴⁰⁰ That is, if asked, why Jill raised her eyebrows? The answer can be given in neurophysiological terms realized in the physical base alone, due to the fact that the former supervene on the latter. These alone suffice, given the causal closure of the physical. One just doesn't need to say anything about Jill *wanting* to make a bid for the Rubens. This sounds like the same argument that Kim posits against non-reductive materialism. Kim, as noted, hopes to escape by a *reductionist* move. That is, the supervenient properties, including the "want" or the "desire" are *realised at the base* and are thus one with the base and thus causally significant. Lynne Rudder Baker is wary of this but contends that, even if plausible and beliefs and desires etc. are so reduced, *intention-dependent* properties are left irrelevant. For example, "being off-side" (in a football game), "being married", "being the payment of a debt", etc. Take the latter property, the problem with such a property is, first, "it is not at all clear...that the property of being payment of a debt has a single causal role"⁴⁰¹. And, second, she says, "We have no idea what the base properties on which being a payment of a debt supervenes", as required by supervenience. She continues:

Yet, if Kim is right, the causal efficacy of the payment of the debt resides in the nonintentional realizer (whatever that is). So, Kim's view would have us transform a causal connection that we all understand, and that we predict—the causal connection between Jones' payment of his debt and his putting an end to harassing phone calls from his creditor—into a causal connection between we-

⁴⁰⁰ Op cit. Lynne Rudder Baker, (1993), p.83

⁴⁰¹ Op cit. Lynne Rudder Baker, (2008), p.13

know-not-what.⁴⁰²

The problem is not just that we don't know "how to carry out the reduction", she continues, the problem is that "If we have no idea what a reduction would look like, we are in no position to claim that it can be carried out in principle".⁴⁰³ A raft of properties are left causally irrelevant (or caught in limbo) in this manner: "These are properties mentioned in causal explanations of psychology, economics, and political science, as well as in everyday life. They are properties without which we cannot begin to make sense of the world that we encounter."⁴⁰⁴ This is damning, indeed, for not only are intentional properties threatened but most properties making sense of the world we live in based on intentional properties. So not only is mental causation threatened by Kim's reductionism, but the normal causation that we take for granted around us is also put under severe pressure.⁴⁰⁵ She sums up this argument against Kim thus: "Without the "merely epistemological", one has little reason to believe the loftily metaphysical".⁴⁰⁶ And this brings us back to her claim that the order of priority should be reversed, the explanatory should come before the metaphysical, which I have already said is acceptable in the context of the minimal dualism discussed here. The dualist can circumvent absurdity, again, by treading this path.

⁴⁰² Ibid. pp.13-14

⁴⁰³ Ibid. p.14

⁴⁰⁴ Ibid. p.12

⁴⁰⁵ Cf. Op cit. William Seagers, (1999), pp.23-24

⁴⁰⁶ Op cit. Lynne Rudder Baker, (2008), p.14

Well, the main point I am trying to press home is this: mental causation is problematic for everyone, and remains problematic for those who think that they can reject a philosophy of mind, for example, dualism, based on a metaphysically inclined view of mental causation. The worse for them when we discover that the position advocated as a replacement falls foul of the same concerns, which tends thereafter to make their position look absurd, we saw this with Anomalous Monism, Non-reductive Physicalism and a reductionist position.⁴⁰⁷ Well, if there seems to be a choice between mental causal efficacy and mental causal inefficacy, but both lead to rejection, then the choice is a false choice. That is, again, if A or B is your choice, but A is false and B is false, then the choice is a false choice. Above it was contended epiphenomenalism and interactionism with respect to Chalmers and Elitzur, respectively, could both be rejected. Chalmers second premise was rejected, and Elitzur's position didn't seem to actually entail dualism. Now it has been shown that materialist advocates of *mental causation* tend to *epiphenomenalism*, which is absurd. Taking a cue, then from Chalmers' caution and Hume's scepticism, and the false choices that seem to emerge when discussing mental causation, the dualist can take a wary (Chalmers-like), if not outright sceptical (Humean), stance on causation. Or more temperately following Lynne Rudder Baker, one can advocate explanation over metaphysics. And, then, given the argument for a difference in explanation advanced here, provide reasons to think that consciousness has a role to play in explanations, and just stop there; or continue down the road to mental causation grounded upon explanation rather than say the physical laws. For such *explanations* do all the work necessary to explain why, for example, one scratches one's head, without

⁴⁰⁷ Searle's position looks likely to suffer the same fate, given that a carving off of the subjective aspects of experience *could* be done without disrupting the metaphysical story.

having to evoke metaphysical laws or anything else foreign to ordinary human explanation. For these reason, I do not see causality as a burden that dualism has to face alone or has to face at all. Thus, there can be no more reason to reject dualism as opposed to any other system of philosophy that treats the mind, at least as opposed to those considered here.

In this section, so far, I have been considering dualism in a basic form, a minimalist set of assumptions. Some of the problems for dualism, the other moves and strategies against dualism that Kripke mentioned, have been considered. None of them has managed to provide a reason to rule dualism out as a theory of consciousness: no psychological confusion; no privacy conclusion; no better alternatives; no necessary causality problems. It seems to me, then, that Dennett's claims that dualism *has been* found to be wanting and his other charges are more rhetorical than anything, perhaps, in Kripke's terms, *ideological*. For sure he holds to them religiously. This concludes the substantive part of this thesis. I want to end with a brief section which gives a summary of the arguments thus far and draws one or two final conclusions.

Section F: Conclusions

This section will summarize the arguments that have gone before and draw some conclusions. Some thoughts on physicalism will also be registered.

There are genuine modal possibilities and non-genuine modal possibilities. The former are logical, the latter are not. A logical possibility is a possibility that does not entail an absurdity. A logical impossibility, one that does. Therefore, a genuine possibility does not involve absurdity and a non-genuine possibility does. Three related themes in Kripke's work are important: modality, a priori knowledge of modality, and psychology. Facts determine modality; logic determines a priori knowledge of modality, for example, in the form of conditionals. When the facts are in, the conclusions can be drawn. That is, when the facts are in, we learn what is a genuine possibility and what is not. The facts, also, determine psychology, that is, the facts determine what can and what cannot be imagined. When the facts are in, and the logic done, we understand what can really be imagined and what can never be imagined. So, for example, logic relating to essentialist statements tells us, $A \rightarrow \Box A$, $\Diamond \neg A \rightarrow \neg A$, etc. Say, then, the facts tells us A . We can conclude $\Box A$. This is a genuine possibility. On the other hand, $\Diamond \neg A$ is not. The facts, as said, also determine what can be imagined. Thus, it follows that $\Diamond \neg A$, and even $\neg A$ cannot be imagined. The imagination causes confusion, human psychology is a kind of noise. Unfortunately, Kripke says little about that. So one is left wondering how it is that the facts determine what can be imagined, yet the imagination causes modal confusion, that is, allows us to think that what we imagine is something that is genuinely not possible? The picture theory of the imagination developed in this thesis tries to explain this.

Two functions are associated with pictures. The representation function and the representation-as function. So if P is a picture and it represents x , then it denotes x . If P is a picture and it represents x as F , then it refers to, or exemplifies, a set of descriptions (including labels), F , and F describes what x is represented as. If P is a picture and it represents x , and $x = y$, then it denotes y . Unfortunately, if P is such a picture and it represents x as F , where F is a set of descriptions that describe x , less its essential attribute, y -ness, then P can lead us to think that we imagine that x is not necessarily y . But this is not so. Because the facts are in, and the maths is done, and they determine $\Box x = y$. Rather, the representation-as function does not represent x as y , so does not represent x as x , but represents x as x -like given it represents x as F , where F is a set of descriptions that describe x , less its essential attribute, y -ness. That is, P represents x and represents x as x -like, less y -ness, which is called a qualitative analogue of x . If we understand pictures we can understand how thinking that x is not necessarily y can occur, despite the facts, despite the logic, despite it being genuinely *impossible* and impossible to imagine. One needs to make sure the representation and representation-as function are clearly distinguished in order to make sure we understand what a picture truly represents. Imagining is taken as mental picturing. Mental picturing is taken as viewing mental pictures in the mind's eye, etc. Thus, imagining is prone to pictorial confusions. This explains how, even if one cannot say or imagine $\Diamond \neg A$, one can come to think what one talks about or imagines is a state of affairs that is not genuinely possible. But really one is just confusing a representation of A , for what it is represented as, $\neg A$.

There is another way that pictures may confuse us. This has to do with impoverished

knowledge or conceptual paucity. If P is a picture and it represents x as F, where F is a set of descriptions that describe y, less *inessential* attributes commonly associated with x-ness, then P can lead one to think that one imagines that x is not necessarily y, when *in fact* $x = y$, and so $\Box x = y$. In fact, due to impoverished knowledge or conceptual paucity, we might think that what is being represented is anything but x. But this is not so. Because the facts are in, and the maths is done, and they determine $\Box x = y$. Rather, the representation-as function does represent x as y, but we just can't see it. This time imagining is prone to pictorial confusions because we cannot see that the representation and the representation-as function connect our imagining to one object. This explains how if one thinks one cannot say or imagine $\Box A$, that thought is based on a lack of understanding or conceptual paucity.

Now, the materialist claims: $\Box(P \leftrightarrow Q)$, where P abbreviates some statement of the physical facts and Q abbreviates a statement about the presence of an associated mental state, and where this is based on an identity claim. That further implies, $\Box(Q \rightarrow P) \ \& \ \Box(P \rightarrow Q)$. But the dualist claims $\Diamond(Q \ \& \ \neg P) \ \& \ \Diamond(P \ \& \ \neg Q)$. That entails, $\neg(P \leftrightarrow Q)$. Specifically, say P abbreviates a statement about some neurophysiological facts and Q abbreviates a statement about the presence of pain: The materialist wants to say that the dualist is not really talking about pain, but the qualitative aspects of pain alone, and the materialist may want to *justify* this by showing that the psychological rain of the imagination is interfering with the dualist's modal judgements. But he, the materialist, cannot do this employing the Kripkean strategy discussed above. And this can be shown by the use of my pictorial theory of the imagination. I will briefly go over why. Given what has been said, if P represents pain as pain-like but not neurophysiological, then the

materialist wants to say, P doesn't represent pain as actual pain, rather P represents pain as a qualitative analogue of pain. But pain = felt-pain. So pain is not being represented as a qualitative analogue of pain, at all, but the real thing. So, P represents pain as pain-like, but not neurophysiological, *and* it is false that pain is not represented as actual pain. In the opposite direction, if P represents pain as neurophysiological but not pain-like, then the materialist wants to say pain is *still* being represented as actual pain, so pain is represented as pain actually is. But pain = felt-pain. So, P doesn't represent pain as pain actually is. So, P represents pain as neurophysiological, but not pain-like, *and* it is false that pain is *still* being represented as actual pain. Therefore, the materialist cannot say that the dualist is wrong to say that, $\diamond(Q \ \& \ \neg P) \ \& \ \diamond(P \ \& \ \neg Q)$, and $\neg(Q \leftrightarrow P)$. However, just because this non-identity is not ruled out by a psychological confusion, ignorance or conceptual paucity identified in the Kripkean manner, it doesn't mean it is ruled in. For this, and other reasons (psychological and logical) modest conclusions are drawn. Such propositions may, on modest grounds, therefore be taken in the form of a hypothesis or on a conditional basis in order to work out what commitments are entailed or assumed in order to falsify them. And this is what the rest of chapter two undertook to do.

A dualist needs to hold a number of assumptions. First, there is something it is like to be conscious; second, a proposition like $\diamond(P \ \& \ \neg Q)$, which can't be ruled out by appeal to a psychological confusion as shown above, that is, where P abbreviates a statement about some physical facts and Q a statement about the presence of some associated mental state; third, I held, that a dualist should accept that it is possible to know what it's like to be another, for anti-sceptical, existential, ethical, social, scientific and phenomenological reasons. In fact, I think, that if a dualist should reject that proposal, dualism would face a

barrage of criticism and, perhaps, on each of those grounds be falsified. So, these are the minimal number of assumptions a dualist should accept, of course, in a hypothetical manner. Does this set of assumptions entail absurdity? Well, in conjunction with the physicalist contention that knowledge is determined by physics, a contradiction is entailed. For it will follow that if there is physical duplicate of, say, Thomas Nagel, then that duplicate will know everything that Thomas Nagel knows, and that duplicate will not know something that Thomas Nagel knows, for example, what it's like to be Thomas Nagel. Therefore, the dualist must reject the physicalist contention. That is, knowledge of qualitative aspects is not essentially physical. Now, this leads to a kind of private knowledge, and that tends to absurdity from what seems like any number of angles. There is, however, a solution. Knowledge of the qualitative aspects of experience is *public* but one is just not acquainted with it through the third-person. The problem is that if xKq^x , that is, if x knows what it's like to be in pain for x such knowledge is not determined neurophysiologically, functionally or behaviourally so far as one rejects a physical determination of such knowledge and so is not deterministically or evidentially known in such fashion by any other. That tends towards the conclusion that such knowledge is private, which can be represented by the following formula: $(\exists x)[xKq^x \ \& \ \square(y)(yKq^x \rightarrow y = x)]$. This is read, for some x , x knows what it's like to be in pain for x , and necessarily for any y who knows what it's like to be in pain for x , y is identical to x . What is the solution to this problem? It was said that one can know what it's like to be another *merely from one's own experiences*. Here is the formula: $(x)[xKq^x \rightarrow \diamond(\exists y)((yKq^x \rightarrow yKq^y) \ \& \ y \neq x)]$, which reads, for all x , if x knows what it's like to be in pain for x , then it is possible that someone, y , knows what it's like to be in pain for x only if y knows what it's like to be in pain for y , even though y is not x . Therefore, the rejection of a physical determinant

of knowledge, together with the acceptance of the minimal requirements of dualism, does not lead to the absurdity of private knowledge. For sure, there may be a private *way* of knowing, but no private knowledge is entailed. This is intuitive and fits the neuroscience.

Some alternatives to the kind of dualism formulated were considered and rejected. For example, giving up (K) for (P) led to absurdity and obscurity. Next, Davidson and Searle's respective positions. Davidson's position was found not to be necessitated, and, further, if it were accepted it would come with a grave number of problems: mystery, scepticism and absurdity! Searle's position seems to twist language in order to explicate his view leading to a number of serious issues: logical truths are dependent on description and seem disconnected from metaphysical truths, which is false on a Kripkean account of things; accepting physical determinants seems to tend to absurdity as shown by the reconstruction of Jackson's argument in terms of epistemic logic, an argument that Searle accepts as valid; if a causal reduction of mind doesn't lead to ontological reduction like the reduction of solidity does, *by definition*, this suggests two different types of thing are being talked of; last, the mere fact of talking about two things that are *ontologically* distinct and irreducible suggests two different types of thing. Searle's position offers too many problems, turns of speech and artifices to be confident of accepting. The dualist has no need, thereof, to give way to its obscurities.

Finally, causality was considered. Dualism is thought to have to face a choice between epiphenomenalism and interactionism (parallelism mentioned in chapter one was not mentioned as an alternative because it can be considered as a kind of epiphenomenalism because parallelism is often defined as the rejection of mind to body and body to mind

causation, which is, then, just a mind to body and body to mind epiphenomenalism). The choice is a false choice: if A or B are your choices, and A is false and B is false, then the disjunction is false. The general discussion surrounding dualism suggests that the choice is false. I looked, first, at two dualistic views which do seem to indicate that the choice in question is a false one. Chalmers and Elitzur's views was first considered. Chalmers argues for explanatory epiphenomenalism, which threatens causal epiphenomenalism. Elitzur argues for interactionism, in contrariant form to Chalmers' view. But both views are found wanting. Chalmers view doesn't just lead to a paradox, but leads to absurdity, for the very same reasons as holding that the assumptions I associated with dualism when conjoined with the proposition that knowledge is determined by physics leads to absurdity. Furthermore, Kripke's strategy reconstructed by the picture theory of the imagination developed in this paper explains a confusion that might be attributed to Chalmers. Basically, Chalmers thinks x and y if physically indistinct, even where only x enjoys and suffers the qualitative aspects of experience, nevertheless share judgements, including those about the qualitative aspects of experience. But I argue that Chalmers is confusing a quantitative analogue with the real thing, that is, Chalmers is confusing judgements that physically resemble each other, but are essentially distinct.

Recall, If P is a picture and it represents x as F, then it refers to a set of descriptions, F, and F describes what x is represented as. If P is a picture and it represents x, and $x = y$, then it denotes y. Unfortunately, if P is such a picture and it represents x as F, where F is a set of descriptions that describe x, less its essential attribute, y-ness, then P can lead us to think that we imagine that x is not necessarily y. But this is not so. Rather, the representation-as function does not represent x as y, so does not represent x as x, but

represents x as x -like given it represents x as F , where F is a set of descriptions that describe x intimately, less its essential attribute, y -ness. That is, P represents x and represents x as x -like, which is, in this case, called a *quantitative analogue* of x . So Chalmers has a picture of a judgement of consciousness made by x represented as a judgement of consciousness made by y , where the latter judgement lacks conscious content, which makes him think that he imagines such judgements do not entail such conscious content. But so far as it is held that the physics does not determine such judgements but that consciousness is what essentially determines such judgements, or essentially distinguishes judgements, Chalmers cannot really be said to imagine such a thing. Chalmers mistake is, then, to confuse the representation of a judgement of consciousness, which is essentially related to qualitative content, with a *quantitative analogue* of such a judgement in the picture before his mind's eye. That is, he thinks that he can picture a judgement about consciousness without any relation to consciousness because he pictures such a judgement *as* something quantitatively (physically) indistinguishable from the real thing, but qualitatively, and therefore *essentially*, distinct from the real thing. That is, he pictures the real thing as a quantitative analogue of itself. Confusion is honest here because of the pronounced physical resemblance. This has a consistent symmetry to the arguments deployed earlier.

Now, Elitzur's argument is rejected because it doesn't entail dualism, in fact, it seems to rule in a necessary entailment from the physics to the presence of the subjective aspects of experience. To hold $\diamond(P \ \& \ \neg Q)$, that is, the possibility of the occurrence of a physical processes, for example, and the absence of an associated qualitative property is to hold $\diamond(\exists x)(\exists y)(x \text{ }_{p=} \text{ } y \ \& \ \neg x \text{ }_{q=} \text{ } y)$, that is, where $_{p=}$ abbreviates 'is physically indistinguishable

from', and $q=$ abbreviates 'is experientially indistinguishable from', where that difference is understood to be in terms of one of the parties in question lacking the capacity to enjoy and suffer the qualitative aspects of experience. But according to Elitzur, two beings that are *near* physical duplicates, but where one lacks consciousness, are ultimately distinct due to a physically identifiable difference. But that implies $\neg\Diamond(\exists x)(\exists y)(x_{p=} y \ \& \ \neg x_{q=} y)$, or $\Box(x)(y)(x_{p=} y \rightarrow x_{q=} y)$, which implies, $\Box(P \rightarrow Q)$. This doesn't imply dualism. Elitzur's position is consistent, therefore, with materialism, for example, a non-reductionist materialism.

The conclusion reached was that the choice between epiphenomenalism and interactionism for dualism was a false choice. But it was argued that these problems could be identified everywhere, which rather pointed to a defect with the idea of mental causation as a whole. For example, Davidson rejects epiphenomenalism, and opts for mental causation, but his position entails epiphenomenalism. So his view of mental causation is absurd. Non-reductive materialism rejects epiphenomenalism, but, in turn, is easily reduced to absurdity. Reductionists, then, reject epiphenomenalism, but likewise end up holding an absurd position. This suggests that something is wrong with the whole idea of mental causation. A Humean style rejection of causality was one possible outcome, or a more temperate reversal of priorities as advocated by Lynne Rudder Baker was another option; explanations to be given more weight than causation. The main point was not the solution to the problem, but to show that the problem was ubiquitous and could not be solely aimed at dualism, but could be aimed at any rival and any alternative theory. Because the dualism in question could avoid taking any position on causality, and such problems were ubiquitous for other alternatives, there was no reason to give dualism up

on that basis.

Well, these were the charges that I thought might be thrown at the kind of dualism, minimal in form, postulated here. But it was able to successfully evade these charges, it was not reduced to absurdity and was not falsified.

However, what about physicalism? What has been learned about physicalism on our journey thus far? First, physicalism cannot defend the identity between some physical fact and an associated mental fact by appealing to a psychological error; second, appealing to such an error tends to suggest the *falsity of physicalism*; third, if dualism is accepted, physicalism not only fails to determine consciousness, but fails to determine knowledge, judgements, sayings, etc. about consciousness; fourth, rejecting dualism in order to save knowledge for physical determination leads to absurdity unless an epistemic connection between the physical and consciousness is denied under ideal circumstances; fifth, accepting that leads to mystery upon mystery upon mystery upon absurdity; sixth, alternative physicalist views to the dualism inspired by Kripke's argument are either mysterious, skeptical and absurd, like Davidson's, or seem obscure and threaten absurdity, like Searle's; last, mental causation is an extreme problem for physicalism in that it tends to absurdity. The questions, then, that I want to finish with are these: *Why would anyone choose to be a physicalist? Are there social and psychological factors, even "ideological" factors in Kripke's words, which keep philosophers in the grip of a physicalist picture? Isn't it incumbent on philosophers to take this possibility seriously and at least ask the question why physicalists remain physicalist in spite of the tremendous pressures, which approach absurdity, on their doctrine?* As for the alternative, dualism, as hypothesized

here, it remains very much alive and a much strengthened alternative to the absurdities and obscurities identified with physicalism.

References

J R. **Anderson**, (1978), “Arguments Concerning Representations of Mental Imagery” in *Psychological Review*, 85, pp.249-277

George **Bealer**, (2002), “The Origin of Modal Error” in *Dialectica*, Vol. 58, No. 1, 11-42

Ned **Block**, (1978), Troubles With Functionalism, *Minnesota Studies in the Philosophy of Science*, 9, pp.261-325

Ned **Block**, (1995), “On a confusion about a function of consciousness” in *Behavioural and Brain Sciences*, 18, pp.227–47

Ned **Block**, (2011), “Perceptual Consciousness Overflows Cognitive Access” in *Trends in Cognitive Sciences*, December 2011, Vol. 15, No. 12, pp.567-575

Peter **Bokulich**, (2005), “Does Black Hole Complementarity Answer Hawking’s Information Loss Paradox?” in *Philosophy of Science*, 72 (December 2005), pp.1336–1349

C. D. **Broad**, (1925), *The Mind and Its Place in Nature*, (London: Routledge and Kegan Paul)

Alex **Byrne**, (2007), “Possibility and Imagination”, extracted <http://web.mit.edu/abyrne/www/possandimag.pdf> 20/07/2013

A. J. **Calder**, J. **Keane**, F. **Manes**, N. **Antoun**, and A.W. **Young**, (2000), “Impaired Recognition and Experience of Disgust Following Brain Injury”, in *Nature Neurosci.* 3, pp.1077-1078

Tracy **Cassels**, Sherilynn **Chan**, Winnie **Chung** and Susan A. J. **Birch**, (2010), “The Role of Affective Empathy: Cultural and Bicultural Differences”, in *Journal of Cognition and Culture* 10, (2010), pp.309–326

David **Chalmers**, (1995), Facing up to the problem of Consciousness”, in *Journal of Consciousness Studies* 2(3), pp.200-219, extracted <http://consc.net/papers/facing.html>, 15/11/2013

David **Chalmers**, (1997), *The Conscious Mind: In Search of a Fundamental Theory*, (New York and Oxford: Oxford University Press)

David **Chalmers** and Frank **Jackson**, (2001), “Conceptual Analysis and Reductive Explanation”, in *The Philosophical Review*, 110, pp.315-361

David **Chalmers**, (2002), “Does Conceivability Entail Possibility”, in (Eds.) T. Gendler & J. Hawthorne, (2002), *Conceivability and Possibility*, (Oxford University Press), pp.145-200, extracted <http://consc.net/papers/conceivability.html> 17/11/2013

David **Chalmers**, (2004), “Phenomenal Concepts and the Knowledge Argument”, in (Eds.) Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, (2004) *There’s Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson’s Knowledge Argument*, (Cambridge, Massachusetts: The MIT Press), pp.269-299

David **Chalmers**, (2006), “The Foundations of Two-Dimensional Semantics”, in (Eds.) Manuel Garcia-Carpintero and J Macia, (2006). *Two-Dimensional Semantics: Foundations and Applications*, (Oxford: Clarendon Press), pp.55-140

David **Chalmers**, (2009), “The Two-Dimensional Argument Against Materialism”, in (Eds.) Ansgar Beckermann, Brian McLaughlin, and Sven Walter, (2009), *The Oxford Handbook of the Philosophy of Mind*, (Oxford: Oxford University Press), extracted <http://consc.net/papers/2dargument.html> 17/11/2013

Noam **Chomsky**, (1965), *Aspects of the Theory of Syntax*, (Cambridge, Massachusetts: The MIT Press)

Paul **Churchland**, (1981), “Eliminative Materialism and the Propositional Attitudes”, in. *Journal of Philosophy* 78, 67-90

Michael A. **Cohen** and Daniel C. **Dennett**, (2011), “Consciousness Cannot Be Separated From Function”, in *Trends in Cognitive Science*, 15, pp.358-364

Michael A. **Cohen** and Daniel **Dennett**, (2012), Response to Fahrenfort and Lamme: Defining Reportability, Accessibility and Sufficiency in Conscious Awareness” in *Trends in Cognitive Sciences*, Vol. 16, No. 3,

Samuel **Clarke**, (1717), *A Collection of Papers, Which Passed Between the Late Learned Mr. Leibnitz and Dr. Clarke, In the Years 1715 and 1716* (London: 1717), extracted <http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00236>, on 15/01/2013

M. J. **Cresswell** and G. E. **Hughes**, (1996), *A New Introduction to Modal Logic*, (London: Routledge)

Donald **Davidson**, (1970), “Mental Events” in Donald Davidson, (2001), *Essays on Actions and Events*, (Oxford: Clarendon Press), pp.206-227

Daniel **Dennett**, (1978), *Brainstorms*, (Cambridge, Massachusetts: The MIT Press)

Daniel **Dennett**, (1991a), “The Brain and Its Boundaries”, extracted <http://ase.tufts.edu/cogstud/papers/mcginn.htm>, 15/11/2013

Daniel **Dennett**, (1991b), *Consciousness Explained*, (New York, Boston and London: Back Bay Books, Little, Brown and Company)

Daniel **Dennett**, (1992), Back from the Drawing Board, extracted <http://ase.tufts.edu/cogstud/papers/backdraw.htm> 15/11/2013

Daniel **Dennett**, (1993a), “Living on the Edge”, in *Inquiry* 36, (1-2), pp.3-16 extracted <http://cogprints.org/270/1/liveedge.htm> 15/11/2013

Daniel **Dennett**, (1993b), “Caveat Emptor”, in *Consciousness and Cognition*, vol. 2, no. 1, 1993, 48-57, extracted <http://cogprints.org/276/1/caveatem.htm>, 15/11/2013

Daniel **Dennett**, (2003), “Who’s on First: Hetrophenomenology Explained”, in *Journal of Consciousness Studies, Special Issue: Trusting the Subject*, (Part 1), 10, No. 9-10, October 2003, pp. 19-10

Daniel **Dennett**, (2006), “Hetrophenomenology Reconsidered”, extracted ase.tufts.edu/cogstud/papers/hetroreconsidered.pdf, 20/11/2013

Rene **Descartes**, (1985), *Meditations on First Philosophy*, (Ed.) (Trans.) John Cottingham, (1985), (Cambridge: Cambridge University Press)

Sir John **Eccles** and Karl **Popper**, (2003), *The Self and the Brain*, (London and New York: Routledge)

Sir John **Eccles**, (1994), *How the Self Controls Its Brain*, (Berlin: Springer-Verlag)

Catherine Z. **Elgin**, (1983), *With Reference to Reference*, (Indianapolis, Indiana: Hackett Publishing Company)

Catherine Z. **Elgin**, (2009), “Exemplifications, Idealization, and Understanding”, in (Ed.) Mauricio Suarez, (2009), *Fictions in Science: Essays on Idealization and Modelling*, (London: Routledge), pp.77-90

Avshalom **Elitzur**, (2009), “Consciousness Makes a Difference: A Reluctant Dualist’s Confession”, in A. Batthyany & A. C. Elitzur (eds.), *Irreducibly Conscious: Selected Papers on Consciousness*, (Heidelberg and New York: Winter), and extracted <http://a-c-elitzur.co.il/site/siteArticle.asp?ar=67> 5/11/2013

Johannes J. **Fahrenfort** and Victor A.F. **Lamme**, (2012), “A True Science of Consciousness Explains Phenomenology” in *Trends in Cognitive Sciences*, Vol. 16, No. 3, pp.138-139

Paul **Feyerabend**, (1993) “Materialism And The Mind-Body Problem” in M. S. Christensen and D. R. Turner, (eds.), *Folk Psychology and the Philosophy of Mind*, Psychological Press, pp.3-17

G. W. **Fitch**, (2004), *Saul Kripke*, (Montreal and Kingston: McGill-Queens University Press

Francisco **Valera** and Jonathon **Shear**, (1999), “First-Person Methodologies: What Why How?”, in *Journal of Consciousness Studies*, **6**, No. 2–3, pp.1–14

Vittorio **Gallese**, (2001), “The ‘Shared Manifold’ Hypothesis: From Mirror Neurons to

Empathy”, in *The Journal of Consciousness Studies*, 8, No. 5-7, 2001, pp.33-50

Vittorio **Gallese**, (2003), “The Manifold Nature of Interpersonal Relations: The Quest for a Common Mechanism”, in *Philosophical Transactions of the Royal Society*, (2003) 358, pp.517-528

Vittorio **Gallese**, (2009a), “Mirror Neurons, Embodied Simulation, and the Neural Basis of Social Identification”, in *Psychoanalytic Dialogues*, 19, pp.519-536

Vittorio **Gallese**, (2009b), “We-ness, Embodied Simulation, and Psychoanalysis: Reply to Commentaries” in *Psychoanalytic Dialogues*, 19, pp.580-584

Nelson **Goodman**, (1976), *Languages of Art*, (Indiana, Indianapolis: Hackett Publishing Company, Inc.)

Christopher **Hill**, (1997), “Imagination, Conceivability, Possibility and the Mind-Body Problem”, in *Philosophical Studies*, 87, pp.61-85

Ted **Honderich**, (1984), “Donald Davidson’s Anomalous Monism and the Champion of Mauve”, in *Analysis* 44, extracted <http://www.ucl.ac.uk/~uctytho/AnMon.htm>, 29/10/2013

Terence **Horgan**, (1993), “From Supervenience to Superdupervenience” in *Op cit.*, Chalmers (2002), pp.150-162

David **Hume**, (1910), *An Inquiry Concerning Human Understanding*, (Harvard Classics Vol 37)

David **Hume**, (1968), *Treatise of Human Nature*, (Oxford: Clarendon Press)

W. D. **Hutchinson**, K. D. **Davis**, A. M. **Lazano**, R. R. **Taskev**, and J. O. **Dostrovsky**, (1999), "Pain Related Neurons in the Human Cingulate Cortex, in *Nature Neurosci.* 2, pp.403-405

T.H. **Huxley**, (1992), "On the Hypothesis That Animals Are Automata," in Brian Beakley and Peter Ludlow (Eds.), *The Philosophy of Mind*, ed. (Cambridge, MA: the MIT Press, 1992), pp.133-136

Ray **Jackendoff**, (1987), *Consciousness and the Computational Mind*, (Cambridge, Massachusetts: The MIT Press)

Frank **Jackson**, (1982), "Epiphenomenal Qualia" in (Eds.) Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, (2004) *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, (Cambridge, Massachusetts: The MIT Press), pp.39-51

Frank **Jackson**, (2003) "Mind and Illusion", in (Eds.) Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, (2004) *There's Something About Mary: Essays on Phenomenal*

Consciousness and Frank Jackson's Knowledge Argument, (Cambridge, Massachusetts: The MIT Press), pp.421-433

William **James**, (1879), "Are We Automata?", extracted 15/08/2013, from <http://psychclassics.yorku.ca/James/automata.htm>

Jaegwon **Kim**, (1993), "The Nonreductivist's Troubles with Mental Causation", in *Supervenience and Mind*: 309-335

Jaegwon **Kim**, (2000), *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, (Cambridge, Massachusetts: The MIT Press)

Amy **Kind**, (2001), "Putting the Image Back into Imagination", in *Philosophy and Phenomenological Research* 62, pp.85-109

Saul **Kripke**, (1971), "Identity and Necessity" in (Ed.) M. K. Munitz, (1971), *Identity and Individuation*, (New York: New York University Press), pp.135-164

Saul **Kripke**, (1980), *Naming and Necessity*, (Cambridge, Massachusetts: Harvard University Press)

Stephen **Kosslyn**, (1983), *Ghosts in the Mind's Machine*, (New York: Norton)

S. M. **Kosslyn**, W. L. **Tompson**, W. L., G. **Ganis**, (2006), *The Case for Mental Imagery*,

(Oxford: Oxford University Press)

Mark **Kulstad** and Laurence **Carlin**, (2013), "Leibniz's Philosophy of Mind", *The Stanford Encyclopaedia of Philosophy* (Winter 2013 Edition), Edward N. Zalta (ed.), extracted <http://plato.stanford.edu/archives/win2013/entries/leibniz-mind/> 17/11/2013

Gottfried **Leibniz**, (1686a), *First Truths*, (Trans.) Jonathon Bennett, extracted http://www.earlymoderntexts.com/f_leibniz.html

Gottfried **Leibniz**, (1686b), *Monadology*, (Trans.) Jonathon Bennett, extracted http://www.earlymoderntexts.com/f_leibniz.html 01/11/2013

Joseph **Levine** (1983), "Materialism and Qualia: The Explanatory Gap" in *Pacific Philosophical Quarterly*, 64, pp.354-361

Joseph **Levine**, (1998), "Conceivability and the Metaphysics of Mind", in *Nous* 32-4, pp.449-80

David **Lewis**, (1972), "Psychophysical and Theoretical Identifications" in *Australasian Journal of Philosophy*, 50, pp.249-58

David **Lewis**, (1988), "What Experience Teaches", in (Eds.) Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, (2004) *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, (Cambridge,

Massachusetts: The MIT Press), pp.77-105

Benjamin **Libet**, (1999), “Do We Have Free Will”, in *Journal of Conscious Studies*, 6, No. 8-9, pp.47-57

Benjamin **Libet**, (2002), “The Timing of Mental Events: Libet’s Experimental Findings and Their Implications” in *Consciousness and Cognition* 11, pp.291–299

Benjamin **Libet**, (2005), *Mind Time: The Temporal Factor in Consciousness*, (Cambridge, Massachusetts: Harvard University Press)

Benjamin **Libet**, (2006), “Reflections on the Interaction of the Mind and Brain” in *Progress in Neurobiology* 78 (2006), pp.322-326

Colin **McGinn**, (1989), “Can We Solve the Mind-Body Problem”, in *Mind*, New Series, Volume 98, Issue 391 (Jul., 1989), pp.349-366

Colin **McGinn**, (1993), *Problems in Philosophy: The Limits of Inquiry*, (Oxford: Blackwell Publishers)

Colin **McGinn**, (1999), *The Mysterious Flame: Conscious Minds in a Material World*, (New York: Basic Books)

Colin **McGinn**, (2006), *Mindsight: Image, Dream, Meaning*, Boston, (Massachusetts:

Harvard University Press)

J. J. **Meyer**, (2001), "Epistemic Logic" in (Eds.) L. Goble, (2001), *The Blackwell Guide to Philosophical Logic*, (Oxford: Blackwell), pp.183-203

G. E. **Moore**, (1922), "The Conception of Intrinsic Value",
<http://www.ditext.com/moore/intrinsic.html> extracted 01/11/2013

Thomas **Nagel**, (1974), "What is it like to be a bat?", *Philosophical Review*, 83, pp.435-50

Martine **Nida-Rumelin**, (1995), "What Mary Couldn't Know: Belief About Phenomenal States", in (Ed.) Thomas Metzinger, (1995), *Conscious Experience*, (Exeter, UK: Imprint academic), pp.219-241

David **Papineau**, (2007), "Kripke's Proof is Ad Hominem not Two-Dimensional", in *Philosophical Perspectives*, Vol. 21, pp.475-494

David **Papineau**, (2009), "Kripke's Proof That We Are All Intuitive Dualists" extracted
<http://www.kcl.ac.uk/ip/davidpapineau/Staff/Papineau/OnlinePapers/Kripke's%20Proof.htm>
27/10/2013

Ian B. **Phillips**, (2011), "Perception and Iconic Memory: What Sperling Doesn't Show" in *Mind & Language*, Vol. 26, No. 4 September 2011, pp.381-411

Gualtiero **Piccinini**, (2009), “First Person Data, Publicity, and Self Measurement”, in *Philosophers Imprint*, Volume 9, No. 9, October 2009

Ullin **Place** (1956), “Is Consciousness a Brain Process”, in David Chalmers, (2002), *Philosophy of Mind: Classical and Contemporary Readings*, (Oxford: Oxford University Press), pp.55-60

Alvin **Plantinga**, (1970), “World and Essence”, in *The Philosophical Review*, 79, No. 4, (Oct. 1970), 461-492

Thomas **Polger**, (1999), “Kripke and the Illusion of Contingent Identity”, (APA Eastern Division (Boston, Massachusetts) Colloquium Paper), extracted <http://homepages.uc.edu/~polgertw/Polger-APA1999.pdf>, 25/11/2013

Hilary **Putnam**, (1975a), “The Nature of Mental States”, in Hilary Putnam, (1975), *Mind, Language and Reality*, (Cambridge: Cambridge University Press), pp.429-441

Hilary **Putnam**, (1975b), “The Meaning of ‘Meaning’” in Hilary Putnam, (1975), *Mind, Language and Reality*, (Cambridge: Cambridge University Press), pp.215-272

Z. W. **Pylyshyn**, (2003), “Return of the Mental Image: Are There Pictures in the Brain?” in *Trends in Cognitive Sciences*, 7, pp.113-118

V. S. **Ramachandran** and William **Hirstein**, (1997), “Three Laws of Qualia: What Neurology Tells Us about the Biological Functions of Consciousness, Qualia and the Self”, in *Journal of Consciousness Studies*, 4, No. 5-6, pp.429-58

Thomas **Reid**, (1785), *Essays on the Intellectual Powers of Man*, extracted www.earlymoderntexts.com/authors/reid.html, 01/12/2013

Richard **Rorty**, (1965), “Mind-Body Identity, Privacy, and Categories”, in *Review of Metaphysics* 19, pp.24-54;

Richard **Rorty**, (1970), “In Defence of Eliminative Materialism” in *Review of Metaphysics* 24, pp.112-121.

Jean-Michel **Roy**, Jean **Petitot**, Bernard **Pachoud** and Francisco **Varela**, (1999), “Beyond the Gap: An Introduction to Naturalizing Phenomenology”, in (Eds.) Jean Petitot, Francisco Varela, Bernard Pachoud, and Jean-Ichel Roy, (1999), *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*, (Stanford, California: Stanford University Press), pp.1-83

Lynne **Rudder Baker**, (1993), “Metaphysics and Mental Causation”, in (Ed.) John Heil and Alfred Mele, (1993), *Mental Causation*, (Oxford: Clarendon), pp.75-95

Lynne **Rudder Baker**, (2008), “Non-Reductive Materialism”, in Brian McLaughlin and Ansgar Beckermann (Eds.), *The Oxford Handbook for the Philosophy of Mind*, (Oxford: Oxford University Press), pp.109-120

Bertrand **Russell**, (1927), *The Analysis of Matter*, (London: Kegan Paul)

Stephen **Schiffer**, (1987), *Remnants of Meaning*, (Cambridge, Massachusetts: The MIT Press)

William **Seagers**, (1999), "Generalised Epiphenomenalism", extracted <http://www.uts.utoronto.ca/~seager/mywork.htm>, 10/01/2013

John **Searle**, (1994), *The Rediscovery of Mind*, (Cambridge, Massachusetts: The MIT Press)

John **Searle**, (2002a), "Animal Minds" in *Consciousness and Language*, (Cambridge: Cambridge University Press), pp.61-77

John **Searle**, (2002b), *Consciousness and Language*, (Cambridge: Cambridge University Press)

John **Searle** (2002c), "Why I Am Not a Property Dualist", *The Journal of Consciousness Studies*, 9, No. 12, pp.57-64

John **Searle**, (2004), *Mind: A Brief Introduction*, (New York and Oxford: Oxford university Press)

Simone G, **Shamay-Tsoory**, Judith **Aharon-Peretz** and Daniella **Perry**, (2009), “Two Systems for Empathy: A Double Dissociation Between Emotional and Cognitive Empathy in Inferior Frontal Gyrus Versus Ventromedial Prefrontal Lesions”, in *Brain: A Journal of Neurology*, 2009: 132; pp.617–627

Sydney **Shoemaker**, (1982), “The Inverted Spectrum”, in *The Journal of Philosophy*, Volume LXXIX, No. 7, July, pp.367-81

Tania **Singer**, Ben **Seymour**, John **O’Doherty**, Holger **Kaube**, Raymond J. **Dolan**, Chris D. **Frith**, (2004), “Empathy for Pain Involves the Affective but Not Sensory Components of Pain”, in *Science* Vol. 303 20 February, pp.1157-1162

Peter **Singer**, (1990), Do Animals Feel Pain?, Excerpted from *Animal Liberation*, 2nd edition, New York: Avon Books, 1990, pp. 10-12, pp.14-15, extracted <http://www.animal-rights-library.com/texts-m/singer03.htm> 16/09/2013

Scott **Soames**, (2006), “The Philosophical Significance of The Kripkean Necessary *Aposteriori*” in *Philosophical Issues: A supplement to Nous*, Volume 16, Issue 1, September 2006, pp.288-309

Quentin **Smith**, (1999), “A More Comprehensive History” in (Eds.) Paul W. Humphreys and James H. Fetzer, (1999), *The New Theory of Reference: Kripke, Marcus, and Its Origins*, (Dordrecht, The Netherlands: Kluwer Academic Publishers), pp.235-285

George **Sperling**, (1960), “The Information Available in Brief Visual Presentations”, in *Psychological Monographs: General and Applied*, 74, (11, Whole No. 498), pp.1-29

Baruch **Spinoza**, (1677), *Ethics*, (Trans.) R.H.M Elwes, (1883), (MTSU Philosophy Webworks Hypertext Edition), extracted <http://capone.mtsu.edu/rbombard/RB/Spinoza/ethica-front.html> 17/01/2013

Daniel **Stoljar**, (2001), “Two Conceptions of the Physical”, in *Philosophy and Phenomenological Research* 62: 253-281

Daniel **Stoljar** and Yujin **Nagasawa**, (2004), “Introduction”, in in (Eds.) Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, (2004) *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, (Cambridge, Massachusetts: The MIT Press), pp.1-37

Leonard **Susskind**, (2003), “Black Holes and the Information Paradox”, in *Scientific American*, Scientific American Inc.

Leonard **Susskind**, (2008), *The Black Hole War: My Battle with Stephen Hawking to Make the World safe for Quantum Mechanics*, (New York: Little, Brown)

N. J. T **Thomas**, (1999), “Are Theories of Imagery Theories of Imagination? An Active Perception Approach to Conscious Mental Content” in *Cognitive Science* Vol 23 (2), pp.207-245

Evan **Thompson**, (2001), “Empathy and Consciousness”, in *Journal of Consciousness Studies* 8, No. 5-7, pp.1-32, extracted www.imprint.co.uk/pdf/Thompson.pdf

Michael **Tye**, (2000), “Knowing What it is Like: The Ability Hypothesis and the Knowledge Argument,” in M. Tye, (2000), *Consciousness, Color, and Content*, (Cambridge, Massachusetts: The MIT Press), pp.3-20

Francisco **Varela**, 1996, “Neurophenomenology: A Methodological Remedy for the Hard Problem” in *Journal of Consciousness Studies*, 3, No. 4, pp.330-49

Francisco **Valera** and Jonathon **Shear**, (1999), “First-Person Methodologies: What, Why, How?” in *Journal of Consciousness Studies*, 6, No.2-3, 1999, pp.1-14

Bruno **Wicker**, Chrisitan **Keysers**, Jane **Plailly**, Jean-Pierre **Royet**, Vittorio **Gallese**, and Giacomo **Rizzolatti**, (2003), “Both of Us Disgusted in *My* Insula: The Common Neural Basis of Seeing and Feeling Disgust”, in *Neuron*, Vol. 40, October 30, 2003, pp.655-664

Ludwig **Wittgenstein**, (1963), *Philosophical Investigations*, (Tans.), G. E. M Anscombe, (Oxford: Blackwell)

Ludwig **Wittgenstein**, (1968), “Notes for Lectures on ‘Private Experience’ and ‘Sense Data’”, (1968), in (Eds.) Rush Rhees, (1968), *Philosophical Review*, 3, July LXXVII, pp.275-320

Stephen **Yablo**, (1992), “Mental Causation”, in *Philosophical Review*, 101, pp.245-80

Stephen **Yablo**, (1993), “Is Conceivability a Guide to Possibility”, in *Philosophy and Phenomenological Research*, (53) 1, pp.1-42

Stephen **Yablo**, (2000), “Textbook Kripkeanism and the Open Texture of Concepts”, in *Pacific Philosophical Quarterly*, Volume 81, Issue 1, March 2000, pp.98–122⁴⁰⁸

⁴⁰⁸ “Water” is the answer to the quiz question earlier. Here is the wiki cite with information about a hoax related to water under the description used above:
http://en.wikipedia.org/wiki/Dihydrogen_monoxide_hoax