



Title	Zombies and their possibilities
Authors(s)	Egan, Ruth P. K.
Publication date	2003
Publication information	Egan, Ruth P. K. "Zombies and Their Possibilities." University College Dublin. Department of Philosophy, 2003.
Publisher	University College Dublin. Department of Philosophy
Item record/more information	http://hdl.handle.net/10197/9553
Publisher's version (DOI)	http://hdl.handle.net/10197/9553

Downloaded 2023-03-15T17:09:45Z

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)



© Some rights reserved. For more information

Zombies and Their Possibilities

by

Ruth Egan

The thesis is submitted to University College Dublin for the degree of MLitt
in the Faculty of Arts.

November 2003

Department of Philosophy

With the acknowledgement of the Irish Research Council for the Humanities
and Social Sciences

Head of Department: Dr. Gerard Casey

Supervisor: Dr. Jim O'Shea

Acknowledgements

I wish to acknowledge the support for this research of a Government of Ireland Research Scholarship administered by the Irish Research Council for the Humanities and Social Sciences.

Abstract

This thesis is a critical examination of the basis of some arguments in contemporary philosophy of mind against a materialist view of phenomenal consciousness, as proposed by David Chalmers (1996) in his book *The Conscious Mind*. I address Chalmers' "zombie" argument in particular, disputing the soundness of the argument itself and its basis, and examining some of the salient concepts involved.

I argue that logical possibility claims only carry as much weight as the background framework against which the claim was made. I propose therefore that Chalmers only succeeds in showing the epistemic possibility of zombies (i.e. they only seem logically possible given our current ignorance in the area) and this, I contend, is not strong enough to refute materialist claims with respect to consciousness. In addition I try to show that he does not adequately answer objections to his argument from *a posteriori* considerations since I argue that logical entailment of a given phenomenon by its (physical) basis is generally something that only begins to emerge during the process of discovery of what that phenomenon is *a posteriori*.

I explore Chalmers' notion of a zombie and propose that it suffers from a basic incoherence which arguably places a question mark over its logical possibility. I also query Chalmers' claim that the essence of phenomenal consciousness is not explainable in terms of function/structure and, consequently, in physical terms. I suggest that by analysing our mental life into phenomenal and psychological aspects whereby the latter is associated with mental functioning, Chalmers already prejudices the question of whether there could be a function of phenomenal consciousness. Arguably experience may be essential for our kind of functioning and may be at least partially so explainable.

TABLE OF CONTENTS

INTRODUCTION	1
CHAPTER 1: THE ZOMBIE ARGUMENT	5
THE “HARD” PROBLEM	5
SUPERVENIENCE	10
REDUCTIVE EXPLANATION	12
INTENSIONS	16
CHALMERS’ ZOMBIE ARGUMENT.....	23
CHAPTER 2: POSSIBILITY, NECESSITY AND THE <i>A PRIORI</i>	28
CONCEIVABILITY	30
POSSIBILITY	36
Physical Possibility	36
Logical Possibility.....	38
Epistemic Possibility.....	49
Metaphysical Possibility	51
A POSTERIORI NECESSITY	55
A PRIORITY, NECESSITY AND INTENSIONS.....	59
CHALMERS VS. KRIPKE	66
TRUTH IN VIRTUE OF MEANING REGARDING THE <i>A PRIORI</i> ASPECT OF A CONCEPT	68
EXPLANATION	80
CHAPTER 3: PHENOMENAL CONSCIOUSNESS	93
PHENOMENAL VS PSYCHOLOGICAL CONSCIOUSNESS	96
AWARENESS	104
ZOMBIES	114
FUNCTIONS FOR PHENOMENAL CONSCIOUSNESS?.....	131
CONCLUSIONS	144

INTRODUCTION

Contemporary philosophy of mind is currently embroiled in much debate about the nature of consciousness – that is, of our subjective experience. Some philosophers think that consciousness can ultimately be explained in terms of the physical and that, as J.J.C. Smart put it in “Sensations and Brain Processes” in 1959: “Sensations are nothing over and above brain processes” (Smart, 1995, p. 95). Others are less convinced that a physical explanation would ever be enough to fully capture the qualitative nature of our experience. Many of these latter philosophers have proposed strong arguments to show this: for example the zombie argument; the inverted spectrum argument; and the knowledge argument - as exemplified by Jackson’s (1982) example of neuroscientist Mary brought up in a black and white room who sees colour for the first time, or Nagel’s (1974) query of what it is like to be a bat.

In his book *The Conscious Mind*, David Chalmers (1996) has taken a number of these “classical” arguments against materialism in philosophy of mind and presented them from a new angle in terms of the *a priori* and of logical necessity. In this manner he manages to make a comprehensive and credible case against the materialist or physicalist view of consciousness. Believing himself to have successfully defeated materialism, he then proposes his own theory of “naturalistic dualism” to explain consciousness by means of non-physical phenomenal properties. However, although an impressive assault on the materialist account, I do not believe that his arguments ultimately succeed. In this thesis I wish to examine the reasons why in my opinion Chalmers’ case against a materialist view of consciousness fails. I will concentrate on his zombie argument in particular for he puts the most effort into developing this one. He does note that each of the five arguments he presents stands on its own, but that together they make a very strong case. However, I intend to query some aspects of the basis to all his arguments.

Although Chalmers presents a variety of arguments against a materialist view of phenomenal consciousness in his book, he bases them all on the same foundations. In chapter one of my thesis I try to clarify and explain the basis of these arguments, concentrating on the zombie argument. I examine his notion of supervenience, the different types involved and his formulation of the “doctrine of materialism” in terms of logical supervenience. I investigate his notion of the primary and secondary intensions of a concept, which he proposes correspond to the *a priori* and *a posteriori* aspects of the concept respectively. I look at why he frames his arguments in terms of the *a priori* aspects of the concepts concerned and at his notion of what a reductive explanation entails. I then put forward a clarification of the form his zombie argument takes in terms of its premises and conclusions.

In chapter two I try to show that a number of claims and arguments Chalmers uses in building up his main arguments (concentrating on his zombie argument) against a materialist view of consciousness do not succeed. Since Chalmers’ zombie argument is essentially a modal one based on a claim of logical possibility, I initially examine various types of possibility – physical, logical, epistemic and metaphysical – and Chalmers’ views of them. I try to show that logical possibility judgements depend on the degree of knowledge that we have with regard to the phenomenon in question, and in cases where we know very little about the relevant phenomenon (e.g. how/if experience is entailed by the physical), such judgements are highly fallible. I argue that Chalmers does not succeed in demonstrating the logical possibility of zombies but merely their epistemic possibility (i.e. it only seems logically possible given our current ignorance in the area) and this, I believe, is not strong enough to refute materialist claims with respect to consciousness.

As Chalmers propounds his arguments in terms of the *a priori* aspects of the concepts involved, I dispute his claim that the *a priori* aspect of a concept constitutes a

necessary *a priori* truth in relation to that concept. I argue that this *a priori* aspect of a concept seems to be formed by contingent properties of the concept's referent and to be revisable in the light of further empirical discoveries and thus does not seem to be a candidate for forming a necessary truth. If the *a priori* aspect of a concept does not back a logically necessary truth (truth in virtue of meaning) then arguably analysis at the level of primary intensions (*a priori*) carries no logical entailment. I try to show further that Chalmers does not adequately answer objections to his argument from *a posteriori* considerations by questioning his claims that it is the *a priori* realm which is the most important when it comes to reductive explanation. I argue that logical entailment of a given phenomenon by its (physical) basis is not something we can know only *a priori* but rather that it is during the process of the *a posteriori* discovery of what a phenomenon is that we begin to see the conceptual connections. Thus I suggest that Chalmers is demanding too much of a materialist view of consciousness.

Having concentrated on the soundness of Chalmers' zombie argument against materialism in the first two chapters of the thesis, in chapter three I explore the concept of a zombie and of phenomenal consciousness in general, both in Chalmers and in other philosophers. I clarify Chalmers' notion of phenomenal consciousness and query his analysis of mind into a psychological, or functional, aspect and a phenomenal aspect. To this end I compare Chalmers' explanation of mental functioning in terms of the psychological aspect of mind with a similar division made by Ned Block. By criticising Block's analysis of a functional non-phenomenal aspect of consciousness I try to show that the phenomenal cannot be so obviously separated from the functional as philosophers who make such divisions propose. Similarly I argue, Chalmers' division along the same lines is questionable. I also query the coherence of Chalmers' notion of a zombie which leads to some seemingly bizarre consequences. In the last section of this chapter I look at some plausible functions of phenomenal consciousness. Even if

the essence of phenomenal consciousness seems to be qualitative this does not preclude it from being to some extent at least, explainable in functional or structural terms, and arguably exploring any possible function of the phenomenal will give us better insight into its nature. Analysing phenomenal consciousness in such a way as to divorce it from any functional role in one's mental life, as I argue Chalmers does, would prematurely cut off this avenue of exploration.

In the conclusion I will summarise and highlight some of my main arguments in the thesis.

CHAPTER 1: THE ZOMBIE ARGUMENT

In his influential book *The Conscious Mind*, David Chalmers (1996) puts forward a number of arguments against a materialist explanation of phenomenal consciousness. The arguments he proposes are some of the “classical” arguments used against a materialist view of consciousness in philosophy of mind: the zombie argument, the inverted spectrum argument, and the knowledge argument, to name a few. However, Chalmers manages to present these arguments anew in a very strong light. In this thesis I wish to concentrate on Chalmers’ version of the zombie argument against a materialist view of phenomenal consciousness. Thus this first chapter of my thesis will be one of clarification where I will try to illuminate various aspects of and background to Chalmers’ zombie argument. I will attempt to explain and elucidate the problem of consciousness as Chalmers sees it; the notions of which he makes use in setting up the background for his arguments (e.g. supervenience, reductive explanation, intensions); and the zombie argument as he proposes it in his book.

First of all I would like to take a look at what Chalmers calls the “hard” problem of consciousness, namely explaining why it is that all the processing that goes on in our minds is accompanied by an experienced inner life.

THE “HARD” PROBLEM

Chalmers distinguishes between what he terms the “easy” and “hard” problems of consciousness. He claims that we can conceptually differentiate between two different aspects of mind – the psychological and the phenomenal aspects – and that these form the basis of the problems. The “easy” problem of consciousness he claims involves such questions as: “How does the brain process environmental stimulation? How does it integrate information? How do we produce reports on internal states?” (p.

xi-xii)¹. Chalmers associates such “easy” problems with what he terms the “*psychological* concept of mind” and says that this is “the concept of mind as the causal or explanatory basis for behaviour” (p. 11). Explaining *this* aspect of mind, although a monumental task, is relatively easy in the sense that it is possible *in principle*, according to Chalmers. As he puts it, there are “no deep metaphysical enigmas” (p. 24) in relation to the psychological aspect of mind.

The “hard” problem of consciousness on the other hand, involves explaining the “experienced inner life” we have, according to Chalmers. He calls this experiential aspect of mind the “phenomenal concept of mind” and says that it is the “concept of mind as conscious experience” and that it is characterised by “the way it *feels*” (ibid.). Without wishing to beg any questions as to whether these two aspects of mind “will turn out to be the same thing” (p. 12), Chalmers argues that we can at least *conceptually* distinguish between the two. He summarises this conceptual difference as “what it *means* for a state to be phenomenal is for it to feel a certain way, and what it means for a state to be psychological is for it to play an appropriate causal role” (p. 12).

Let us take a closer look at Chalmers’ idea of the psychological “aspect” or concept of mind. Clearly this aspect of mind is strongly linked to a functionalist model of the mental since it is primarily concerned with causal roles of mental processes and resulting behaviour. According to the functionalist view, as Chalmers summarises, “a mental state is defined wholly by its *causal role*” (p. 14). He gives a brief sketch of the development of psychology: from the early introspective phenomenal days; through the discovery of the “unconscious” by Freud; to the “behavioural dispositions” of the behaviourists and consequent irrelevance of internal (mental) states; and then finally to functionalism with its causal role for mental states. Chalmers points out that Freud’s discovery of unconscious desires (leading to particular behaviour) showed that

¹ All page numbers in brackets without the author’s name refer to Chalmers’ (1996) book *The Conscious Mind*.

“accessibility to consciousness is not essential to a [mental] state’s relevance in the explanation of behaviour” implying a notion of mentality that is “independent of phenomenal notions” (p. 13). Similarly, Chalmers claims, the behaviourists, by analysing the mental in terms of “external” factors such as behaviour, brought about a “shift in emphasis from the phenomenal to the psychological” (p. 14) in philosophy. He argues that “these two developments established as orthodoxy the idea that explanation of behaviour is in no way dependent on phenomenal notions” (ibid.). This is an orthodoxy which he suggests is still alive in philosophy of psychology today.

Even though functionalism tries to account for the mental state, according to Chalmers, it is not the phenomenal aspect of the state that is important here but rather its causal role. Mental activities such as learning, he claims, can be explained thus in terms of their function and causal roles in our cognitive systems: the phenomenal aspect to learning (what it feels like to learn something) is not essential to explaining learning as a cognitive ability.

So Chalmers sees that great headway has been made in explaining mental phenomena in terms of “functional properties, characterized by causal roles” (p. 24). He believes that this allows a relatively straightforward route to arriving at a physical explanation of the psychological aspect of mental states: “so the question ‘How could a physical system have a psychological property *P*?’ comes to the same thing as ‘How could a state of a physical system play such-and-such a causal role?’” (p. 24). This is a question that he believes can *in principle* be answered by examining the organisational set-up of the physical system, its interaction with its environment and its resulting behaviour. Such an investigation is, he admits, a huge undertaking and is by no means anywhere near completed. Nevertheless he sees it as, theoretically at least, entirely possible: “There is no great mystery about how a state might play some causal role, although there are certainly technical problems there for science.” (p. 15). For example,

he suggests that learning is largely psychological (as opposed to phenomenal) and can be essentially explained in functional terms as (roughly) adaptation of an organism's cognitive capacities "in a certain way to various new circumstances and stimuli" (p. 18). Arriving at a suitable explanation of the underlying processes may not be easy, but Chalmers argues that in "explaining learning, the central thing we have to explain is how the system manages to adapt in the appropriate way" (p. 19).

Once we have such a functional analysis in the form of causal roles and resulting behaviour, he believes, we have something that can be explained in physical terms. This is because, he argues, such a functional analysis allows a reductive explanation of the phenomenon in question, that is, an explanation "wholly in terms of simpler entities" (p. 42). With such a functional analysis of learning in hand, all we have to do (roughly) is explain in physical terms, perhaps neurophysiological terms, how certain brain states/neural activity/etc., bring about the organism's behavioural adaptation to environmental change. In this manner, he proposes, if we give "an appropriate account of lower-level processes, an explanation of the higher-level phenomenon falls out" (ibid.). So he argues, in principle at least, the psychological aspect of mind does seem to be explainable in terms of the physical.

However, Chalmers points out that mental states have to a greater or lesser extent phenomenal aspects to them as well, i.e. what it feels like to be in that state. He gives many examples of such phenomenal "feels" or conscious experience, ranging from bodily sensations through to conscious thought and sense of self (pp. 7-10). With regard to such experiences, e.g. seeing colour, hearing a sound, etc., he asks: "Why should it feel like *that*? Why should it feel like anything at all?" (p. 7). Although he initially claimed that there was a *conceptual* distinction between the two notions of the psychological and phenomenal aspects of mind, Chalmers soon comes to the much stronger conclusion that "both the psychological and the phenomenal are *real and*

distinct aspects of mind” (p. 16, my italics). He argues that it is this phenomenal aspect or “feel” of mental states that does not seem to be adequately captured in the functionalist or cognitive science descriptions of mind which, he claims, have not “shed significant light on the question of how and why cognitive functioning is accompanied by conscious experience” (p. 25). According to Chalmers this is because “the progress in the understanding of the mind has almost entirely centred on the explanation of behaviour” (ibid.). The problem is that this phenomenal “feel” of mental states does not seem to be central to any functional explanation of the mental states, for he claims it is “a conceptually coherent possibility that something could be playing the causal role without there being an associated experience” (p. 15). So he says the question remains of how the physical could give rise to the phenomenal “feel” of our mental life, our conscious experience.

Since he believes we theoretically at least can get an explanation of “how a physical system can have psychological properties” (p. 25), we have “dissolved” the “psychological” mind-body problem. All that remains is the question of how and why phenomenal properties accompany the psychological ones, i.e. the “mind-mind” problem². Explaining this phenomenal aspect of our mental life is exactly the “hard” problem of consciousness that Chalmers wishes to address in his book and he puts forward a number of arguments to show why he thinks this problem cannot be solved by materialist appeals to the physical to explain the phenomenal aspects of consciousness (subjective experience). Before I examine the particular argument I wish to discuss, his zombie argument, in the following sections I will look at some of the key notions of which Chalmers makes use in constructing his arguments.

² Chalmers credits this phrase to Jackendoff.

SUPERVENIENCE

In order to clarify what he means by a reductive explanation of phenomenal consciousness (in terms of the physical), Chalmers introduces the concept of “supervenience”. Supervenience of one set of (generally high-level) properties on another set of (generally low-level) properties means that the latter set fully determine the former. He describes the concept in terms of “A-properties” and “B-properties” which correspond to the lower-level properties and higher-level properties respectively. By lower-level properties or A-properties, Chalmers says that he will usually mean physical properties: “more precisely, the fundamental properties that are invoked by a completed theory of physics” (p. 33). Thus he defines supervenience as follows: “B-properties *supervene* on A-properties if no two possible situations are identical with respect to their A-properties while differing in their B-properties” (ibid.). David Lewis (1986) puts it somewhat similarly: “Supervenience means that there *could* be no difference of one sort [e.g. B-properties] without difference of the other sort [A-properties]” (Lewis, p. 15). Chalmers also talks in terms of “facts” rather than “properties” as a kind of shorthand, for he claims that facts are just instantiations of properties and can “ultimately be cashed out in terms of patterns of co-instantiation of properties” (note 2, p. 33).

Chalmers describes various different kinds of supervenience: local versus global supervenience and logical versus natural supervenience. Local supervenience involves individuals and occurs “if ... any two possible individuals that instantiate the same A-properties instantiate the same B-properties” (pp. 33-34). Global supervenience, according to Chalmers, involves whole worlds or universes. He defines it as follows: “B-properties *supervene globally* on A-properties...if there are no two possible worlds identical with respect to their A-properties, but differing with respect to their B-properties” (p. 34). More important for the basis of his main arguments is the distinction

between natural and logical supervenience. Of logical supervenience he says “B-properties supervene *logically* on A-properties if no two *logically possible* situations are identical with respect to their A-properties but distinct with respect to their B-properties” (p. 35). I will examine Chalmers’ notion of logical possibility in much more detail in the next chapter, but for the purposes of this explication suffice to say that he defines it as logical possibility in the broadest sense. He notes that “logical supervenience is defined in terms of logically possible *worlds*” and says that in “determining whether it is logically possible that some statement is true, the constraints are largely *conceptual*” (ibid.). He also describes logical supervenience in terms of “facts”, saying that generally when B-properties supervene logically on A-properties this is another way of saying that the A-facts *entail* the B-facts. In other words, “*all there is* to the B-facts being as they are is that the A-facts are as they are” (p. 36).

Natural supervenience, on the other hand, “arises when two sets of properties are systematically and perfectly *correlated* in the natural world” (p. 36). The correlation involved although contingent is nevertheless lawful, according to Chalmers. Thus with regard to cases of natural supervenience only, there is no logical necessity or entailment, just natural necessity. This happens, he explains, when “the same clusters of A-properties in our world are always accompanied by the same B-properties, and when this correlation is not just coincidental but *lawful*” (p. 37). Thus logical supervenience of B-properties on A-properties requires (logical) entailment of the former by the latter, whereas natural supervenience does not require such entailment, merely a contingent lawfulness.

The kind of supervenience in which Chalmers is interested for his discussion of the materialist position in philosophy of mind is *global logical* supervenience. Thus he arrives at a formulation of the doctrine of materialism regarding consciousness against which he is going to argue. He says “materialism is true if all the positive facts about

the world are globally logically supervenient on the physical facts” (p. 41).³ In other words, materialism is true if it is the case that once the physical facts (A-facts) of the world are given (globally) the high-level facts (B-facts) too are “automatically” fixed – i.e., the “B-facts merely redescribe what is described by the A-facts” (ibid.). By using *logical* instead of natural supervenience in the above formulation, Chalmers is saying that the materialist claim is that the lower-level physical facts about the world *entail* the higher-level facts (of consciousness, etc.) about the world - as opposed to the latter merely being correlated with the former. Chalmers argues that “if all that is required for materialism is that all facts be lawfully [but contingently] connected to the physical facts, then materialism becomes a weak doctrine indeed” (p. 126). Presumably this is because mere natural supervenience of higher-level facts on lower-level facts is not enough to support a reductive explanation of consciousness by the physical. But I will look more closely at Chalmers’ notion of reductive explanation in the following section.

REDUCTIVE EXPLANATION

Indeed Chalmers argues that “reductive explanation requires a logical supervenience relation” because it “requires some kind of analysis of the phenomenon in question, where the low-level facts imply the realization of the analysis” (p. 48). However, he is careful to stress that even if high-level facts may be explained in terms of low-level ones this does not mean that the former have to be *identified* with the latter. That is, a reductive explanation of a (high-level) phenomenon does not require that the phenomenon itself be completely “reduced away” as it were. Returning to the example of learning, he says that a complex higher-level phenomenon such as learning could be realised in various physical ways as it is, he indicates, primarily a function. Learning, he says, “might not be reducible in that we cannot *identify* learning with any specific lower-

³ Chalmers notes (pp. 38-41) that there is a small caveat with regard to “negative existence claims”, hence his use of the term “positive facts” in the above description of what materialism entails. Since he regards this as only a minor complication I will not pursue the matter further.

level phenomenon. But this multiple realizability does not stand in the way of reductively *explaining* any instance of learning in terms of lower-level phenomena” (p. 43). So he is not supposing that the materialist claim is that consciousness should be able to be reduced in a very strict way to low-level facts only, but rather that the low-level (physical) facts should be able to explain how consciousness occurs. The important point about reductive explanation, according to Chalmers, is that once “we have told the lower-level story in enough detail, any sense of fundamental mystery goes away: the phenomena that needed to be explained have been explained” (p. 42).

Of course he also sees that the (reductive) explanation need not always be in terms of the lowest-level properties possible, as he says “a practical reductive explanation of a phenomenon does not usually go all the way to the microphysical level” (p. 50). Instead, he says, “high-level phenomena are [usually] explained in terms of some properties at a slightly more basic level” (ibid.), so that “biological phenomena may be explainable in terms of cellular phenomena, which are explainable in terms of biochemical phenomena” (p. 51) and so on down to the level of physical phenomena. Thus he concludes that even if it turns out not to be *practically* feasible, nevertheless, given “logical supervenience, along with the simplicity and autonomy of the lowest level, this sort of explanatory connection between the sciences ought to be possible *in principle*” (p. 51, my italics). Such a view is of course open to dispute: both the notion that each level of explanation can be explained in terms of a slightly lower level and so on all the way down to the microphysical; and the notion of physics as a basic science to which all others more or less reduce.⁴ In his defence, Chalmers explains that he is “not suggesting that high-level facts and laws are entailed by microphysical *laws*”, but is rather making the weaker claim that they are “entailed by all the microphysical *facts* (perhaps along with microphysical laws)” (p. 71).

⁴ See J.A. Fodor (1974) on the disunity of science and Tim Crane and D.H Mellor (1990) on the difficulties with physicalism. I will not dispute Chalmers’ notion of reductive explanation as such, but will assume a roughly similar one.

Chalmers believes that it is because we *can* functionally analyse most of what he calls our psychological mental concepts (and indeed he believes, theoretically, most other higher-level phenomena in the world), that we may conclude that these phenomena can in principle be reductively explained. Once we have defined the function of a psychological state, such as learning, then “if we can explain how certain neurophysiological states are responsible for the performance of the functions in question, then we have explained the psychological state” (p. 46). He thinks a functional account of, say, learning explains it fully because that is *all* there is to learning: “Because all it *means* to learn is to function like this” (p. 47). He does allow that consciousness may play some role, saying that the functional account is all there is to learning “except perhaps insofar as learning requires consciousness” (ibid.). However, he suggests the concept of learning is “largely psychological [functional]” (p. 18).

In the case of (phenomenal) consciousness, however, we have no functional analysis of this concept. In fact, according to Chalmers, we have no good analysis at all of consciousness, functional or otherwise, and therefore cannot explain it reductively. This is because, he claims, the phenomenal is essentially qualitative and therefore cannot be explained in terms of function or structure, neither of which he thinks tell us anything about the quality. As he puts it:

consciousness is surprising. If all we knew about were the facts of physics, and even the facts about dynamics and information processing in complex systems, there would be no compelling reason to postulate the existence of conscious experience (p. 5).

Thus, he argues that the *essence* of phenomenal consciousness (subjective experience), its particular qualitative “feel”, cannot be captured by functional or structural descriptions. Chalmers maintains that, with regard to natural phenomena, “most [high-level] properties supervene logically on physical properties” (p. 71). So we can generally (roughly) analyse most other high-level concepts in functional or structural

terms, he claims, and it is “in virtue of this analyzability that high-level facts are in principle derivable from microphysical facts and reductively explainable in terms of physical facts” (p. 81). In other words he is postulating that explanation of a phenomenon in terms of the physical requires some kind of analysis of that phenomenon in terms of function and/or structure, and he maintains that such terms cannot explain the essentially qualitative aspect of our mental life. Hence his conclusion that phenomenal consciousness cannot be explained in terms of the physical.

So, with regard to a psychological concept like learning, once we have a functional account of it, he claims, it is “simply *logically impossible* that something could instantiate that account without learning” (p. 47), because this is what learning *is*: this is what we have found it to be through analysis, therefore it is true by definition. But no matter what functional account we have of any psychological state or cognition, it *does* seem logically possible to Chalmers that “that account could be instantiated without any accompanying consciousness” (ibid.). Again, this is because the essence of (phenomenal) consciousness is not explained by any functional description of cognitive states – its essentially qualitative aspect, he maintains, cannot be defined in terms of function or causal role. Therefore it does not follow *by definition* that consciousness should “accompany” the state. Thus, he argues, it is logically possible that consciousness does not accompany the cognitive state in question.

In fact, this emphasis on logical supervenience, entailment and logical possibility outlined above seems to suggest that Chalmers sees the (correct) functional analysis of a given higher-level phenomenon as logically necessary of the related concept. As he says himself: “In my account of supervenience and explanation, I have relied heavily on the notions of logical possibility and necessity” (p. 52). He relates logical necessity to “truth in virtue of meaning” but notes that it is important to be careful in interpreting the term “meaning”. He proposes to understand meaning in terms of *intension*. In the next

section I will try to clarify Chalmers' notion of intensions and how this notion is used in framing his arguments against the possibility of a reductive explanation of phenomenal consciousness.

INTENSIONS

Since Chalmers has argued that materialism involves *logical* supervenience, then, in order to defeat an argument for a materialist view of phenomenal consciousness all he believes he has to do is to show that consciousness is not *logically* entailed by the physical. And this, he claims, can be shown in an *a priori* manner. One of the principal ways he in which he brings out the *a priori* aspect of the concepts involved is by means of his analysis of the meaning of a concept in terms of a "two-dimensional framework" of intensions.

Prior to introducing the notion of intension, Chalmers explains logical necessity, saying that the "basic way to understand the logical necessity of a statement is in terms of its truth across all logically possible worlds" (p. 52). As I have mentioned above, he also describes the logical necessity of a statement as "truth in virtue of meaning" (ibid.). According to Chalmers these definitions involve the notion of *conceptual truth*, which he defines as "the notion that some statements are true or false simply by virtue of the meanings of the terms involved" (ibid.). In analysing the meanings of concepts, Chalmers notes the frequent absence of a clear definition for many concepts but claims that "the sort of 'meaning' of a concept that is relevant in most cases is not a definition, but an *intension*", which he describes as "a function specifying how the concept applies to different situations" (p. 54).

Chalmers relates intensions to the Fregean "sense" of a concept: "In Frege's own view, every concept had a *sense*, which was supposed to determine the reference of the concept depending on the state of the world; so these senses correspond closely to

intensions” (p. 56). However he points out that *two* intensions are needed to reflect what he claims (building on well-known distinctions from Kripke and Putnam) are “two quite distinct patterns of dependence of the referent of a concept on the state of the world” (p. 57). These two patterns are: the way reference is fixed in the actual world; and the way reference is determined in possible worlds, given that it is already fixed in the actual world. Thus he defines the *primary intension* of a concept as “a function from worlds to extensions reflecting the way that actual-world reference is fixed” (p. 57). The *secondary intension*, he says, corresponds to the “dependence by which reference in *counterfactual* worlds is determined, given that reference in the actual world is already fixed” (ibid.). These two intensions form the core of the “two-dimensional framework” which he uses for dealing with meaning and necessity.

According to Chalmers, the primary intension gives the means by which one picks out the referent of a term and is independent of empirical facts and therefore *a priori*. As he points out, the primary intension of a concept “*specifies* how reference depends on the way the external world turns out, so it does not itself depend on the way the external world turns out” (p. 57). So analysis of the primary intension of a concept would appear to be an *a priori* enterprise insofar as it involves questions about that to which our concept *would* refer regarding the various ways the actual world *could* turn out. He gives the example of the concept of water, of which he says its primary intension roughly picks out “the dominant clear, drinkable liquid in the oceans and lakes; or more briefly, ...it picks out the *watery stuff* in a world”⁵ (p. 57). Thus had the actual world turned out to be one with XYZ in its oceans and lakes then, according to Chalmers, “water” would refer to XYZ. However, “water” turns out to refer to H₂O in the actual world, so he says “the primary intension of ‘water’ maps the XYZ world to XYZ, and the H₂O world to H₂O” (ibid.). Thus the primary intension of water, watery

⁵ Of course he notes that such a brief characterisation is a major simplification. I will follow him in this and use “watery stuff” to abbreviate the primary intension of “water”.

stuff, tells us what the term “water” would refer to in various different worlds considered as actual and “we have the ability to engage in this reasoning independently of how the world turns out” (p. 58), according to Chalmers.

The secondary intension of a concept, on the other hand, picks out that to which a concept, after empirical investigation in the actual world, is *discovered* to be referring. It is determined empirically and is therefore *a posteriori*. As Chalmers describes it (following Kripke on rigid designation), “the secondary intension is determined by first evaluating the primary intension at the actual world, and then *rigidifying* this evaluation so that the same sort of thing is picked out in all possible worlds” (p. 59). Thus the secondary intension of a concept becomes a rigid designator (i.e., picks out the same referent across all possible worlds) for that concept. In the case of the concept of water, its referent in the *actual* world turns out to be H₂O, so its secondary intension is H₂O in *every* possible world.

But obviously the primary intension of a concept *also* depends on how the actual world is, insofar as we must first encounter water in the actual world to form any concept of it at all. The primary intension of water as Chalmers explains it is related to characteristics of water that we *discover* in the actual world: clear, liquid, drinkable, etc. As the primary intension is dependent on such empirical information for its formation, he cannot mean that it is *a priori* in the Kantian sense of “pure” *a priori*. So it would appear that Chalmers’ notion of *a priori* in this case concerns a *relative* sense of *a priori*. The primary intension involves our initial notion of what a concept is in the world (watery stuff) *before* we know what its underlying nature is (H₂O). So it is *a priori* presumably in the sense that it does not depend on what that underlying nature (secondary intension) turns out to be. At least this seems to be what Chalmers means when he says that the primary intension is independent of the way the world “turns out”. He does however add that matters relating to evaluation of a concept by its primary

intension are “in principle accessible from the armchair” (p. 68) which would seem to imply more a sense of “pure” *a priori*. However, given the empirical means of formation of the primary intension, one can only conclude that he means *a priori* in a *relative* sense: that is, *a priori* relative to evaluation by the secondary intension (*a posteriori* discovery of underlying nature).

Chalmers equates both intensions with meaning and suggests that they can be regarded as “*a priori* and *a posteriori* aspects of meaning, respectively” (p. 62). Referring to the water example, this suggests that the (relative) *a priori* aspect of the meaning of water is its primary intension, watery stuff, and that the *a posteriori* aspect of its meaning is its secondary intension, H₂O. On the basis of these two aspects of meaning of a concept, he notes that each intension, primary and secondary, “backs” a different kind of conceptual truth (“truth in virtue of meaning”). The primary intension is related to *a priori* truths and the secondary to *a posteriori* truths. Both, he claims, “qualify as truths in virtue of meaning; they are simply true in virtue of different aspects of meaning” (p. 62). He also links such truths to necessity, maintaining that both are “varieties of *necessary* truth” (p. 63). Chalmers connects the *a posteriori* necessary truth of a concept evaluated by its secondary intension with Kripkean *a posteriori* necessity.

The necessity associated with *a priori* truths however is “unaffected by *a posteriori* considerations” (p. 63), according to Chalmers. He claims that an *a priori* true statement “will be true no matter how the actual world turns out, although it need not hold in all nonactual possible worlds” (p. 62). There seems to be a contradiction here in that if such a statement (e.g. “water is watery stuff”) will be true no matter what way the actual world turns out, then does that not imply that it is true in *every* possible world? To say that the statement need not hold in all nonactual possible worlds is surely to evaluate it by its secondary intension (i.e. water is H₂O) and thus say that in a possible world where watery stuff is XYZ that the statement “water is watery stuff” does not hold

because of *a posteriori* considerations? However, Chalmers clarifies this by suggesting that we need a “construal on which possible worlds are *considered as actual*” (p. 63), in order to say that “water is watery stuff” is a necessary truth. To this end Chalmers talks of “centered” possible worlds in relation to primary intension evaluations. As he explains:

Such a center is necessary to capture the fact that a term like “water” picks out a different extension for me than for my twin on Twin Earth, despite the fact that we live in the same universe. It is only our position in the universe that differs, and it is this position that makes a relevant difference to the reference-fixing process (p. 60).

So Chalmers appears to be indicating that the necessary truth associated with the primary intension of a concept (e.g. water is watery stuff) will hold in each possible world considered as actual (“centered”) but will not necessarily hold in each possible world considered as counterfactual.

This seems to imply that *before we know* what water is, the statement “water is watery stuff” is a necessary truth, as evaluated according to the primary intensions of the terms. (Or, the way Chalmers puts it, if we consider the given possible world in which the statement is being evaluated as *actual*, then the statement is necessarily true.)

However, once we know what water turns out to be (H₂O) in the actual world then in all counterfactual worlds this is what water is. Therefore considering a possible world where the only watery stuff is XYZ and not H₂O, the supposedly necessarily true *a priori* statement “water is watery stuff” turns out to be false when we consider that water is H₂O in the actual world. So the *a priori* truth, which should be true by definition, it would appear can turn out to be false in a possible world due to *a posteriori* considerations. Such a conclusion seems strange as one would think that “true in virtue of meaning” could never be false, but rather the meaning of the concept would have to be re-evaluated (which presumably would lead to a better determination of meaning thereby allowing “truth in virtue of meaning” again).

However, as Chalmers propounds his two-dimensional intensional framework, it would appear to dissolve many of the problems and ambiguities between *a priori* and *a posteriori* considerations mentioned above. He basically proposes to keep completely separate *a priori* considerations of meaning (and hence of necessity and possibility) from *a posteriori* ones, thus avoiding any contradictions between the two. So, in relation to *a priori* considerations, he claims that a concept can be evaluated according to its primary intensions and that one can speak of what he terms 1-necessity, 1-conceivability and 1-possibility of the concept accordingly, where these terms relate *only* to such *a priori* evaluation of the concept. Similarly, where a concept is evaluated by its secondary intensions, his *a posteriori* terms 2-necessity, 2-conceivability and 2-possibility apply. Chalmers claims that *both* 1-necessity and 2-necessity are *logical* necessities: “we have two varieties of logical necessity of statements, depending on whether we evaluate truth in a possible world according to primary and secondary intensions” (p. 65). Looking at the example of water again, presumably the statement “water is XYZ”, when evaluated according to the *primary* intension of water (watery stuff), is 1-necessary, 1-conceivable and 1-possible. Similarly the statement “water is H₂O”, when evaluated by the *secondary* intension of water, is 2-necessary, 2-conceivable and 2-possible. However, the first statement “water is XYZ”, evaluated by primary intensions, presumably would *not* be 2-necessary, 2-conceivable or 2-possible.

Chalmers equates the two kinds of possibility resulting from this two-dimensional intensional framework with logical and metaphysical possibility as follows:

A statement is “logically possible” ... if it is true in some world when evaluated according to primary intensions [1-possibility]; a statement is “metaphysically possible” if it is true in some world when evaluated according to secondary intensions [2-possibility] (p. 68).

Thus he proposes that 1-possibility is equivalent to logical possibility (of statements) and 2-possibility is equivalent to metaphysical possibility.

Chalmers believes that this two-dimensional framework sorts out the often mixed comparison of terms evaluated using primary intensions with those evaluated using secondary intensions, which can be a source of tension in modal talk. He claims, for example, that the conceivability of a statement is often equated with 1-conceivability but that its possibility is equated with 2-possibility. For example, as discussed above, the statement “water is XYZ”, when evaluated according to the *primary* intension of water, is 1-conceivable, however it is not 2-possible, because the latter involves secondary intensions and when evaluated according to *secondary* intensions, water is H₂O in *every* possible world. So in this case, he says, conceivability does not imply possibility. However he maintains that “it remains the case that 1-conceivability implies 1-possibility, and 2-conceivability implies 2-possibility” (p. 67). In other words, if we remain in the *a priori* “realm” and use the appropriate terms (1-conceivability, 1-possibility, etc.) when evaluating a concept according to its primary intension, and similarly, remain in the *a posteriori* “realm” (2-conceivability, etc.) when evaluating according to secondary intensions, we can presumably avoid the ambiguities and tensions I raised in previous paragraphs. We must simply separate the two aspects and evaluate a concept *either* according to primary intensions *or* according to secondary intensions.

Chalmers says that he will present his arguments against a materialist view of consciousness in terms of the *primary* intensions, or the *a priori* aspect, of the concepts involved. Thus, in his zombie argument, he is not concerned with *a posteriori* considerations since he claims that it is enough to establish the *logical* possibility (1-possibility) of zombies in an *a priori* way (i.e. with respect to primary intensions) to show that materialism is false. I will now examine Chalmers’ zombie argument in the following section.

CHALMERS' ZOMBIE ARGUMENT

Chalmers argues that we can conceive of beings, zombies, physically identical to humans but lacking any form of conscious experience, and that this shows that consciousness is not entailed by the physical. He wants to frame this argument in terms of primary intensions, as he claims “it is the primary intension that enters into reductive explanation, so it is this that we are most concerned with” (p. 78). The secondary intension he argues is less important because, for example in the case of water, it is “precisely in virtue of its satisfying this [primary] intension that we deemed that H₂O was water in the first place” (ibid.). So he expounds his argument in the *a priori* “realm” in terms of 1-conceivability, 1-necessity and 1-possibility. In relation to his zombie argument Chalmers describes conceivability as “a statement is conceivable (or conceivably true) if it is true in some conceivable world” (p. 66) and says this involves “first, the conceivability of a relevant world, and second, the truth of the statement in that world” (p. 67). Essentially he appears to be claiming that in order for a statement to be conceivable it must be logically coherent, as he says further that “every conceivable world is logically possible” (p. 66).

Thus Chalmers argues that a world which is physically identical to ours but, say, biologically different would seem to be inconceivable. For once the physical facts are fixed then, according to Chalmers, the biological facts are automatically fixed as well because they are entailed by the physical facts. He says, “These biological facts are not the sort of thing that can float free of their physical underpinnings even as a conceptual possibility” (p. 73). In other words he is claiming that, given the same set of physical facts, it is logically impossible that the biological facts be other than what they are. However, he maintains this is not the case with consciousness since he claims a “zombie world”, with beings physically identical to humans but lacking any conscious experience, *is* logically possible.

Chalmers describes a zombie as identical to a human being at the most basic physical level and therefore also psychologically and functionally identical – it “will be processing the same sort of information, reacting in a similar way to inputs, with his internal configurations being modified appropriately and with indistinguishable behaviour resulting” (p. 95). The zombie is also able to perceive things, turn its attention to various things and report its internal states. However it has no subjective experience, according to Chalmers: “There is nothing it is like to be a zombie” (ibid.).

He does admit though that the logical possibility of such zombies might not seem at all obvious to some people and that in some ways “an assertion of this logical possibility comes down to a brute intuition” (p. 96). However, he also claims that he can discern no (logical) contradiction in the description of such a creature and that “a certain burden of proof lies on those who claim that a given description is logically *impossible*” (ibid.). He goes on to support his argument by appealing to the consideration of *nonstandard* realisations of our functional (psychological) organisation, such as the example originally given by Block of all the people in China organising themselves so as to realise a functional organisation similar to a human brain. It is much more likely that people could accept that such a “brain”, made up of people simulating neural activity etc., would conceivably *not* have consciousness. So Chalmers makes the same appeal for the logical coherence of his zombie example and argues that “there is no more of a *conceptual* entailment from biochemistry to consciousness than there is from silicon or from a group of homunculi” (p. 97). Thus he concludes that there is a logically possible (1-possible) world where the physical facts are identical with our world but the phenomenal facts are not (they are nonexistent in the zombies). Therefore consciousness fails to logically supervene on the physical, his argument goes, and so materialism is false.

Of course the basis of this argument seems to rest on his earlier claim of a lack of

an analysis for the concept of consciousness. He argues that in order to give some indication of how physical facts might logically account for consciousness we first need to have a good analysis of the notion of consciousness. That is “we should be able to *analyze* what it takes for an entity to satisfy the intension of a high-level concept, at least to a sufficient extent that we can see why those conditions for satisfaction could be satisfied by fixing the physical facts” (p. 77). He explains, “The role of analysis here is simply to characterise the intensions in sufficient detail that the existence of an entailment becomes clear” (p. 78).

However, Chalmers argues, we in fact have no adequate analysis of consciousness at all. He claims that the only “remotely tenable” analysis of consciousness we could have would be a functional or possibly structural one. As I mentioned in earlier sections, he maintains that a functional analysis of the notion of consciousness fails to capture the essence of the experience: “Although conscious states may play various causal roles, they are not *defined* by their causal role” (p. 105). A conscious state, he says, is defined by how it *feels* to be in that state and if such a “phenomenal feel” is not captured in an analysis of consciousness then we have missed the essence of the concept. Similarly, a structural analysis of consciousness does not seem to hold out hope of explanation either. As Chalmers point out: “Whether or not consciousness *is* a biochemical structure, that is not what ‘consciousness’ *means*” (p. 106). Thus if we cannot even properly analyse the concept of consciousness then we cannot begin to give a reductive explanation of the phenomenon. In fact he concludes that the phenomenon of consciousness is “characterizable only in terms of concepts that themselves involve consciousness” (ibid.) and is therefore irreducible. Conversely if consciousness were essentially explainable in terms of function or structure, then it would not seem logically coherent that a being could be functionally identical to a human and yet not have phenomenal consciousness.

I would suggest that Chalmers' zombie argument, in terms of its premises and conclusions, is basically structured as follows:

- P1 the *essence* of phenomenal consciousness or subjective experience lies not in any functional or causal role it may play or in its structure, but in the phenomenal "feel" of what it is like to experience something
- C1 therefore no functional or structural analysis adequately defines subjective experience
- P2 physical facts/phenomena/properties are constituted in such a way that they are completely explainable in terms of their structures and/or functions – i.e. structure and function are the terms of physical explanation
- C2 to show that a phenomenon is entailed/explained by (logically supervenes on) physical facts one must have some kind of functional or structural analysis of the phenomenon in question
- C3 therefore it is not possible to show that phenomenal consciousness is logically entailed by any physical facts (i.e. that it logically supervenes on the physical) – there is no *logical* connection (C1, C2)
- C4 therefore it is *not* logically *impossible* that a being exactly physically identical to a human could nevertheless lack phenomenal consciousness; in other words, it is *not true in virtue of meaning* (of "phenomenal consciousness") that a being exactly physically identical to a human *must* have phenomenal consciousness.
- P3 a materialist theory of consciousness requires that consciousness be entailed by (logically supervene on) the physical
- C5 therefore materialism is false (C3, P3)

There are a number of points to note about the argument in the above form.

Firstly, it seems as if the whole zombie part of the argument can be taken out since conclusion C4 is not really necessary to reach the end conclusion about materialism, C5. Chalmers however would not agree with this, as he clarifies in an article subsequent to his book "Materialism and the Metaphysics of Modality" (1999a): "The direction of support is the other way around: the conceivability of zombies is used as one argument for the failure of logical supervenience" (Chalmers, 1999a, p. 475). I am not sure this is actually the case in his book, but at any rate I will argue against the conceivability of zombies (i.e. examine C4 in the light of P1) in chapter three and against the lack of

logical entailment of the phenomenal by the physical (i.e. C3 & C4 in the light of P3) in chapter two. A second point to note about the argument in the above form, is that it seems as if the notion of logical supervenience can be left out of the argument. I have put the argument in terms of “entailment” and “explanation” (Chalmers also talks in these terms) and have only included “logical supervenience” in brackets. However, I am not going to dispute Chalmers’ notion of logical supervenience and will use it interchangeably with the notion of (logical) entailment as I believe he does in a broad sense.

In this chapter I have tried to elucidate the basis of and background to Chalmers’ zombie argument for the purposes of clarification. In the next chapter I will critically analyse some of the notions of which he makes use in his argument. I will discuss conceivability and possibility, in particular logical possibility and its role in the zombie argument. I will argue against Chalmers’ particular notion of conceptual truth with regard to the *a priori* aspect of a concept and query his view of the role of this *a priori* aspect (primary intension) in reductive explanation.

CHAPTER 2: POSSIBILITY, NECESSITY AND THE A PRIORI

In this chapter I wish to investigate in depth some of the notions and claims behind Chalmers' main argument against a materialist view of phenomenal consciousness, the "zombie" argument. As explained in the last chapter, his argument is that it is logically possible that there exist a world physically identical to ours yet lacking phenomenal consciousness (a "zombie" world). The basis of his claim is that since there is no logical entailment from the physical to the phenomenal then it is logically possible that the former could exist without the latter. Therefore, he concludes, phenomenal consciousness does not logically supervene on the physical and therefore materialism is false.

Of course *a posteriori* considerations could lead one to ask why phenomenal consciousness must logically supervene on the physical and not just metaphysically supervene. However, Chalmers claims that the materialist needs logical supervenience of the phenomenal on the physical in order that materialism be true. And this, arguably, is a reasonable claim because in order to see that the physical gives rise to the phenomenal we have to see *how* this occurs – we need conceptual entailment. But I would argue that Chalmers is putting the cart before the horse, so to speak, since I believe we normally only get such conceptual links once we have at least begun to discover *a posteriori* what a phenomenon *is*. So I propose that what Chalmers is demanding of a materialist theory is unreasonable: he is looking for logical entailment *before* we have the means of getting it, i.e. before we have at least started to discover *a posteriori* what consciousness is.

In what follows I will first briefly look at Chalmers' notion of conceivability and its connection to possibility in the first section of this chapter. I will try to show that Chalmers' notion of (ideal) conceivability is knowledge-dependent and therefore as fallible as any other notion. Then I will investigate various types of possibility,

Chalmers' interpretation of them and their relation to his zombie argument. In this context I will argue that he equivocates on two different types of (or ways of using) logical possibility.

In the three sections that follow the discussion of types of possibility, I will critically examine Chalmers' two-dimensional intensional framework for dealing with the case against his zombie argument based on Kripkean *a posteriori* necessity. I will therefore compare his framework with Kripke's one. In addition I argue against Chalmers' claim that a zombie world is also metaphysically possible, for since he has put his argument in terms of the *a priori*, I propose that consequently according to his own two-dimensional framework, he can only claim logical possibility and not metaphysical possibility of such a world.

Chalmers wants to maintain a link between *a priority* and necessity because he proposes his "zombie" argument in terms of the *a priori* aspects of the concepts involved, arguing that these are more important for reductive explanation than the *a posteriori* aspects. However, in the subsequent section on "truth in virtue of meaning", by critically analysing his notion of conceptual truth with regard to the *a priori* aspect of a concept I will try to show that he does not convincingly manage to demonstrate this link. I argue that the *a priori* aspect of a concept seems to pick out accidental properties of the related phenomenon and is also open to being erroneous. It therefore appears to be only contingently, rather than necessarily, true of the relevant concept.

As a result, I will argue, when it comes to the *a priori* aspect of a concept, Chalmers' notion of logical possibility is really only one of epistemic possibility (i.e. it seems possible because currently we do not know any better) and that in general logical entailment is not found at an *a priori* level alone. I do not dispute his claim that a "natural phenomenon is reductively explainable in terms of some low-level properties precisely when it is logically supervenient on those properties" (pp. 47-8). But I will

argue that we only come to see a phenomenon as logically supervenient on certain low-level facts through at least starting the process of discovering (roughly) those very facts *a posteriori*, a feat which has not yet been accomplished (or even sketched in sufficient detail) with respect to phenomenal consciousness. I will try to show in the last section that it is largely (or at least partly) the *a posteriori* discovery of what a phenomenon is that gives us *conceptual* entailment, and that therefore any claim that there is a lack of *a priori* logical entailment in the concepts involved does not *prove* that a materialist view is false. Arguably the same kind of *a priori* logical entailment was absent in many cases before we had discovered *a posteriori* what the relevant phenomenon actually was.

CONCEIVABILITY

Many “zombie arguments” about consciousness in philosophy of mind are controversial because of differences of opinion regarding the conceivability of such a zombie world. Güven Güzeldere (1995) remarks in his paper “Varieties of Zombiehood” that “[b]elief in zombies has become a litmus test for intuitions in recent philosophy of mind” (Güzeldere, p. 326). In his paper “The Unimagined Preposterousness of Zombies”, Daniel Dennett (1995) complains that “when philosophers claim that zombies are conceivable, they invariably underestimate the task of conception (or imagination), and end up imagining something that violates their own definition” (Dennett, p. 322). So I believe it is worth looking more closely at Chalmers’ notion of conceivability, even though it may be the way the concept is normally used in philosophy of mind, since he actually goes to the trouble of explaining what he means by the term “conceivable”. In fact, Chalmers has since written an article on the subject, “Does Conceivability Entail Possibility?” (2002), defending the notion that conceivability is a guide to possibility.

In *The Conscious Mind*, Chalmers qualifies his own use of the term “conceivable” as compared to some other philosophers’ use of the term. He gives the

example of another sense of “conceivable”, different to his, “according to which a statement is conceivable if for all we know it is true, or if we do not know that it is impossible” (p. 66). Such a sense of the term “conceivable” would presumably be an epistemic notion: that is, if we do not *know* if something is true or not, we conclude that it is a possibility as we know of nothing that precludes it. In contrast, it would appear that Chalmers is proposing a more “ontological” notion of conceivability: roughly, that something is conceivable only if it actually *is* (logically) possible and not just if for all we *know* it is possible. In his book, Chalmers does not explicitly define his notion of conceivability as such but it seems to amount to such as definition.

This impression is reinforced by his comparison of the definition he wishes to use for conceivability with a similar one put forward by Stephen Yablo that Chalmers mentions in his notes. For Chalmers the conceivability of a statement involves the conceivability of a relevant world and the truth of the statement in that world. He claims this definition is related to that of Yablo (1993). However Yablo’s definition of conceivability seems to include an epistemic dimension, according to Chalmers, as he claims that Yablo relates conceivability to imagining a possible world that one “takes to verify” the thing to be conceived (note 32, p. 67). Chalmers argues that this “taking to verify” allows for “misdescription of conceived situations” (ibid.), presumably because “taking to verify” is basically a judgement that can depend on the degree of knowledge one has about a given situation (i.e. is epistemically based) and is therefore fallible. He claims that with the definition of conceivability of which he (Chalmers) makes use “this source of defeasibility is removed” (ibid.). He then says, “[i]t follows that in making conceivability judgements, one has to make sure that one describes the world that one is conceiving correctly” (p. 67).

However it is not at all obvious how such an approach to conceivability should remove a “source of defeasibility” that could be present in the more epistemic

definitions of conceivability such as Yablo's definition apparently is. Despite the utmost care, one could still "misdescribe" a given possible world (e.g., through lack of knowledge of certain relevant facts) and thus incorrectly evaluate the truth of a particular statement at that world. Even though Chalmers seems to be arguing that the definition he uses for conceivability (the conceivability of a relevant world and the truth of the statement/situation in that world) should remove a source of defeasibility which he implies is present in other definitions that rely on epistemic, subjective notions as "taking to verify", it is hard to see how this difference could ever come to light in practice. He gives the example of thinking it conceivable that Goldbach's conjecture were false and claims that if, in fact, Goldbach's conjecture is true, then one is "misdescribing" that (possible) world. However, if we do not *know* whether Goldbach's conjecture is true or not, how do we *know* we are "misdescribing" a possible world in which we take the conjecture to be false? Thus for all intents and purposes the two different senses of the term "conceivability" (Yablo's and Chalmers' own) seem to be indiscernible in practice.

As I mentioned above, Chalmers (2002) has written an article since publication of his book, wherein he proposes and examines various different notions of conceivability. The type of conceivability that he claims is the most relevant as a guide to possibility is what he terms "ideal conceivability".⁶ Of this notion he says that a statement S is "ideally conceivable when S is conceivable on ideal rational reflection" (Chalmers, 2002, p. 147). Of course ideal rationality can be difficult to characterise, but Chalmers suggests that we "simply invoke the notion of undefeatability [sic] by better reasoning" (Chalmers, 2002, p. 148). With respect to the notion of good reasoning, Chalmers takes certain rational notions as primitive and says that he is appealing to "our

⁶ In fact, he proposes ideal primary positive conceivability as the type of conceivability to be used as a guide to possibility, where primary conceivability of S is when "it is conceivable that S is *actually* the case" (Chalmers 2002, p. 157) – i.e. *a priori* evaluation of S. Positive conceivability of S is when "one can imagine a situation that verifies S" (ibid., p. 150).

intuitive grasp of notions of reasoning and of when one reasoning process defeats another” (ibid.). With regard to the notion of undefeatability, he claims that this is “implicit in our concept of knowledge: it is generally held that if one’s justification for a belief that P is defeatable by better reasoning, then one does not know that P” (ibid.). Thus the notion of conceivability is no worse off than the notion of knowledge, according to Chalmers.

It is not quite clear, though, which kind of possibility Chalmers (2002) is referring to when he proposes that what he calls “ideal primary positive conceivability” is the best candidate as a guide to possibility. Initially when he introduces the concept of primary conceivability, he ties this notion in with *epistemic possibility*, claiming that when we “consider situations as actual, we consider and evaluate them in the way that we consider and evaluate epistemic possibilities” (Chalmers, 2002, p. 157). He also ties it to *primary possibility* which he describes as follows: “S is *primarily possible* (or 1-possible) if its primary intension is true in some possible world” (Chalmers, 2002, p. 164). In his book, however, Chalmers ties 1-possibility to logical possibility. So despite the initial connection to epistemic possibility, Chalmers seems to be saying that primary conceivability implies logical possibility.

In two other articles Chalmers (1999a&b) also clarifies some of the above notions of conceivability and possibility. In “Materialism and the Metaphysics of Modality”, referring to his book *The Conscious Mind*, he explains: “On my usage, ‘logically possible’ is near enough to interchangeable with ‘conceptually possible’ ... and is tied by stipulation to [ideal] conceivability” (Chalmers, 1999a, p. 477). Thus he is *stipulating* that conceivability of a statement implies (or, perhaps, ensures) a logically possible world in which that statement is true and *that* is his notion of (ideal) conceivability. Arguably, however, conceivability would imply the more subjective and thus fallible epistemic possibility, rather than logical possibility of a world, for we could

be mistaken in our conceiving. Indeed, with respect to the primary intension of a concept (with which Chalmers is concerned when he talks of ideal *primary* positive conceivability), Chalmers (1999b) in “Précis of *The Conscious Mind*” describes it as delivering “a concept’s referent in a centered world when the world is considered as actual (i.e., considered as an *epistemic possibility*)” (Chalmers, 1999b, p. 435, my italics). Thus ideal primary positive conceivability, which involves the primary intension of the relevant concept, also involves epistemic possibility. So how does this give us a conceivability that is effectively *logical* possibility (of a world)? I believe that for Chalmers the shift from epistemic possibility in evaluating according to primary intensions, to the stronger logical possibility comes from this evaluation being found to be *true* in some world. He says, “A statement is conceivable (or logically possible) when its primary intension is true in some world” (Chalmers, 1999b, p. 435).

So effectively Chalmers is stipulating that conceivability be logical possibility, as he says in his book, “there seems to be no reason to deny that conceivability of a world implies possibility. I will henceforth take this for granted as a claim about logical possibility” (p. 68). While I would agree that conceivability is really just about working out possibilities, it is not at all clear that this leads to *logical* possibility. Chalmers proposes that ideal conceivability involves ideal rational reflection, but it would seem that such a notion is a non-starter when it comes to applying it in practice since none of us are ideally rational. If he wants to say that ideal rational reflection is something we humans can attain then, in principle, *all* of us are (potentially) ideal rational thinkers; but we know that historically most of our rational reflection is far from ideal and is often superseded by better rational thinking, or is improved upon when more of the relevant (empirical) facts are known. If ideal rational reflection (on a more natural reading, in my opinion) is something we can only ever strive towards but never attain, then *none* of us are ever ideally rational thinkers and so the notion becomes inoperable, and we

cannot rely on it for working out possibilities. Thus I would argue that the notion of ideal conceivability upon which Chalmers relies is no better for avoiding “misdcriptions” than other notions such as Yablo’s, and hence no better for determining possibility, because it is either an unobtainable idealisation, or it is just our everyday rational thinking which is notoriously fallible and/or dependent on what empirical facts are known to us.

Even taking Chalmers’ notion of ideal rational reflection as our best possible attempts at rational thinking, or “undefeatability by better reasoning”, it is still not obvious that such reasoning will lead to the establishment of *logical* possibility. Again, the term “undefeatability” is another idealisation – we have no guarantees that any of our (current) reasoning is, or will remain, “undefeatable”. If reasoning regarding a particular conceived world and truth of a statement within that world cannot currently be defeated by better reasoning, this does not *prove* that the situation really is logically possible because we could still be mistaken in our reasoning, only not (currently) know it. In particular since often in philosophy the conceived situations in which we are interested are in areas of which we have little knowledge, then it is arguably safer to conclude that even ideal conceivability leads only to epistemic possibility (i.e. seems possible according to current knowledge) and not any more substantial kind of possibility. Or, if it does lead to logical possibility, such possibility is revisable and thus may be subject to change in the future.

This is not to suggest however that Chalmers’ definition of conceivability is unusual in philosophy. On the contrary, it is probably along the lines of a “standard” one, especially in the area of modal logic. (Although Yablo does identify five other senses of “conceivable”, none of which he claims are what *he* calls “philosophical conceivability” (Yablo, p. 26), yet some of which are used by other philosophers (e.g., Putnam’s sense (Yablo, p. 22)); so apparently even the “standard” philosophical sense of

“conceivable” is anything but standard!) However, if we rely on Chalmers’ (or perhaps *any*) notion of conceivability, I think we are still left with problems as to how we can know if something is genuinely conceivable or not, and that we are not simply “misdescribing” a possible world, especially when we have very little knowledge of what the proposed situation would entail. At any rate, as I have pointed out, Chalmers (1999a) stipulates that conceivability is logical possibility (of a world) and so the burden of the argument is shifted from establishment of conceivability to establishment of logical possibility, thus it is the latter concern that I need to address. In the following section I try to clarify various kinds of possibility and look at Chalmers’ view with regard to each type discussed.

POSSIBILITY

Physical Possibility

Natural possibility, as Chalmers calls it, or physical, or even *actual* or *real*⁷, possibility as it is also known as, is relatively straightforward and uncontroversial. It is what we normally think of as possible, that is, something that could be or could happen according to the laws of nature. Chalmers defines it thus: “A naturally possible situation is one that could actually occur in nature, without violating any natural laws” (p. 36). Although our notion of natural or physical possibility may be fairly uncontroversial, the process of estimating whether something is or is not physically possible would obviously be less so. The reason being that we are often interested (especially in the sciences as well as in philosophy) in estimating possibilities with regard to phenomena at the limit of our knowledge and not with respect to things with which we are familiar, whose possibility we already know. In other words, our estimation of physical possibility is dependent on the degree of our knowledge in the relevant area.

⁷ However some philosophers refer to *metaphysical* possibility as real possibility, as G. Randolph Mayes (1990) in his paper “Ross and Scotus on the Existence of God”, talks of “a sense of possibility one might reasonably call ‘real’ or ‘metaphysical’ possibility” (p. 100).

One would be inclined to think of physical possibility as being bounded, as Chalmers put it, by the natural laws of our (actual) world – that there is a *real* physical boundary to what is physically possible in the actual world. However, in a sense, physical or natural possibility has *two* boundaries: one is the aforementioned actual physical boundary, as determined by the laws of nature (whatever they actually are). The second is determined by what we currently *know* to be physically possible – i.e. what we currently know the natural laws to be. Thus, the former would seem to be a real, physical boundary on what is physically possible and the latter merely an epistemological limit. Obviously we would like to think that in judging whether something is physically possible or not, we are using the former boundary – that is, that our judgements reflect actual physical possibilities. In practice though, our judgements of (physical) possibilities will always be based on our (current) knowledge and thus, on what *seems* to be physically possible to us, given our current level of knowledge of the world. Therefore I would argue that our judgements of physical or natural possibility are always *limited* by our (best) current knowledge.

Of course usually our *knowledge* of what is physically possible and what actually *is* physically possible will coincide as, for example, when an architect or engineer calculates whether a building or bridge of a particular shape, length, etc., would be stable and strong enough to take the forces to be exerted on it. Normally we expect that their estimation of possibility in such cases would be correct (although there certainly have been failures in this respect!). But this is an area where there are few mysteries for us in general – it is fairly well established what kinds of constructs in those sorts of domains are physically possible and which ones are not, and there are a number of tried and tested ways of calculating much of this. Frequently however, we are interested in figuring out the physical possibility of phenomena that are at the edge of our knowledge, so to speak. For example, with new discoveries in science we put forward possibilities

in the form of theories but ideally we want to know which one is correct; i.e., which one is *actually* the case. And it is often in *this* area, at the limits of our knowledge, that there can be wide gaps between what we *think* is (physically) possible and what actually *is* physically possible, something we sometimes discover years later when new discoveries and/or conceptual frameworks show that our initial proposals of what was possible were actually incorrect.

So, while Chalmers is correct to point to the natural boundaries of what is physically or naturally possible – i.e. the natural laws of our world – I believe for the reasons given above that this is an idealisation and that *in practice* our estimations of physical possibilities will often have an epistemological component.

Logical Possibility

Another commonly used type of possibility in philosophy is logical or conceptual possibility. This possibility is generally seen as having to do with the conceptual coherence of the proposed phenomenon. When trying to determine if something is logically or conceptually possible it usually is not seen to matter whether the thing is *physically* possible or not. In fact, sometimes we know that it is not physically possible, but we try to see whether the concept involved is logically *coherent* or not anyway. Often this boils down to discerning that there is at least no logical contradiction in the concept. For example, considering the notion of a flying horse, one could judge there to be no logical contradiction in the concept of a horse with wings and capable of flying and, therefore, judge the phenomenon to be logically possible, even though one might judge it to be physically impossible.

On the subject of logical possibility Chalmers says that it is “quite unconstrained by the laws of our world” and that we can think of it “loosely as possibility in the broadest sense” where the constraints on determining such possibility are “largely

conceptual” (p. 35). He also describes logical possibility in terms of possible worlds: “It is useful to think of a logically possible world as a world that it would have been in God’s power (hypothetically!) to create...God could not have created a world with male vixens, but he could have created a world with flying telephones” (p. 35). The reason God could not have created male vixens is because they are logically impossible – that is, a male vixen is a contradiction in terms as a vixen is *by definition* female. So this could lead us to conclude that any (possible) world that does not contain a logical impossibility, a logical contradiction, is a logically possible world. In fact this is precisely the way in which Chalmers uses the notion later on in his book when arguing for the logical possibility of the existence of zombies: “While this is probably empirically impossible, it certainly seems that a coherent situation is described; I can discern no contradiction in the description” (p. 96).

Thus it would appear to be the case that if we can discern no logical contradiction in a given proposed situation, then that situation is logically possible. Of course, the controversial area here lies in discerning a logical contradiction in the situation. Some would argue that, while the description of a particular situation is not explicitly logically contradictory, it is nevertheless *implicitly* so. One tends to think of logical or conceptual possibility as encompassing physical possibility but also including much more than just the physically possible. There seems to be no reason, logically speaking, why the laws of nature, say, could not have been other than what they actually are. Arguably, logical possibility is limited by our conceptual *abilities* – i.e. the limit of our capacity to think. This is contrary to estimations of physical possibility which, as I argued above, would appear to be limited by our *knowledge* (ideally, by the actual physical limits of the world, but our *estimations* of physical possibility depend upon our current knowledge). However, it seems that in determining more precisely the limiting factors for the estimation of logical possibility, one could go two ways.

Firstly one could use the limit of our conceptual abilities as the limiting factor and say, as I have just described, that as long as we can find no *formal* logical contradiction in the proposed notion or concept (for example involving a claim that *A* is *not A*, a clear logical contradiction) then the notion or concept⁸ is logically possible. And, as I mentioned above, this seems to be the kind of thinking behind Chalmers' claim for the logical possibility of the existence of zombies, when he says that he can discern "no [logical] contradiction in the description".

The second way of determining what is logically possible, I suggest, would entail using the limits of our (current) *conceptual frameworks*, rather than merely our general conceptual abilities, as a boundary when estimating such possibility. Although Chalmers describes logical possibility according to the *first* way I have mentioned – where the "constraints are largely conceptual" – he also seems to use the notion in this second sense, i.e. possibility constrained by conceptual frameworks. At least this appears to be the case when he illustrates what he means by the notion of logical possibility. He gives the example of biological properties supervening *logically* on physical properties, claiming that once we fix the physical facts then (at a global level) there is "simply no logical space for the biological facts to independently vary" (p. 35). In other words, once the physical facts are fixed, it is simply not *logically* possible for the biological facts to be other than what they actually are in the actual world.

Using *this* way of determining logical possibility seems to tie it to physical possibility, at least in this example given by Chalmers of biology. This is because the limiting factor for estimating logical possibility that this second way of looking at the problem draws on, is not so much one of mere logical coherence but rather one that is also bounded by, and thus must fit in with, our current conceptual frameworks. This may seem like an odd usage of the term "logically possible", as we do not tend to think

⁸ Chalmers talks about the logical possibility of *statements* rather than of concepts, but, either way, the *limiting* factors with respect to determining logical possibility would be the same.

that logic should be bounded by our current conceptual schemes of how the physical world is, but it is not an uncommon notion in philosophy, as I will try to show later in this chapter.

Returning to Chalmers' example of biological facts being *logically* dependent on the physical facts, if we go with the first notion of logical possibility I mentioned above – limited purely by what is logically coherent, i.e. that there is no logical contradiction – then I suggest, to the contrary, it plausibly does seem logically possible that we could have different biological facts for the same physical facts. At least there seems to be no logical *contradiction* in the concept of having the same physical facts hold as in the actual world but nevertheless having different biological facts hold. We could simply stipulate a possible world where some of the laws of nature had changed, for example. Compare this with Chalmers' other example, the possibility of a male vixen. Now this of course does not seem logically possible in any sense because by definition a vixen is female and therefore a male vixen is a contradiction in terms, as if we were saying a thing is both not-A and A, and thus a logical contradiction⁹. With the biology example, however, there is no such *obvious* logical contradiction.

However, taking the second notion of logical possibility mentioned above, where the limits to what is logically possible are determined by our current conceptual frameworks, then, as Chalmers himself argues, it does not seem logically possible that exactly the same physical facts could give rise to different biological facts. If we know roughly *how* the actual physical facts lead to the actual biological facts then, as Chalmers puts it, there is no “logical space” for the latter to vary. This is not to say that we need to know every detail of how the higher level facts are entailed by the lower level facts. In fact arguably we need only know *that* the former are entailed by the latter, but we generally only come to know this by discovering the connections between the

⁹ Setting aside Quine's qualms about analyticity, of course, something Chalmers seems to do himself since he claims that even God could not have created a male vixen.

two levels: the “how”. Thus, when we have discovered in broad lines how the higher level is related to the lower level, then we can see *in principle* how the biological is entailed by the physical. Basically this second kind (or interpretation) of logical possibility ties it to what is possible within our current conceptual frameworks and therefore, effectively, to what is physically possible. And it is exactly between these two interpretations of logical possibility that intuitions vary: some believing anything, even wildly outside physical possibility, should be considered logically possible as long as there is no obvious contradiction in the notion, or at least none that cannot be explained by stipulating that the laws of nature could have been different¹⁰ (without always supplying any detail of how); others looking for explanations, or at least indications, of *how* something impossible within the confines of our current conceptual frameworks could nevertheless be deemed logically possible.

It does appear from my comments above that Chalmers himself is equivocating on these two interpretations of logical possibility. While he pleads for logical possibility in the “broadest sense”, “quite unconstrained by the laws of our world”, in the next breath he argues that even “God could not have created a world that was physically identical to ours but biologically distinct” (p. 35). Of course he would presumably argue that a world physically identical to ours includes the laws of physics – he does say that it is “useful to stipulate that the world’s physical facts include its basic physical laws” (p. 33). Therefore, he could argue that with exactly the same physical properties *and laws*, one could not possibly get different biological facts – it is not even a logical possibility. But if he holds everything (all physical properties, laws, etc.) the same, then his argument for the logical possibility of zombies should fall under the same reasoning. The only difference in the zombie case is that we do not *know* if consciousness is determined by the physical facts or not. In other words, *pace* Chalmers, different

¹⁰ See Alan Sidelle’s article “On the Metaphysical Contingency of Laws of Nature” (2002) for such a viewpoint.

(phenomenal) consciousness facts from the same physical facts may *not* be logically possible at all, in the same way that different biological facts are not (according to Chalmers), but we simply do not yet *know* whether this is the case or not. Our knowledge in the area of phenomenal consciousness is far from complete. One could imagine the same argument as Chalmers' plea for the logical possibility of zombies being proposed, say, hundreds of years ago (in particular, prior to Darwin) for the logical possibility of different biology from the same physical facts. When very few of the connections between the biological and the physical were known and when, in fact, we did not even have the *conceptual framework* wherein we could view biological life in terms of function and/or structure, then it could certainly have *seemed* logically possible for the same physical facts to lead to a different biology. People did not *know* then that biology (life) was completely entailed by the physical (assuming Chalmers' proposal that it is).

I believe much of the problem in determining possibility claims in general, logical or physical, is tied to our current knowledge of the relevant issues or, more pertinently, our lack thereof. As I mentioned in the section on physical possibility, there can sometimes be wide gaps between what we believe to be possible and what actually is possible and this occurs mainly because we are most often interested (particularly in science and philosophy) in estimating possibilities at the limits of, or beyond, our knowledge. I think this is true even more so of logical possibility, especially in philosophy of mind. We are usually trying to estimate the possibility of hypothetical phenomena (e.g. zombies) about which we have no facts, or phenomena about which we know very little, such as how consciousness occurs in humans. It is arguably this question of *knowledge* that leads to the two different interpretations of the boundaries of the logically possible mentioned above.

On the one hand, once we know roughly *how* (and, therefore, *that*) the physical

facts make up the biological facts, then it does seem logically impossible, within our current conceptual frameworks, that the same physical facts could give rise to a very different biology. Let us term this kind of logical possibility, logical possibility in the *framework sense*. At the very least, if someone were to claim it to be logically possible, I believe we would be justified in demanding *some* kind of explanation as to how, under what conditions, laws, etc., such a thing could come about. Assuming, as Chalmers seems to do, that we do in principle know the physical underpinnings of the biological facts, then this fits into whatever conceptual frameworks we use for explaining our universe – our universe as we currently see it. If we try to claim the possibility of things being otherwise (different biological facts for same physical ones) then, within those frameworks, that possibility claim makes no sense whatsoever (unless a countering theory is supplied, of course) and is in this sense logically impossible.

On the other hand, one can say that there is nevertheless no formal logical contradiction in the notion that we could get a different biology from the same underlying physical facts: that is, there is no obviously contradictory claim of the type, *A is not A*. Let us call this logical possibility in the *formal sense*. Why could our laws of nature not have been such that when one had the same physical facts that we have now, one would have different biological facts? It may not be physically possible but surely, from a purely *logical* point of view, it is a possibility? For example, we could imagine that biological laws were determined by certain emergent properties (as perhaps some vitalists once held) that could vary independently of low-level physical facts. Indeed we make such possibility claims all the time, imagining that the world is quite a different place to what it actually is. But this kind of possibility claim seems completely unbounded by any kind of knowledge. As I mentioned, it is often with respect to areas where our current knowledge is severely lacking that we make such postulations and, arguably, it is our very *lack* of the relevant knowledge that makes the proposed notion

seem logically possible in this formal sense. If one is to inquire further into the proposed logical possibility and look for explanations or even indications, as to how such a thing could arise, they are usually not supplied. In other words, we do not have enough knowledge of the relevant phenomenon (we often have no developed framework for it) with which to *find* a logical contradiction in the proposed scenario and therefore it seems as if there is none.

Generally estimations of this kind of logical possibility (*viz.*, in the *formal* sense) do not include stipulations as to how the proposed possibility could come about. Arguably suggesting that some laws of nature be different does not completely solve the problem as questions still remain about the way in which such laws should be different. And what effect will even a small change in a particular law actually have? The answer to these questions is simply that we do not know. At any rate, if it is stipulated that the basic physical laws are also fixed along with the physical facts, as Chalmers seems to do and as is indeed a reasonable stipulation, then arguably it is not logically possible in any sense for the biological facts to vary given the physical facts. One could still try to claim however that there are *two* different possible sets of biological facts that could be entailed by the *same* set of physical facts (and laws) and that, therefore, it is logically possible in the *formal* sense (i.e. there is no logical contradiction in the proposal) that we could get different biological facts from the same physical ones. At first glance this seems like a possibility but, especially given that we are to suppose that we have all the physical facts, including laws, one should be able to give some kind of indication as to *how* these same physical facts could give rise to a different, second set of biological facts in order to claim that such a proposal *is* logically possible. Otherwise we seem to be left with “anything goes” claims of logical possibility, the basis of which is founded only on a superficial *apparent* lack of contradiction in the proposed possible situation. More detail about the proposal might show inherent logical contradictions which were not

initially obvious. However, the problem here is usually that such details are not available as we are speculating in areas where we have little or no knowledge. If we try to judge the logical coherence of a situation of which we have very little knowledge, then arguably we cannot *see* implicit logical contradictions (if they are there) precisely because we do not know enough about the proposed situation to be able to find them. Ignorance is bliss and it is easy to claim there are no logical contradictions when one does not have the means (informational and/or conceptual) to see them.

Therefore, I would argue that whether we accept the claim that, say, it is logically possible to have a different biology given the actual existing physical facts of the world, depends on our degree of knowledge about the proposal in question and how it embeds in our current conceptual schemas. If we *know* that the physical facts of our world give us the biological facts, i.e. if we have an idea as to how they are related, then we will probably have more difficulty accepting a claim that it is logically possible that the same physical facts could give us different biological facts. We will want to know *how* this is possible, what mechanisms could account for this, and so on, because it would have to *fit in* with the rest of our conceptual frameworks. If we have little idea as to how a particular phenomenon works or what it is, for example what consciousness is, we tend to allow that it be logically possible that the phenomenon be anything at all really, as we have only sketchy knowledge in relation to the phenomenon in question with which to *find* a contradiction in the proposed description. In reality these two means of determining logical possibility that I have described, really reflect a difference between estimating the logical possibility of situations where the facts, including natural laws, are (more or less) known and so stipulated as fixed, and situations where the facts are not fixed or known, because we do not have many facts yet.

Although, as I have pointed out, Chalmers describes his notion of logical possibility as a “broad” one based on lack of formal logical contradiction, he

nevertheless appears to use both ways I have described for estimating logical possibility. On the one hand he seems to use logical possibility in the *framework* sense when he maintains that it is logically impossible that we could get different biological facts from the same physical facts, presumably because according to our current conceptual schemes the biological is entailed by the physical. So in this case he is allowing that our conceptual schemes determine what is deemed to be logically possible. However, on the other hand, in relation to the zombie case, he claims that it *is* logically possible in the *formal* sense to get different (phenomenal) consciousness facts from the same physical facts because there is, he claims, no logical contradiction in the notion. This is presumably because we do not have any idea how, or even if, consciousness is entailed by the physical. As Chalmers has argued, we have no logical entailment of the phenomenal by the physical within our current conceptual frameworks because we have no functional or structural description of phenomenal consciousness. But we do not yet have an adequately developed framework for dealing with phenomenal consciousness. We do not yet know how it is to fit into our conceptual frameworks – or how these frameworks may have to change to accommodate it – so we cannot yet make a judgement on logical possibility in the *framework* sense in this area.

When Chalmers claims that he can discern no logical contradiction in his notion of a zombie, this is because he has very little information with which to find such a contradiction in the notion, as our knowledge of the underlying nature of consciousness is currently so lacking. The fact that our existing conceptual frameworks and information are apparently not up to the job of giving us logical entailment from the physical to the phenomenal does not *show* that it *is* logically possible to have the same physical facts without the phenomenal ones. Rather this only *seems* logically possible, given our current (rather inadequate) knowledge/conceptual schemes, and *could* be an illusion that may be corrected by further empirical discoveries in the area and perhaps

further development or change in the relevant conceptual frameworks. This would be in much the same way that our notion that life is entailed by the physical developed slowly through empirical discoveries and eventual changes in our concept of life.

So, arguably, such logical possibility in the formal sense with regard to phenomena about which we know very little is really only apparent logical possibility – a superficial, *prima facie* judgement of lack of logical contradiction in the proposal – and is thus highly fallible. Therefore I would urge use of the more cautious term “epistemic possibility” in such situations because it acknowledges the fallibility that arguably lies behind many uses of the term “logical possibility”.

So making a claim for the logical possibility of the existence of zombies at *this* stage, when we do not know enough facts to discern whether there *really* is a logical contradiction in the notion or not, seems rather premature as such claims are based on a lack of pertinent knowledge and may be shown to be unfounded in the future. Certainly using such a claim as *proof* that there is no entailment between the phenomenal and the physical, as Chalmers wants to do, seems entirely unjustified. If the very correctness of the alleged logical possibility of different phenomenal facts from the same physical facts (e.g. zombies) is itself questionable then this surely cannot be used as *proof* that there is no entailment from physical to phenomenal. Thus arguably any claim for the logical possibility of zombies in the *formal* sense is more a case of epistemic (or apparent) possibility and, in the *framework* sense, is as yet uncertain since we do not yet have an adequate framework to deal with the facts of consciousness. But these are points that need to be teased out further and I will return to them later on in this chapter. In the following section I will examine the notion of epistemic possibility, particularly with regard to the notions of logical possibility I have discussed here.

Epistemic Possibility

The apparent reliance that I propose we have on our current knowledge and conceptual frameworks for estimations of logical and physical possibility, thus leads directly to the subject of another type of possibility: epistemic possibility. A situation (or statement, to use Chalmers' language) is epistemically possible if it is possible according to the best of our current knowledge – in other words there is nothing in our current knowledge or conceptual schemes that rules the possibility out. However, often the reason that there is nothing to rule out the possibility of a situation is precisely because our knowledge with regard to the relevant situation is lacking. In such cases, one could say that epistemic possibility is just *apparent* possibility. For example, when trying to estimate the logical possibility of a hypothetical scenario, the proposed situation, or statement, may seem logically possible because we do not know enough about it to see either that it is definitely logically impossible or that it is definitely logically possible. As in the case of something like Goldbach's conjecture: one may think it logically possible (in the *formal* sense) that this is (say) true because nothing currently known rules it out. However, if it turns out to be false, then it was never really logically possible at all that the conjecture be true. So such an estimation of logical possibility with regard to situations about which we know very little (logical possibility in the *formal* sense) could arguably better be called epistemic possibility to more aptly reflect the degree of fallibility or unreliability in the judgement.

In his book, Chalmers does not discuss epistemic possibility. However, in his later article "Does Conceivability Entail Possibility?" (2002), he does talk about this kind of possibility in relation to primary conceivability (i.e. conceivability with respect to primary intensions, or 1-conceivability). He says that we can think of "ways the world might be as *epistemic possibilities*, in a broad sense according to which it is epistemically possible that S if the hypothesis that S is not ruled out a priori" (Chalmers,

2002, p. 157). This sounds quite similar to my definition above, although it is not clear if Chalmers' notion of the *a priori* would be as knowledge-dependent as I have explicitly stipulated epistemic possibility to be (such knowledge would very much include empirical *a posteriori* knowledge in my definition).

At any rate, many philosophers do acknowledge the subjectivity of epistemic possibility and connect it directly with what is known either by a speaker or community of speakers. Keith DeRose (1991) in his paper "Epistemic Possibilities" proposes that he will "defend a knowledge-based analysis of epistemic possibilities" (DeRose, p. 581). Lloyd Reinhardt (1978) in his paper "Metaphysical Possibility", to illustrate epistemic possibility, gives the example of a pupil mistakenly thinking it possible that there be five prime numbers between 10 and 20. He says: "Speaking epistemologically, it is possible that there are five. But speaking metaphysically, it is impossible that there should be five" (Reinhardt, p. 211). He goes on to say: "My pupil may be said honestly to *think it possible* that there is a fifth prime between 10 and 20. But this does not show that it *is possible* any more than honestly thinking something dangerous shows that it *is dangerous*" (ibid.). Thus the notion of epistemic possibility seems to allow for fallibility, in that it is dependent on the knowledge of the speaker or of the community of speakers.

Arguably all our possibility claims are to some extent based on the best of our current knowledge. However one would hope and expect to find little difference between our possibility claims and what actually *is possible* in areas in which we are quite knowledgeable. But, as I have argued in the previous section, many of our claims of logical possibility with regard to phenomena at the edge of, or beyond, our current knowledge, are really just epistemic possibility claims in the sense that they are claims of *apparent* logical possibility. In other words, with our current lack of knowledge of the proposed phenomenon in such cases, we can discern no logical contradiction in the

proposed scenario. But when more empirical information is in we may well begin to see logical contradictions in the proposal (in the context of the new developing framework or changes in the way we view this information). Thus, it does not seem that a claim of logical possibility (in the *formal* sense) with respect to such a phenomenon at the limits of our knowledge is justified. The *absence* of any reason to rule out the proposed possibility, with respect to an area where our knowledge is poor or non-existent anyway, is certainly not enough to show that it really is logically possible – i.e., that there is no hidden logical contradiction which we will uncover as we discover more later on. At most, I would suggest, we can claim epistemic possibility until more is known or discovered about the proposed situation. This I think would more accurately reflect our lack of knowledge with which to find a logical contradiction in the proposed possibility, and therefore the inherent fallibility of such a claim of logical possibility, until we have developed a better relevant knowledge framework.

Metaphysical Possibility

Another type of possibility that is also discussed by Chalmers, is metaphysical possibility. This kind of possibility, although widely used in philosophy, is regularly ill-defined or not defined at all. When it is defined, it is frequently contradictorily described, with some philosophers likening it more to a kind physical possibility and others to a kind of logical possibility. For example, in his paper “Ross and Scotus on the Existence of God: Two Proofs from Possibility”, G. Randolph Mayes describes metaphysical possibility as “possibility that we associate with being, rather than thought” (Mayes, p. 100). This would seem to be contrasting metaphysical possibility with logical possibility, the latter normally being more associated with thought or logic.

Associating metaphysical possibility with being would arguably tie it to physical possibility, although not all philosophers would agree. Often appeal is made to some

kind of “rational insight” by means of which we can recognise metaphysical possibility or necessity, as for example in judging that every event has a cause. This involves a tendency to see metaphysical possibility, and the metaphysical in general, as more than just physical possibility. In other words, to see it as that which could actually be if, say, our laws of nature were different. In such a case presumably metaphysical possibility would differ from physical possibility since the latter is bound by the natural laws of the actual world. However, describing metaphysical possibility in this way seems to make it more akin to logical possibility in the sense that, in order to determine if something is metaphysically possible in a situation where the laws of nature are different to those of the actual world, the principal tool on which one has to rely is that of logical coherence. But the latter is normally the means of determining *logical* possibility.

In order to be able to judge metaphysical possibility one must be able to determine the limiting factors for estimating such possibility, as for example with physical possibility where I suggested that what is physically possible is bounded (ideally) by the laws of nature. The boundary for what is logically possible I have argued is the limit of our conceptual abilities or (perhaps in cases where the facts and conditions involved are known and can be stipulated) of our current conceptual frameworks. However, there does not seem to be any kind of boundary in between the laws of nature and the limits of our conceptual abilities/frameworks with which to “mark off” metaphysical possibility, as it were, from physical or logical possibility. As I mentioned above, if we say that metaphysical possibility concerns that which could be *regardless* of the natural laws of our world, then the only means we have of estimating such possibility is *exactly* the same means we have of estimating *logical* possibility: i.e. that the described situation is logically coherent (or fits into our conceptual schemes). So, in this case, how do we tell metaphysical possibility from logical possibility? If, however, we say that metaphysical possibility concerns that which could actually be, i.e.

in our world, then such possibility is determined by the laws of nature. In this case, how do we tell metaphysical possibility from physical possibility?

For example, one might wish to claim that it is metaphysically impossible that whatever has shape does not have size or that two things of the same type can occupy the same point in spacetime. But our only means of determining the (im)possibility of such statements is either to use the laws of logic (is there a logical contradiction in the notion?) or to use the laws of nature (is it physically possible?). It would seem to be a logical impossibility that something have shape without having size since the former seems to be inherent in the concept of size. It may be a logical possibility in the formal sense that two things (of the same kind) could occupy the same point in spacetime, in that there is perhaps no logical contradiction in the notion. It is probably not a logical possibility in the framework sense because it would contradict our current conceptual schemes for the laws of nature with regard to four-dimensional space. Certainly if it contradicts the laws of nature, it would not be physically possible. Either way, there does not appear to be an in-between: we either look to the logical coherence of the notion (either in a minimal formal sense or with regard to our conceptual frameworks) or to the laws of nature to determine the possibility of the statement. These methods reflect determining logical possibility or physical possibility respectively. I can discern no other means of determining the possibility of something other than using the natural laws of our world or logical coherence. Therefore I can only conclude that metaphysical possibility must collapse onto one of the other possibilities – either physical or logical (in either the *formal* or *framework* sense).

Nevertheless, as I have pointed out, many philosophers do tend to see metaphysical possibility as a third, separate type of possibility, in between physical and logical possibility, as it were. For Chalmers however, metaphysical possibility apparently has a meaning different to this traditional view, for he says that he will argue

that “the metaphysically possible worlds are just the logically possible worlds” (p. 38). Here he seems to indicate that metaphysical possibility is simply logical possibility, although he does qualify this with respect to possibility of statements (as opposed to worlds). He says that “metaphysical possibility of statements is logical possibility with an *a posteriori* semantic twist” (p. 38). The distinction between logical and metaphysical possibility he claims is “not a distinction at the level of *worlds*, but at most a distinction at the level of *statements*” (p. 68). As explained in chapter one, Chalmers interprets logical and metaphysical possibility in terms of his two-dimensional intensional framework as 1-possibility and 2-possibility respectively. He describes both 1- and 2-possibility as “two sorts of logical possibility” (p. 67). Thus it would appear that he sees the difference between logical and metaphysical possibility as merely one of semantics, “stemming from the two-dimensional nature of our concepts” (note 34, p. 68).

However he does show that he is aware of the traditional usage of the term “metaphysical possibility” as he refers to the term in the way many other philosophers would, that is, as “possibility *tout court*”. He discusses objections to his argument that it is logically possible that there exist a zombie world, saying that a “popular response to this sort of argument is to object that it only establishes that a zombie world is *logically* possible, which is quite different from being *metaphysically* possible” (p. 131). Indeed, he acknowledges Kripke’s use of the term: “the oft-cited distinction between ‘logical’ and ‘metaphysical’ possibility stemming from the Kripkean cases – on which it is held to be logically possible but not metaphysically possible that water is XYZ” (p. 67-8).

However, Chalmers believes that logical possibility is the most crucial kind of possibility when it comes to his argument against a materialist view of consciousness. In formulating his definition of what a materialist view must hold, he makes use of logical possibility when he says: “materialism 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” (p. 42). He then goes on to comment on claims that he should be using the term “metaphysical possibility” by saying:

Some may object to the use of logical possibility rather than possibility *tout court* or “metaphysical possibility”. Those people may substitute metaphysical possibility for logical possibility in the definition above. Later, I will argue that **it comes to the same thing** (p. 42, my bold type).

So it appears that Chalmers is effectively ensuring the irrelevance of metaphysical possibility for his zombie argument by defining it as a logical possibility relating to *a posteriori* considerations (2-possibility) and claiming that his zombie argument need only address *a priori* considerations. In this manner he appears to believe that he can dismiss appeals to metaphysical possibility and *a posteriori* considerations against his argument.

Having looked at the various types of possibility at issue in regard to Chalmers’ zombie argument, in the following sections I will critically examine how Chalmers grounds his assertion that metaphysical and logical possibility come to “the same thing” and his claims of the irrelevance of *a posteriori* considerations. But in order to do this I first need to look at Kripke’s notion of *a posteriori* necessity.

A POSTERIORI NECESSITY

In his book *Naming and Necessity*, Saul Kripke (1980) argues that there are *a posteriori* necessities – that is, necessities that are discovered empirically and that are not knowable in an *a priori* way. Whereas previously necessity had been associated with the *a priori*, Kripke argues that not everything necessary is *a priori*: “It’s certainly a philosophical thesis, and not a matter of obvious definitional equivalence, either that everything *a priori* is necessary or that everything necessary is *a priori*” (Kripke, p. 36). Furthermore, he points out that *a priori* and necessity actually deal with *different*

domains: “they are dealing with two different domains, two different areas, the epistemological and the metaphysical” (ibid.). Kripke is concerned with a *metaphysical* notion of necessity: “what I am concerned with here is a notion which is not a notion of epistemology but of metaphysics” (Kripke, p. 35). His claims of *a posteriori* necessity concern actual identities, empirically discovered. So as to get a better insight into Kripke’s arguments for such a necessity and the resulting effects on possibility claims, it is important to look briefly at his theory of naming.

Kripke views a (proper) name as a “rigid designator”, where the latter designates the same object in every possible world. This view is also tied in with his view that objects have *essential* properties. Kripke describes such a property by pointing out that when “we think of a property as essential to an object we usually mean that it is true of that object in any case where it would have existed” (Kripke, p. 48). Presumably¹¹ there must be *some* essential properties because if we change enough of an object’s properties in speculating about its existence in other possible worlds, we can no longer identify the object at all as the same one we started out with. As Kripke puts it, a rigid designator picks out a particular object and we can imagine that certain (contingent) features of *that* object could be different in another possible world but not that the object *itself* would be different, as then we would be talking about *another* object. He gives the example: “although the man (Nixon) might not have been the President, it is not the case that he might not have been Nixon (though he might not have been *called* ‘Nixon’)” (Kripke, p. 49).

Kripke’s notion of a rigid designator stems from the way in which he believes we name objects. Arguing against any kind of “cluster of descriptions” or “description” theory of names, he claims that if a name “*means the same* as that description or cluster

¹¹ Of course, many philosophers, such as Quine, would disagree with Kripke’s analysis. However Chalmers himself does not follow such a route but rather seems to accept the Kripkean analysis of *a posteriori* necessary identities (he does not argue against it *per se*), incorporating them into his own scheme as necessity with respect to secondary intensions.

of descriptions, it will not be a rigid designator” (Kripke, p. 57). For example, if we use the description “the greatest man who studied with Plato” as a *definition* of “Aristotle”, then, in any other possible world, *another* man might have filled this description and thus might have been Aristotle. Likewise we can imagine that the man Aristotle might not have become a philosopher at all, and so would not fit the description but would nevertheless still be the same man. Therefore, in Kripke’s view, a proper name cannot be synonymous with such descriptions, and such descriptions, in general, cannot be part of the *meaning* of the name. If they were part of the meaning of a name, we would have no way of using the name to *consistently* refer to the *same* object because such descriptions are not rigid designators and could point out different objects across different possible worlds. The confusion, he argues, lies in the fact that such descriptions do not actually give a synonym for a name but are really used to *fix the reference* of that name (in the actual world). He claims that it is often the case that the reference of a name is fixed by a *contingent* property of the entity to which the name refers. Hence a description of this contingent property could not be part of the *meaning* of the name, as the entity might not have had this property at all.

So according to Kripke, various descriptions associated with a given name (often picking out contingent properties of the named entity) are not part of the *meaning* of that name but are used to *pick out the referent* of that name. Thus he allows that a person could have had some different properties in another possible world, for example Aristotle might not have been a philosopher, but asserts that the *identity* of that person is *necessary*. Although we may “pick the man out by an accidental property of the man, still we use the name to designate that man ... in all possible worlds” (Kripke, pp. 75-6). Similarly, with respect to “natural kind” terms, such as water, tigers, gold, etc., Kripke argues that once we have established what, say, water is (H₂O) then this is a *necessary* identity. Even though it is an empirical discovery that water is H₂O or that heat is

molecular motion, these are nevertheless necessary truths and not, as it might appear, contingent ones. Indeed he claims they are necessary “in the highest degree” (Kripke, p. 99). With regard to the phenomenon of heat for example (and theoretical identifications in general), when we eventually discovered it to be molecular motion, then Kripke argues, we have “discovered a phenomenon which in all possible worlds will be molecular motion – which could not have failed to be molecular motion, because that’s what the phenomenon *is*” (Kripke, p. 133). Thus such (*a posteriori*) identities are *necessary* truths according to Kripke because once we have discovered a given *identity*, we have discovered what the given phenomenon actually *is*. If we were to speculate about that given phenomenon being something else (e.g., heat being something other than molecular motion) then we are already talking about some *other* phenomenon, as the given phenomenon is what it has been discovered to be.

However Kripke admits that there can be a kind of confusion in the sense that we seem to think that we can imagine possible worlds where, say, Hesperus is *not* Phosphorus, even though both names refer to the *same* planet Venus in the actual world. Since Hesperus actually *is* Phosphorus in the actual world then this identity is necessary and so, accordingly, Hesperus must be Phosphorus in every possible world. But he points out that when we imagine that Hesperus is not Phosphorus in some possible world, we are imagining a situation in which there are two different “heavenly bodies” visible in the same position in the sky at different times, to which we give the names “Hesperus” and “Phosphorus”. But in such a case the names (or at least one of them) are being used differently to the way we actually use them: “in a counterfactual world in which ‘Hesperus’ and ‘Phosphorus’ were not used in the way that we use them, as names of this planet [Venus], but as names of some other objects, one could have had qualitatively identical evidence and concluded that ‘Hesperus’ and ‘Phosphorus’ named two different objects” (Kripke, p. 104). (Kripke argues that we must use the one

language, “our” language, when talking about other possible worlds – which makes sense for if terms changed their meaning/usage between worlds, we would have no common frame of reference and no means of comparing two worlds.) Thus the sense in which we think that we can imagine Hesperus being different to Phosphorus is merely epistemic, a “qualitatively identical epistemic situation” (ibid.). In other words, it is an *epistemic* possibility only that Hesperus might not be Phosphorus, but, according to Kripke, given that Hesperus *is* Phosphorus, it is not a *real* possibility that they are different. (Although Kripke does not use the term “epistemic possibility”, it seems reasonable to characterise “qualitatively identical epistemic situation” as a case of epistemic possibility.)

A PRIORITY, NECESSITY AND INTENSIONS

Chalmers incorporates Kripke’s notion of *a posteriori* necessity into his own intensional framework. In chapter one I explained Chalmers’ notions of primary and secondary intensions and how they both “back a certain kind of conceptual truth”. The primary intension of a concept “backs” *a priori*¹² truths, such as, for example, “water is watery stuff”, according to Chalmers. This is true in virtue of meaning presumably because watery stuff (clear drinkable liquid found in oceans, lakes, etc.) is what water comes to mean to us before we ever discover the underlying nature of water. The secondary intension of a concept “backs” *a posteriori* truths, such as “water is H₂O”. Chalmers also calls this a conceptual truth presumably because of Kripkean *a posteriori* necessity according to which water is necessarily H₂O. Even though this is an empirical discovery, it becomes the (*a posteriori*) meaning of water and thus we have truth in virtue of meaning.

¹² As I noted in chapter one, this appears to be a *relative* sense of *a priori* since formation of the primary intension of a concept is in fact largely an empirical process.

So, how does Chalmers' two-dimensional framework for the meaning (or, more correctly, intension) of a concept affect his notion of metaphysical possibility and *a posteriori* considerations in general? He introduces this two-dimensional framework precisely in order to deal with the consequences of Kripke's notion of *a posteriori* necessity for conceptual and necessary truth. Chalmers relates Kripke's *a posteriori* analysis of concepts to the *secondary* intensions of concepts in his own framework:

However, *given* that "water" turns out to refer to H₂O in the actual world, Kripke notes (as does Putnam [1975]) that it is reasonable to say that water is H₂O in every counterfactual world. The *secondary intension* of "water" picks out the water in every counterfactual world; so if Kripke and Putnam are correct, the secondary intension picks out H₂O in all worlds (p. 57).

So for *a posteriori* concerns, one must evaluate a concept using its secondary intension: i.e. what it turns out to refer to in the actual world. According to Chalmers, the primary intension of a concept, on the contrary, is an *a priori* notion insofar as it is our notion of what that concept means even before we have determined any further empirical facts about what it refers to in the actual world (e.g., H₂O). Chalmers wants to maintain a link between such *a priori* knowable truths and necessary truths. However he sees this link as potentially threatened by Kripke's account of *a posteriori* necessity because this account shows that there are necessary truths that are only knowable *a posteriori* and not *a priori*. Thus Chalmers sees his problem as: "on some accounts these necessary truths are conceptual truths, implying that not all conceptual truths are knowable *a priori*", however according to other accounts "such statements are not conceptual truths, but then the link between conceptual truth and necessity is broken" (p. 56). Thus Chalmers wants the link between conceptual truths and necessity to remain intact while *at the same time* ensuring that conceptual truths (or at least the relevant kinds) are knowable *a priori*.

He tries to do this by analysing concepts in terms of primary and secondary intensions and argues that it is the primary intensions (*a priori* aspect) that are important when it comes to explanation: "for a concept of a natural phenomenon, it is the primary

intension that captures what needs explaining” (p. 57). This is because, according to Chalmers, “If someone says, ‘Explain water’, long before we know that water is in fact H₂O, what they are asking for is more or less an explanation of the clear, drinkable liquid in their environment” (p. 57).

Returning to the notion of metaphysical possibility, as I explained in the section on possibilities, Chalmers equates this with 2-possibility and maintains that both 1- and 2-possibility are types of logical possibility. I believe that this latter assertion is based on his earlier claim that each intension “backs” a conceptual truth or “truth in virtue of meaning” and therefore the associated possibilities would be conceptual, or logical possibilities. Metaphysical possibility for Chalmers then is a logical possibility but related to the *a posteriori* aspect of a concept. As I mentioned previously, he claims that the distinction between logical and metaphysical possibility is one at the level of *statements* and not at the level of *worlds*, saying that the “relevant space of worlds is the same in both cases” (p. 68). Presumably the space of worlds is the space of logically possible worlds, as this would be the largest set of possible worlds and would automatically encompass any physically or metaphysically possible worlds. But it is not clear what Chalmers gains by this move. Instead of the perhaps more common notion that there is the set of physically possible worlds, a perhaps bigger (or the same size) set of metaphysically possible worlds and the largest set of logically possible worlds which contains the other two sets, Chalmers wants there to be one space of possible worlds.

Let us examine this further. If the distinction between logical possibility (1-possibility) and metaphysical possibility (2-possibility) is at the level of *statements* then what does it mean for a statement *not* to be metaphysically possible? Chalmers proposes that in such a case we have simply “misdescribed” the conceived world. For example, if we think that we can conceive of a world where water is XYZ then all that Kripkean *a posteriori* considerations (i.e. that water is H₂O, not XYZ) can show us is that we have

incorrectly described that possible world and that we should put our description in terms of the primary intensions, thus claiming something like: “watery stuff is XYZ”. This does nothing to change the *possibility* of that conceivable world itself however – the fact that it was “misdescribed” does not change its status to an impossible world. We just need to describe the world more accurately and it is *still* a possible world. So Chalmers believes that by keeping the space of possible *worlds* the same, regardless of logical or metaphysical possibility of *statements* in relation to those worlds, he can conclude that there “seems to be no reason to deny that conceivability of a world implies possibility”. Thus he says: “I will henceforth take this for granted as a claim about logical possibility” (p. 68). In this manner I believe he thinks he has shown that conceivability (of a world) *does* imply (logical) possibility (of that world).

It is of course important to note that conceivability of a world therefore would imply 1-possibility of that world but *not* 2-possibility (metaphysical possibility, according to Chalmers). Or rather, more accurately, particular statements describing that world may not be metaphysically possible: e.g., “water is XYZ”. However, if we change the statement to reflect the primary intensions of the concepts involved (“watery stuff is XYZ”), then the resulting description is 1-possible (logically possible). So that possible world, now correctly described, is also logically possible (1-possible) according to Chalmers. But he has thereby still *not* shown that conceivability of a world implies *metaphysical* possibility (2-possibility) of that world. Of course he only wishes to discuss his argument in terms of logical possibility as he argues that logical possibility is all that is relevant for his purposes anyway. Nevertheless he does make a claim for metaphysical possibility of *worlds* in the case of consciousness when he says:

if there is a conceivable world that is physically identical to ours but which lacks certain positive features of our world, then no considerations about the designation of terms such as “consciousness” can do anything to rule out the *metaphysical* possibility of the world (p. 134, my italics).

This is puzzling, for although he has argued that the space of possible *worlds* remains the same regardless of whether we are discussing metaphysical or logical possibility of *statements*, that possibility of worlds is logical according to him, *not* metaphysical. It is hard to see how a logically possible world, even correctly “re-described”, in which watery stuff (and not water) is XYZ, could be a candidate for *metaphysical* possibility. While it may be logically possible that there could be watery stuff that is XYZ and not H₂O, it is certainly not metaphysically possible according to Chalmers’ *own* definition of the term, which he equates with 2-possibility and thus secondary intensions. A *posteriori* considerations (the secondary intension of water) tell us that water is H₂O, not XYZ. Thus, contrary to what Chalmers asserts, it seems clear even according to his own framework that *a posteriori* necessities do put constraints on the *metaphysical* possibility (2-possibility) of any worlds.

The case of consciousness is no different. We may say what we like when evaluating according to primary intensions – “consciousness-like stuff is not physical” – and this world, according to Chalmers’ framework, will be logically possible. However, when evaluating according to secondary intensions we are completely constrained to whatever it is that the secondary intension of consciousness turns out to be in the actual world, since this is what it is across all possible worlds. Thus possible worlds where consciousness-like stuff is stated to be something other than the actual secondary intension of consciousness (whatever that is) are according to Chalmers’ *own* framework logically possible (1-possible), but *not* metaphysically possible (2-possible).

Metaphysical possibility, according to Chalmers’ interpretation of the notion, is a logical possibility concerning *a posteriori* aspects of the concepts involved and such aspects concern extensions in the *actual* world – the secondary intension is based on what the concept turns out to refer to in the actual world. So by making a claim for the metaphysical possibility of the zombie world (quoted in the previous paragraph),

Chalmers seems to be taking an unwarranted step from logical possibility to conclusions about the actual world (*a posteriori* considerations).

But perhaps Chalmers is simply excluding *a posteriori* considerations altogether from such a possible world. For example, maybe he is suggesting that there is no water whatsoever in a possible world where *watery stuff* is XYZ, and thus *a posteriori* considerations do not come into the equation at all. That is, in such a 1-possible world there is only a water-like substance (XYZ) but no H₂O, and thus no water. Therefore we do not need to concern ourselves with *a posteriori* considerations of what water should be (H₂O) because water (H₂O) does not feature in such a world at all, so we can say what we like about watery stuff. Perhaps in this sense he believes that such a world could be claimed to be metaphysically possible (in his terms) because there are no *a posteriori* considerations at all so there is nothing at that level to rule out that world's metaphysical possibility (i.e. no claims of "water is XYZ", only of "watery stuff is XYZ"). But regarding any discussion against a materialist view of consciousness, excluding *a posteriori* considerations seems question-begging since it is these which are of central concern: i.e. what consciousness is, its secondary intension.

However, Chalmers has another point to make with regard to the secondary intension of consciousness. He claims that the "irrelevance of *a posteriori* necessity can be further supported by the observation that with consciousness, the primary and secondary intensions coincide" (p. 133). We describe conscious experience by its phenomenal feel, so that its primary intension simply is this phenomenal feel. Chalmers argues that its secondary intension is *also* phenomenal feel, as he says "what it takes for something to be a conscious experience in a counterfactual world is for it to have a phenomenal feel" (ibid.). He illustrates this by considering the concept of water again. There could be something that looks and feels like water (watery stuff) in a counterfactual world and thus accords with the primary intension of the concept. But

this watery stuff could be XYZ in the counterfactual world and so not accord with the secondary intension (H₂O) of the concept. However, he claims, this is not possible in the case of phenomenal consciousness since “if something feels like a conscious experience, even in some counterfactual world, it *is* a conscious experience”¹³ (ibid.). This is because, he argues, “All it means to be a conscious experience, in any possible world, is to have a certain feel” (ibid.).

However, granted that the primary intension of consciousness is its “phenomenal feel”, then to determine its secondary intension, according to Chalmers’ own proposed intensional framework, we must *first* evaluate “the primary intension at the actual world” and *then* rigidify this evaluation “so that the same sort of thing is picked out in all possible worlds” (p. 59). However, *evaluating* the “phenomenal feel” aspect (primary intension) of consciousness at the actual world is exactly the problem at hand. What consciousness actually is, *a posteriori* (its secondary intension), has not yet been satisfactorily determined. This *could* turn out to be very different from its primary intension, we simply do not know at the moment. Chalmers argues that with consciousness the secondary intension is the same as the primary intension, but the secondary intension determines a reference in counterfactual worlds *given that reference in the actual world is already fixed*. So if one follows Chalmers’ *own* description of how the secondary intension of a concept is determined, then the resulting secondary intension of consciousness does *not* necessarily coincide with its primary intension, as Chalmers claims. It may *seem* to coincide because in the case of consciousness it may appear, as he argues, that there is no appearance-reality dichotomy, as there is for example with water and most other natural kind terms. However, even if we were to agree that there is no such dichotomy with consciousness, this still does not mean that

¹³ Chalmers notes that Kripke also made a similar point. This point was made in Lecture III of *Naming and Necessity*.

one can imagine its underlying nature to be anything at all. That is to say, this does not mean that one can ignore *a posteriori* considerations.

In the same way, even though we can think it logically possible (1-possible) to have a world with watery stuff that is XYZ and not H₂O when evaluating according to primary intensions, this does not imply that it is a logical possibility with regard to *a posteriori* considerations (2-possible, also a logical possibility according to Chalmers). Evaluated according to secondary intensions water is not XYZ but is H₂O. One could conclude that Chalmers' notion of 1-possibility (logical possibility) is really just one of epistemic, or apparent, possibility. It may seem like the statement "watery stuff is XYZ" is logically possible but this is only superficial for, when we examine it further, we have no idea how XYZ (whatever that is!) could explain the properties of watery stuff (stipulated as similar to water). By separating out the two intensions of a concept Chalmers has created a framework which he believes enables him to ignore *a posteriori* considerations altogether. However, if the Kripkean analysis of *a posteriori* necessities holds any weight, then it is questionable whether one can completely ignore *a posteriori* considerations in this way, especially when it is exactly *these* considerations (the underlying nature of consciousness) which are at the centre of the debate about phenomenal consciousness.

CHALMERS VS. KRIPKE

Interestingly, Kripke's analysis of the way we can conceive things to be other than what *a posteriori* necessity would dictate, is somewhat analogous to Chalmers' manner of describing possibilities in terms of the intensions of a given concept. As described in the last section, Chalmers explains the apparent contradiction between the fact that we seem to be able to imagine possible worlds where certain *a posteriori* necessary identities do not hold and the fact that such identities should hold in all

possible worlds, by claiming that the concept is analysed by its primary intension in the former case and by its secondary intension in the latter case. For example according to Kripkean *a posteriori* necessity water is necessarily H₂O, yet we seem to be able to imagine a possible world where water is not H₂O. This, Chalmers argues, is because we are using the primary intension of the concept water, which, as explained previously, is something like watery stuff (clear, drinkable, liquid, etc.). Thus, he says, we are not really imagining a world where water is not H₂O but rather one where there is some water-like stuff that turns out not to be H₂O. Such a description appears similar to Kripke's notion that one could be in a "qualitatively identical epistemic situation" where one believes one is faced with water which is not H₂O, but, in fact, the term "water" in such a world refers to something other than water in the actual world (as it does not refer to H₂O) – that is, the usage of the term has changed.

Both Kripke and Chalmers are effectively giving the same description of how it occurs that we can think it possible that an *a posteriori* necessary truth (e.g., water is H₂O) might not hold: ultimately *because we are mistaken*. As Chalmers puts it, we are "misdescribing" that possible world if we say that water is not H₂O, because it is actually watery stuff, and not water, that is not H₂O. Kripke says that in such a case the term "water" is not being used in the (correct) way we use it in the actual world but is being applied to some *other* substance. When we think in terms of secondary intensions of the concept, according to Chalmers, then water will always be H₂O in every possible world. Similarly, Kripke implies that when we use the term "water" properly, in the way we use it in the actual world, then water will always be H₂O in every possible world.

So where Kripke's and Chalmers' views diverge seems to be in the area of the *a priori*. As earlier pointed out, Kripke associates the *a priori* with the "epistemological", saying that "the notion of a prioricity is a concept of epistemology" (Kripke, p. 34). Thus the *a priori* deals with what can be *known* and not as such with what *is*,

ontologically speaking. Necessity, according to Kripke, can be used in an epistemological way, but he is interested in necessity (and therefore possibility) of a metaphysical nature. So for Kripke *a priori* and necessity deal with different domains, the epistemological and the metaphysical, and are not necessarily tied to each other. Chalmers, on the other hand, wants to maintain a link between *a priori* and necessity, as he says he uses “*a priori* methods to gain insight into necessity; [and] this is the sort of thing Kripke’s account is often taken to challenge” (p. 56). But he believes that his two-dimensional framework of intensions of concepts can circumvent this problem. However, as I pointed out in the previous paragraph, Chalmers’ analysis of the kind of possibilities we seem to have, depending on which intension of a concept we use when conceiving, nevertheless appears to be very similar to and to give the same conclusion as Kripke’s analysis. But it seems strange that, despite the similar analyses and conclusions described in the previous paragraph, Kripke can maintain that there is not a link between *a priori* and necessity whereas Chalmers can maintain that there is! I will explore this further in the following section.

TRUTH IN VIRTUE OF MEANING REGARDING THE *A PRIORI* ASPECT OF A CONCEPT

I think the main reason that Chalmers’ and Kripke’s analyses above although initially very similar, diverge considerably when it comes to the notion of a link between *a priori* and necessity, is based in Chalmers’ assertion that his analysis of a concept in terms of his two-dimensional intensional framework gives him “truth in virtue of meaning”, particularly with respect to the *a priori* aspect of the concept. As I outlined in previous sections, Chalmers says that the primary intension of a concept backs an *a priori* truth (e.g., water is watery stuff) which is true in virtue of meaning. Similarly, the secondary intension of a concept backs an *a posteriori* truth (e.g., water is H₂O) which is

also true in virtue of meaning. Of course the latter truth is a necessary truth according to Kripke as well. So with respect to *a posteriori* concerns both Chalmers and Kripke seem to agree – at least Chalmers has incorporated Kripke’s account of *a posteriori* necessity into his own framework of intensions. It is in his notion of the status of the *primary intension* or *a priori* aspect of a concept that I believe Chalmers comes to some conclusions contrary to the Kripkean account, as Chalmers argues that this *a priori* aspect of a concept *also* gives necessary truth. He claims that with respect to the primary intension, “a statement S is necessarily true if no matter how the actual world turns out, it would turn out that S was true” (p. 63). Chalmers has defined the primary intension of a concept as picking out the referent of that concept in a given world considered as actual: “In a given world, it picks out what the referent of the concept would be if that world turned out to be actual” (p. 57). So he believes that he can then conclude that the primary intension of a concept gives necessary truth, as for example in the case of water: “according to this construal on which possible worlds are *considered as actual*, ‘Water is watery stuff’ is a necessary truth” (p. 63).

However, I would query Chalmers’ claim that he can get *necessary* truths from *a priori* considerations in this manner – that is, by analysing the concept in terms of its primary intension. As explained in chapter one, Chalmers introduces the notion of intension to correspond, roughly, to the meaning of a concept. He argues that if we equate intensions with meaning in such a way, then both intensions will “back” a certain kind of conceptual truth or “truth in virtue of meaning”. But it is questionable whether intensions as Chalmers describes them *are* to be equated with the meaning of a concept. Chalmers anticipates such an objection when he concedes that meaning “is a many-faceted notion, and some of its facets may not be perfectly reflected by intensions, so one could resist the equation of the two at least in some cases” (p. 66). But he argues that the equation of the two can be thought of as “stipulative” and says: “In any case,

nothing rests on the use of the word ‘meaning’. It is truth in virtue of intension that I am interested in, whether or not intensions are meanings or not” (note 31, p. 66). This seems to be a rather unsubstantiated claim as we have no reason to believe, nor does Chalmers elaborate much, that his “truth in virtue of intension” is the same thing as “truth in virtue of meaning” and will give us conceptual necessity.

Traditionally, one has “truth in virtue of meaning” of a statement when the predicate is part of the definition of, or analytically true of, the subject. For example, the statement “a bachelor is an unmarried man” is true in virtue of meaning because the term “bachelor” simply *means* “unmarried man”¹⁴. Chalmers himself seems to be proposing something akin to this traditional view since he defines conceptual truth as the notion that a statement be true “by virtue of the meanings of the terms involved” (p. 52). In making the link between “truth in virtue of meaning” and “logical necessity”, he says: “If a statement is logically necessary, its truth will be an automatic byproduct of the intensions of the terms (and the compositional structure of the statement)” (p. 66). So in order to show that the intension of a concept “backs” a *necessary* truth, I believe Chalmers must show that the intension is somehow analytically true (in the above-mentioned traditional “pre-Quinean” sense) of the concept in question. At least this seems to be his claim when he says that such a statement (e.g., “water is watery stuff”) gives a conceptual truth (“truth in virtue of meaning”, by his own definition) or is logically necessary: the truth of the statement is an “automatic byproduct” of the intensions. Certainly, an intension may well be true of its concept but is it *necessarily* true? In the following paragraphs I will examine whether Chalmers shows this in relation to the primary intension, as it is this *a priori* aspect of a concept in terms of which he wants to frame his argument against materialism.

¹⁴ Of course not all philosophers would agree with this: some, like Putnam, arguing that this is really truth in virtue of convention.

The intension of a concept as Chalmers defines it, “a function ... from possible worlds to referents”, seems very similar to Kripke’s notion of “fixing a reference”. If we consider the primary intension of a concept in particular, this corresponds to our *a priori* notion of the concept, according to Chalmers. He gives the example of water of which the primary intension roughly “picks out the dominant clear, drinkable liquid in the oceans and lakes” or “watery stuff” (p. 57). In the case of a possible world where water is XYZ instead of H₂O, Chalmers says that the primary intension of water “maps” it to XYZ in the XYZ world. Such a function (primary intension) seems to be the same as what Kripke calls “fixing the reference” of the term “water” in that possible world. The main difference here is that Kripke really only speaks about fixing references in the *actual* world and uses the actual world as his frame of reference when talking about other possible worlds. In fact, Chalmers does define primary intension as something “that fixes the reference in the actual world” (p. 59), so it would seem that the primary intension does substantiate the role of reference fixer. And as I mentioned earlier, Kripke argues that fixing the reference of a term is not the same as giving its meaning. He points out that descriptions of a term (such as Chalmers gives for the primary intension of water, abbreviated to “watery stuff”) cannot be synonymous with that term because they do not *uniquely* describe the referent of the term and could indeed pick out different referents in other possible worlds. This is exactly what Chalmers’ primary intension does: it picks out XYZ for water in the XYZ world and H₂O in the actual world. So it is not what Kripke would call a “rigid designator”. Yet it is the means by which we determine the referent of a term in the actual world, what Chalmers calls the secondary intension of a term. He says the secondary intension of a term, Kripke’s *a posteriori* necessary identity, is determined by first evaluating the primary intension at the actual world.

Arguably the secondary intension of a concept is a necessary truth for all the

reasons that Kripke has given for *a posteriori* necessity. When we find out water is H₂O, we have found out what water actually *is*. However, the primary intension seems to be less straightforward. On the one hand, as I explained above, it is not a rigid designator so will not pick out the same referent for a given concept in every possible world. So in what sense could it be said, as Chalmers claims, to “back” a necessary truth? It does not appear to act like the meaning of a concept, for the primary intension seems to act like a cluster of descriptions with which we fix the reference of the concept. If it were seen as the meaning of the concept, since it picks out different referents in different worlds, it would change the meaning of the concept in different possible worlds. One could argue that this would only change the *referent* of the concept and not its meaning but it seems unlikely that the meaning of a concept is completely independent of its referent. For example, what water is, its secondary intension, also contributes to our interpretation of the term and therefore its meaning. So the primary intension is of course to some extent involved in the meaning of a concept but it is only part of the story. If it picks out different referents in different possible worlds then this will change the overall meaning of the concept for us. Even in the actual world, the primary intension of the concept of gold (hard, yellow metal, etc.) can also pick out “fool’s gold”, which is a completely different thing and does not have the same meaning as “gold” at all. Therefore in this sense the primary intension does not seem to back a *necessary* truth because it is only part of the meaning of the relevant concept and indeed can sometimes be wide of the mark.

At any rate Chalmers argues against viewing the primary intension as a cluster of descriptions and claims: “descriptions play no essential part in this framework...It is the function itself, rather than any summarizing description, that is truly central” (p. 58-9). He says further that “this picture is quite compatible with the ‘causal’ theory of reference” and that the primary intension “may require an appropriate causal connection

between the referent and the subject” (ibid.). So it appears that he is suggesting that the primary intension function does its work via causal connections although he is not by any means committing himself to this suggestion. But even with a function working on causal connections, some sort of set of descriptions is arguably needed because such is the way we recognise the same entity again in our environment. If we are to consistently pick out water via its primary intension (say we do not yet know its secondary intension) then its properties (or a core group thereof) must be consistently similar otherwise we could not pick it out each time. Implicitly this implies a core cluster of descriptions making up the function (primary intension) by which we pick out water in our environment. This does not have to be a well-defined set of descriptions – we can allow for quite a bit of “looseness” here – but it is hard to see how a function like the primary intension could pick out any referents without at least implicitly working with descriptions. The fact that the primary intension *does* pick out different referents depending on which world we are considering as actual also seems to indicate that it is functioning like a cluster of descriptions. So in this sense it does not seem to be a candidate for backing “truth in virtue of meaning”.

However, Chalmers claims that “truth in virtue of intensions” is all he needs. But in order that the primary intension back *a priori* conceptual truths as he claims, it still must be *necessarily* true of the relevant concept – e.g., the statement “water is watery stuff” must be a necessary (*a priori*) truth. Chalmers does not directly argue for this but instead tries to point out the importance of the primary intension for explanation:

Before we even get to the point where rigid designation and the like become relevant, there is a story to tell about what makes an actual-world *X* *qualify* as the referent of “*X*” in the first place. This story can only be told by **analysis of the primary intension**. And this project is an *a priori* enterprise, as it involves questions about what our concept *would* refer to if the actual world turned out in various ways (p. 59, my bold type).

If we consider the case of water again, Chalmers is claiming that analysis of the primary

intension of the concept of water tells us what makes “actual-world” water eligible to be the referent of “water”. The referent of water in the actual world is H₂O, the *a posteriori* secondary intension according to Chalmers. When we discover this, of course we do expect that the properties of H₂O will give us the characteristics of water (wet, liquid, etc.), as outlined by its primary intension. But such an analysis of the primary intension will not tell us everything. Once we discover that water is H₂O, the essence of the meaning of “water” shifts to this molecular structure because this secondary intension will fully explain the primary intension and more: e.g., the behaviour of water in theoretical situations never experienced before and therefore of which we have no *a priori*¹⁵ notion. If H₂O is found to have some properties that do not agree with our previous *a priori* notion of water, then it is this *a priori* notion (primary intension) that has to change and not the secondary intension. This is because the primary intension of a concept is a notion based on our *experience* of the referent of the concept – i.e. how we have found it to be so far – and therefore open to being erroneous.

We generally come to know water in our environment by ostension and use, and gradually we form what Chalmers calls our *a priori* notion of it (watery stuff) as having certain properties and functions (wet, quenches our thirst, etc.). It is this notion that forms the primary intension of our concept of water. This primary intension is *a priori* in the sense that it develops independently of the *a posteriori* discovery that water is H₂O and provides the criteria for the application of the term “water” at a “commonsense” level (before we know what the underlying nature of water is). So we use this *a priori* notion to evaluate whether something in our environment is water or not, relying on the expectation that the core functions and descriptions of the primary intension will be fulfilled if the phenomenon in question is indeed water. However, the functions and descriptions that go towards making up our *a priori* notion of water will of

¹⁵ As I noted in chapter one, this must be a *relative* sense of *a priori*: that is, our notion of the concept before we know what its secondary intension is.

course vary, depending on our environment. For example, in the past, people living in an environment where water never froze, say near the equator, would probably not recognise water in an environment much further north where it regularly froze. Thus the notion of water freezing would not form any part of the primary intension of their concept of water. So analysis of *their a priori* notion (in Chalmers' sense of *a priori*, again) of what water is would fail to tell them that snow and ice were indeed (more or less) water. This seems to indicate that the primary intension is a subjective notion insofar as it would appear to be dependent on the knowledge of a given individual or community. Chalmers acknowledges the dependence but argues that "this sort of problem is irrelevant to the issues" with which he will be concerned and maintains that "we might as well assume ... that all individuals [of a community] are equally well informed" (p. 58). But any community's experience is still limited and so it would appear that the primary intension can be partial or even incorrect in some aspects and therefore does not seem to back a *necessary* truth.

As I mentioned earlier, Kripke has pointed out that the descriptions we use to fix the reference of a term are themselves often *contingent* properties of that term. Hilary Putnam (1975) in his paper "The meaning of 'meaning'", also makes a similar point about descriptions like those that make up Chalmers' primary intensions, although Putnam does not talk in terms of intensions but rather in terms of stereotypes. While he acknowledges that such descriptions are part of the meaning of a term, he points out that with respect to the term "tiger" for example, "it does not follow, in our view, that 'tigers are striped' is analytic" (Putnam, p. 256). He goes on to say that communication presupposes the possession of such stereotypes (rather like primary intensions), such as tigers having stripes, but "it does not presuppose that any particular stereotype be *correct*" (ibid.). So it seems doubtful that such descriptions that go to make up the primary intension of a concept are in any way logically necessary of the concept. They

are often only *contingently* true, and indeed may not be true of the concept in other worlds or even in different parts of the same world. Therefore I would argue the primary intension does not back a *necessary* truth with respect to that concept.

Kripke gives an example of an *a priori* description, somewhat similar in kind to Chalmers' primary intension, that at first sight might appear to be necessarily true. He talks of a stick used as a standard measurement of a metre, so that the stick *S* is exactly one metre long (he adds "at time t_0 " for accuracy). He notes that some people would think that "stick *S* is one metre long at t_0 " is a necessary truth because it gives the *definition* of a metre. However, he points out that such a definition is not used to give the meaning of what a metre is but to *fix the reference* (i.e. it is the length of stick *S*). Furthermore this reference, the length of a metre, is fixed by an *accidental* property of the length of a particular stick – the stick does not *necessarily* have such a length. Kripke observes that "there is an intuitive difference between the phrase 'one meter' and the phrase 'the length of *S* at t_0 '" (Kripke, p. 55). He says: "The first phrase is meant to designate rigidly a certain length in all possible worlds, which in the actual world happens to be the length of *S* at t_0 " (ibid.). However, he points out that the length of *S* (at time t_0) can be known *a priori*, if for example someone decides that they shall fix the reference of one metre in this way, then "he knows automatically, without further investigation, that *S* is one meter long" (Kripke, p. 56). Thus this is the epistemological status of the statement "stick *S* is one metre long at time t_0 ": it is knowable in an *a priori* way. However, Kripke points out, the *metaphysical* status of the statement is that of a *contingent* statement, because "under appropriate stresses and strains, heatings or coolings, *S* would have had a length other than one meter even at t_0 " (ibid.). So Kripke concludes that there are *contingent a priori* truths.

Let us compare Kripke's example of an *a priori* notion of a standard for the length of a metre to Chalmers' example of our *a priori* notion of what water is. The

former would seem to be a *stronger* candidate for providing a *necessary a priori* truth, for the simple reason that we specify everything about it in advance: i.e. that the stick *S* will be one metre long at time t_0 . However we have seen above that even the former is only a *contingent a priori* truth according to Kripke, as the length of the *S* is a contingent property that can vary in reality. With regard to water, considering its primary intension only (before we have discovered the secondary intension), we discover it to be watery stuff (liquid, drinkable, clear, etc.) and this notion forms our concept. We do not specify in advance, as we do with the length of *S*, that water is to have any particular properties – we discover *empirically* that it does have those properties. So the primary intension of water is not an *a priori* specified definition, as in the case of stick *S*, and thus appears to be even *less* of a candidate for a *necessary a priori* truth. Also, the properties of the phenomenon water that constitute its primary intension (“clear drinkable liquid, etc.”) are accidental properties. In another possible world, for example, our biology could be different and water might not be drinkable for us. So Chalmers’ proposal for the primary intension of a concept is a notion involving *contingent* properties of that phenomenon and also seems to be a notion that is empirically discovered and thus open to being erroneous. Therefore it is hard to see how he can get necessity from such a notion.

Chalmers nevertheless proposes that Kripke’s contingent *a priori* is in fact necessary (1-necessary) when worlds are *considered as actual*. So, in the example of a world that has XYZ instead of H₂O, he claims that if the “actual world turns out to be a world in which watery stuff is XYZ, then my statement ‘XYZ is water’ will turn out to be true” (p. 63). Thus he concludes, such a construal makes “water is watery stuff” a necessary truth. However it seems that all he really can conclude from his example where the XYZ world is construed as actual, is that “XYZ is *watery stuff*” and *not* that “XYZ is *water*”. In fact he himself later says of the XYZ world that “Kripke’s analysis

shows us that due to the way the actual world turns out, we are misdescribing this world as one in which XYZ is water” (p. 134). He adds: “Strictly speaking, it is a world in which XYZ is watery stuff” (ibid.). Therefore, following the same logic, it would seem that the statement “water is watery stuff” is not true in the XYZ world because there is no *water* (H₂O) in that world. Thus “water is watery stuff” is not a necessary truth. So it appears to be the case that the *a priori* aspect alone may not “back” a conceptual truth.

Effectively, I have been arguing that Chalmers’ thought experiments about the truth of statements in possible worlds, where the primary intension of the concept involved is considered, seem to be cases of what Kripke calls “qualitatively identical epistemic situations”. So that when we say of the XYZ world that “water is XYZ” we have in fact changed our normal usage of the term “water”. But, as Kripke pointed out, we must describe possible worlds in terms of our “normal” language usage, otherwise we would have no consistent way of referring to such worlds. Even Chalmers acknowledges this by saying the correct description of such a world is that *watery stuff* (and not water) is XYZ. However, this shows that “water” and “watery stuff” are *not* synonymous (because they are not interchangeable in such a case) even according to Chalmers’ own interpretation, as otherwise we have a “misdescription” according to him. According to Quine (1980), synonymy is one way for determining logical necessity of statements as he says that “such a statement [no bachelor is married]...can be turned into a logical truth by putting synonyms for synonyms” (Quine, p. 23). So on this basis, it does not appear that the statement “water is watery stuff” is logically necessary.

The primary intension “watery stuff” may indeed *track* the meaning of the term “water”, for it is in a sense *stipulated*: if water had completely different properties, hard, metallic, green (say), then its primary intension would reflect this and thus watery stuff would pick out hard metallic green stuff. Then the statement “water is watery stuff”

would still be true, but only a *partial* truth because watery stuff does not fully explain water, the secondary intension is also needed. Also the primary intension of a concept, as our experiential notion of the concept, is based on our experience which is limited and can vary, and so may be mistaken – as we sometimes find out when we discover at a deeper level what the phenomenon is (secondary intension). Seen in this the light, the primary intension is only part of meaning of a concept and is also fallible and thus could hardly be said to “back” a *necessary* truth. Therefore I believe that Chalmers cannot substantiate his claim that the primary intension backs a conceptual (logically necessary) truth. If this is the case, then Chalmers has not managed to maintain a link between the *a priori* and necessity, since his primary intensions are contingently true of their concepts.

So if, as I have tried to show above, the primary intension of a concept does not “back” a necessary truth, then I would argue that when Chalmers talks of logical possibility according to the primary intension of a concept what he is really talking about is epistemic possibility – what *seems* to be possible *before* we have further information (before we know what the secondary intension of the concept is). If there is no logical necessity in the relation between a concept and its primary intension, then possibilities regarding that concept analysed in terms of its primary intensions are arguably not logical possibilities because analysis at the level of the primary intension carries no logical entailment.

I would propose that the realm of the *a priori* cannot do the work that Chalmers requires of it and that even following Chalmers’ own framework, it is the secondary intension that determines what is or is not logically possible. For example, before we know that water is H₂O, we could imagine it possible that water could be XYZ, but once we find out that it is in fact H₂O then this conceivability no longer implies possibility – logical possibility is curtailed by *a posteriori* considerations. Certainly it no longer

implies metaphysical possibility in Chalmers' terms (2-possibility, also a logical possibility according to him) – water cannot be other than what it is. If one thinks that it could still imply logical possibility (water could logically have been something else) then one is faced with all the plausible Kripkean arguments in support of *necessary a posteriori* identities: if water were to *be* something other than H₂O then it would not be water and we would be talking about something else. Even Chalmers accepts this *a posteriori* necessity in his analysis (however he relegates it to analysis by secondary intensions which he says is less relevant to explanation than analysis by primary intension). This is why, I believe, Kripke talks about *epistemic* situations in such cases (when imagining water to be something else) as opposed to talking about *logically* possible situations – because estimation of such possibilities is constrained by our *knowledge*. In the following section I will argue against Chalmers' claim that it is the *a priori* level which is the most important for (reductive) explanation.

EXPLANATION

In answer to any objections that he should be addressing *a posteriori* aspects, Chalmers claims that we can effectively ignore the secondary intensions and just follow his argument through in terms of primary intensions (in which terms it is accordingly proposed) because “it is the primary intension that is most relevant to explanation” (p. 132). All we have to do, he says, is to show that it is logically possible (i.e. not logically impossible) that there be a world physically identical to ours but lacking any phenomenal side and “dualism will follow”. This, Chalmers claims, is exactly what his zombie argument has established.

In particular Chalmers claims that it is the primary intension of a concept that is important when it comes to a *reductive* explanation of the concept, which is the type of explanation of phenomenal consciousness against which he is arguing. He says that it is

“logical supervenience according to a primary intension that determines whether reductive explanation is possible” (p. 69) and that, therefore, this is the kind of logical supervenience of the phenomenal on the physical that he will be refuting. Of the secondary intension of a concept he maintains that it (e.g. H₂O) “does not emerge until after an explanation is complete, and therefore does not itself determine a criterion for explanatory success” (p. 69). For example, he says, before we know that water is H₂O, it is the primary intension that “determines whether or not an explanation is satisfactory” because “we have to explain things like its clarity, liquidity, and so on” (p. 69).

However, I would argue that it is actually the secondary intension which does the explanatory *work*, especially with respect to reductive explanation. The primary intension of water tells us, for example, that water is clear, drinkable, liquid, etc., thus enabling us to identify water in our environment. This intension does not tell us *why* water is clear, liquid, etc., only that it (generally) *is* so. But when we discover the chemical composition of water, H₂O, then we learn why it is that water turns solid when it freezes, for example. The chemistry (and perhaps physics) of the H₂O molecule will *explain* that mechanism for us. Previous to discovering the secondary intension of water, we had no explanation of why water was solid at some temperatures and liquid at others. Similarly, for example, with the phenomenon of heat: once we discover what heat is (its secondary intension) then we understand *why* it acts in the way it does (for example, how heat transfers from one body to another). Presumably this is why Kripke attached so much importance to these *a posteriori* identities, because they tell us what the phenomenon *is*, thereby also giving an *explanation* of why it has certain properties. So, on this basis, I would strongly dispute Chalmers’ claim that it is the primary intension that is the more important for explanation, in particular reductive explanation.

In support of his argument, Chalmers says that, with respect to heat: “*explaining* heat involves explaining the fulfillment of the causal role [*a priori* notion], rather than

the motion of molecules [*a posteriori* notion]" (p. 45). In other words, he is saying that it is our *a priori* notion of what heat is (its causal role in the world) that must be explained. Presumably it is in *this* sense that he means that it is the *a priori* notion of a concept that is most important for explanation because it is this which must be explained. However, it is nevertheless our *a posteriori* notion of the concept that *does* the explaining: it is only once we know what the phenomenon actually *is* that we have an explanation of how it fulfils its causal role *and* that we can see the *conceptual* links between what the phenomenon is and its causal role in the world. Chalmers himself acknowledges this to some extent:

The concept of heat that we had *a priori* – before the phenomenon was explained – was roughly that of “the thing that plays this causal role in the actual world.” Once we discover how that causal role is played, we have an explanation of the phenomenon. As a bonus, we know what heat *is* (p. 45).

So it is true that it is the *a priori* notion of a concept which we want explained, and in this sense it is the notion most relevant to explanation as Chalmers claims. But the *only* notion we have initially of a concept before we discover *a posteriori* what it is, is this *a priori* notion, so of course that is the notion we want explained. And in that sense it is the notion most relevant to explanation *at that point in the investigation*, simply because we have no other (*a posteriori*) notion of the concept at that point in time. But I would argue that the explanatory *power* lies in what Chalmers calls the “low-level facts” (that water is H₂O, that heat is molecular motion, etc.) because these give us the “hows” and “whys” of the phenomenon: they *explain* the higher level facts, or *a priori* aspect of the concept. Until we start to piece together an (*a posteriori*) explanation of approximately how (and therefore, that) the low-level facts lead to the high level (*a priori*) facts, we do not have any *conceptual entailment* from the former to the latter. So contrary to what Chalmers claims, I would argue that conceptual entailment between the high-level facts of a particular phenomenon and its underlying low-level facts comes mainly through the

process of empirical discovery of what that phenomenon is – i.e. the secondary intension of the relevant concept. At the very least such conceptual entailment appears to be *as* dependent on secondary intensions as it is on primary ones for its development.

In fact, often the very process of discovering what a phenomenon is (the secondary intension of the related concept) results in the modification of our initial notion of the phenomenon, the primary intension of the concept. For example, in previous centuries it was part of our *a priori* notion (primary intension) of the mind that a particular kind of mad or inappropriate behaviour was often seen to mean that that person was possessed by demons or some such similar thing. Nowadays however, armed with *a posteriori* knowledge (secondary intension) of (certain aspects of) the mind, we can interpret what was previously known as “possession” as mental illness, and we know for example, that there could be a chemical basis to it which can be treated by drugs. Thus, in Chalmers’ terms, the primary intension of this aspect of the concept of mind has been modified to accommodate truths uncovered by the discovery of its secondary intension. The initial *a priori* interpretation of the associated behaviour was simply incorrect and it was the *a posteriori* discovery of the basis of this behaviour that showed that our *a priori* notions were wildly off the mark.

Of course Chalmers is right in suggesting that it is the associated behaviour (*a priori* notion) that needed explaining but I believe it is obvious from examples like this that conceptual entailment does not come purely, or even mainly, from this *a priori* level. And indeed, *a posteriori* discoveries have repercussions on our *a priori* notions and so these two aspects of a concept are really rather *mutually dependent*, contrary to Chalmers’ attempts to handle them independently. It can of course be an indication that a concept can be reductively explained if its primary intension can be analysed in functional terms, as Chalmers suggests. But he claims that “if B-properties are logically supervenient on A-properties according to primary intensions, then the implication from

A-facts to B-facts will be *a priori*” (p. 70). However, sometimes we need to completely change our (so-called) *a priori* notion of the concept before such entailment can be accommodated, and this change is often initiated by various *a posteriori* discoveries. It is, for example, through a long process of various *a posteriori* discoveries of the physical underpinnings of biological life that we started to develop the (*a priori*, in Chalmers’ sense) conceptual framework that describes biological life in functional and structural terms (which are then in turn reducible to physical ones) in the first place.

But this is not to say that Chalmers’ notion of the *a priori* aspect of a concept, its primary intension, is not at all important in discovering conceptual entailment from the low-level facts to the high-level facts. It is just questionable whether *so* much weight should be given to this *a priori* aspect, as Chalmers seems to want to do when he argues that “explanatory connections are grounded in *a priori* entailments from physical facts to high-level facts” (p. 98). He supports his argument that primary intensions are the more important for reductive explanation, observing that since we have “the ability to know what our concepts refer to when we know how the actual world turns out, then we have the ability to know what our concepts would refer to *if* the actual world turned out in various ways” (pp. 59-60). In the case of consciousness though, we do not yet know how the “actual world turns out” in that respect – or, arguably, even what it would take for something to be a conscious state – this is exactly the point in question. So when the question of the nature of consciousness is still unresolved in the *actual* world then it is doubtful whether we can make many useful estimations at this *a priori* level in other possible worlds, beyond an educated guess. Even in the case of a phenomenon well known and understood in the actual world, how are we to know that we are making correct estimations with regard to other possible worlds? We can certainly suggest that water be XYZ in another world but this does not even show, much less *prove*, that it is possible. Even more so with regard to phenomena whose *a posteriori* identity has not

yet been established: judgements in terms of the *a priori* aspect of the concepts involved would appear to be rather speculative. And, while perhaps informative and helpful, such speculations certainly would not seem to be eligible to carry a burden of *proof* of any particular theory about that phenomenon.

I would argue that until we at least *begin* to discover what the secondary intension of a given concept is, we generally have little conceptual entailment between its primary and secondary intensions. For example, if we do not know that water is H₂O, then we have no conceptual link to its primary intension, watery stuff, from anything physical as the primary intension does not give us a conceptual link. So it may *appear* that a reductive explanation of a particular phenomenon is not within our grasp because we cannot see how the higher-level facts could be entailed by the lower-level facts (because we have no conceptual entailment), but this could be due to our current ignorance. (This harks back to my point that Chalmers' logical possibility in the *formal* sense – no logical contradiction - is arguably really only epistemic possibility based on our current lack of the relevant knowledge.) And arguably our current state of ignorance is precisely that we do not know what the secondary intension of the given concept is – that is, what the phenomenon *is*. *Before* we know what water is, it *seems* logically possible that water could be XYZ. However once we *know* that water is in fact H₂O then it becomes logically *impossible* that it could be anything else. Chalmers wants to say that it is *still* logically possible, when we analyse water by its primary intension (watery stuff), that water could be XYZ. However, as I have explained previously, the only logical possibility here is that *watery stuff* is XYZ – it is simply not logically possible that the watery stuff actually be water in this situation (even according to Chalmers' own analysis) because we now know that water is H₂O.

Some might still argue in a similar vein that it is only physically or metaphysically impossible but not *logically* impossible that water could be anything

else, however I wish to make two points against this. Firstly, if we accept Kripkean *a posteriori* necessary identities, then once a concept has been identified to *be* a certain phenomenon (e.g., water is H₂O) it becomes logically impossible in any sense for it to be anything else, because arguably that would be claiming that *A* is not-*A* which is a logical contradiction. Secondly, because the secondary intension (*a posteriori* identity) of a phenomenon explains that phenomenon, it is also arguably logically impossible that the phenomenon be anything else unless that “something else” (e.g. XYZ) can logically account for the phenomenon (in the way, for example, H₂O can explain water).

Of course it could be argued against my first point that discovering, say, that water is H₂O is equivalent to proposing that *A* is *B*, and that there is no logical contradiction in conceiving then that *A* is not-*B*. However, my point is that if we take *a posteriori* identities seriously (if we are swayed by Kripke’s arguments, and in this respect I believe Chalmers is – at least he offers no argument against them) then what we actually discover is *not* that *A* is *B*, but we discover *what A is*. In discovering what the secondary intension of a concept is, we gain a deeper understanding of the relevant phenomenon, we learn what it *is*. So when we then try to say that it is logically possible that, say, water is not H₂O, then we are in a sense trying to claim that it is possible that *A* is not-*A*, a logical contradiction.

With respect to my second point, if we are going to claim that it is logically possible that water is not H₂O, then we have to give some kind of indication at least of how water could possibly not be H₂O. This is because H₂O fully explains the phenomenon of water, as it is simply what water is. Arguably then it is not logically possible that water be anything other than H₂O once we know the explanation because this molecular structure accounts for the properties of water and explains them. In other words, by explaining water in terms of H₂O, a *conceptual* link is established because the latter “low-level fact” shows *how* (and *that*) the former “high-level” fact occurs – we

have no idea of how any other compound could explain water in the way that H₂O does.

This is the *framework* sense of logical possibility I referred to in the beginning of this chapter, when the physical (low-level) properties and laws are known and fixed and logical possibility is bounded by our conceptual frameworks. It may seem like an odd usage of the term “logically possible”, as we do tend to think that logically a phenomenon could have been something other than what it was discovered to be, but I think that necessity of identity does lend credence to the notion that such possibilities as “water is XYZ” are really only *epistemic* possibilities. George Seddon (1972), in his paper “Logical Possibility”, in a similar vein argues that it is not logically possible that certain phenomena be other than what they actually are. He wants to argue that it is *logically* impossible that a bar of iron float on water and notes that in order to show this (or at least one method of doing so), he has to demonstrate that “although the statement that ‘A bar of iron floats on water’ is not explicitly self-contradictory, it is implicitly so” (Seddon, p. 483). He goes on to argue that “we are saying that a mineral with a specific gravity of less than one (i.e. it floats), has a specific gravity in the range 7.3-7.8 (i.e. it is iron), and this is a contradiction, and is therefore logically impossible” (Seddon, p. 483).

Now, if we were to take logical possibility in the *formal* sense, as I described earlier in this chapter, we could argue that there is no apparent logical contradiction in the notion of a bar of iron floating in water – some of the laws of nature could have been different, and so on. Accordingly, Seddon notes that one can argue that the specific gravity of iron is a contingent property and therefore not necessarily true of iron (in other words, the question is: “Is the specific gravity of iron an essential property of iron?”). However, he points out that usually a change in a property like *this* would reflect a change in the underlying atomic structure. As for example with the density of an element, he notes that any change in the density of an element would “not be accompanied by, but in fact *be* either a change in the size of the unit cell or in the

number or kind of atoms of which it is built up” (Seddon, p. 485, my italics). Thus if we discover some strange new water with different properties, as he says we did with “heavy water” (D₂O), then we get a resulting “language shift” because we now have two different things that we mean by the term “water”. But he notes that the “physical properties of H₂O *have not changed*” (Seddon, p. 484, my italics).

Seddon also raises the same question as I have about logical possibility claims being made without any supporting theory or explanation of how the proposed possible situation could obtain. He asks “why suppose that it is possible when *no* context is specified, no theory indicated or evidence cited, without which the probability of the supposition cannot be assessed at all?” (Seddon, p. 490). Arguably the apparent logical possibility of “water is XYZ” is really only an epistemic possibility – it *seems* like water could have been something other than H₂O. To use Chalmers’ (2002) notion, this is only *prima facie*, or superficially, conceivable. If we think *further* about the proposal we must admit that we have no way of even indicating how XYZ could explain the properties of water, or even of watery stuff.

We have certain schemas (often scientific) that explain our world to the best of our knowledge. Arguably either theoretical identifications and explanations have to fit into these schemas or the schema has to change, something which also happens now and again. If we consider the logical possibility of water being something other than H₂O, then within our existing conceptual frameworks, logically this possibility does not make sense. Water is H₂O in our current framework and furthermore its properties are *fully explained* by this fact. Should we propose another chemical composition of water, say CH₄, then the current framework can tell us that *these* molecules have completely different properties to the properties of water and thus cannot do the job of explaining water. Thus either we are saying that something (water) is not what it is (H₂O), which is a logical contradiction, or we have made a “language shift”, as Seddon calls it, and

changed the usage of the term “water” to refer to other things (CH₄, say).

In order to take the proposition of such a logical possibility seriously, we have to respect any implicit logical contradictions and explain, or at least indicate, how the proposed possibility could make sense. Hilary Putnam (1994) suggests something similar in an article entitled “Rethinking Mathematical Necessity” where he discusses the necessity of mathematical truths and the claim by some that even these can be revised. He says that “saying that logic or arithmetic may be ‘revised’ does not have a sense, and never will have a sense, unless some concrete piece of theory building and applying *gives* it a sense” (Putnam, p. 256). (By “sense” of a statement he means “knowing how the words are used in a particular context” (ibid.).) In other words, certain claims of logical possibility (it is logically possible that the truths of logic could be revised, or that water could be XYZ) do not even have a sense for us unless we are also presented with alternative frameworks which would give them a sense for us.

As I mentioned in the earlier section concerning possibility, determining logical possibility even in the “broadest sense”, as Chalmers proposes, must involve some *means* of determination. Whether we are judging logical possibility in the *framework* or in the *formal* sense proposed earlier, this means of determination would seem to be our current conceptual frameworks. (In judging logical possibility in the formal sense, i.e. that there is no logical contradiction in the entertained notion, arguably we still use our conceptual frameworks to find such a contradiction, as these are all we have.) So when we try to judge logical possibilities in an area about which we know little, we do not really have a developed frame of reference against which to make such judgements. If one wants to insist that there is a logical contradiction in Chalmers’ claim that zombies are logically possible, then obviously one will need to employ the relevant conceptual frameworks to show this. But the problem is that these frameworks regarding phenomenal consciousness are still in a state of development and so are not much help to

us. However, by similar reasoning, one cannot employ the relevant conceptual frameworks to show that there is *no* logical contradiction in the notion of a zombie either because these frameworks are not adequate to show it one way or another. So I would argue that any claim made for logical possibility only carries as much weight as the strength of the background framework against which the claim was made. In areas where our knowledge is still quite limited, this is not much weight at all.

In this section I have tried to show that identifying conceptual entailment of high-level facts by the physical (low-level) facts depends at least as much, if not more so, on our knowledge of the secondary intensions of the concepts involved as on our knowledge of their primary intensions (which can sometimes actually change through the discovery of the secondary intension). For example, it is only through the process of discovering *a posteriori* the physical underpinnings of biological facts, such as reproduction, ingestion, etc., that we began to see that the biological facts are *logically* determined by the physical. (In fact, before we knew this, this would have been a controversial claim.) In other words, we must first know (or at least be on the way towards knowing) what the phenomenon is, its secondary intension according to Chalmers, in order to make the conceptual link. This is because, as I have argued, the secondary intension provides an important part of the basis of explanation of the phenomenon. For example, H₂O fully describes the properties, causal role, etc. of water and so once we have the notion of H₂O we have the *conceptual* link to water. At the very least, it would appear that *both* intensions of a concept are needed in uncovering conceptual entailment from low-level facts to high-level ones.

In addition, in this chapter I have tried to show that most concepts do not have logically necessary primary intensions (i.e. “truth in virtue of meaning”). Watery stuff is only contingently true of water, so why should we expect phenomenal consciousness to have any logically necessary relationship to the physical on the basis of primary

intensions alone? Perhaps when, or if, a physical (or indeed, any) secondary intension to phenomenal consciousness is discovered *a posteriori*, then we will see the logical entailment, but, I have argued, we cannot expect to see that now, before we have such knowledge. On the basis then that there is no logical entailment at the level of the *a priori* alone and that the *a posteriori* does much of the explanatory work (thereby giving conceptual entailment), I have argued that Chalmers' claims of logical possibility (1-possibility) are not of logical but of epistemic possibility – i.e. they are highly fallible in areas, such as consciousness, about which we know very little.

Thus in relation to Chalmers' zombie argument against a materialist view of phenomenal consciousness, I propose that all that he can claim is that it is merely epistemically possible that there exist a world physically identical to ours yet lacking phenomenal consciousness (zombie world). The reason that this *seems* possible is due to our current lack of relevant knowledge of what phenomenal consciousness actually is in an *a posteriori* sense. I would argue that the mere epistemic possibility of the existence of zombies (based on our current ignorance of the nature of consciousness) cannot carry the burden of proof against a materialist view of consciousness because this kind of possibility is *itself* fallible. Furthermore, it is only an apparent possibility precisely because of our lack of the relevant knowledge – that is, the secondary intension of phenomenal consciousness. So it seems question-begging to propose that our lack of knowledge, *the very point at issue*, proves a lack of logical supervenience of the phenomenal on the physical. The fact that consciousness at this stage does not *appear* to be of a kind that could be functionally analysed is perhaps a strong indication that it may not be physical, but it is not a proof. All we are entitled to say, I would argue, is that we do not currently *know* whether the phenomenal logically supervenes on the physical. Therefore I would conclude that it appears that Chalmers cannot prove that the phenomenal is not logically supervenient on the physical on the basis of *a priori*

considerations alone.

Having critically examined in this chapter the technical tools of conceivability, possibility, necessity and the *a priori* that Chalmers uses to build up and carry through his zombie argument, in the next chapter I want to look at his argument from another angle. I shall query his view of the functional aspects of consciousness and investigate the coherence of his notion of a zombie.

CHAPTER 3: PHENOMENAL CONSCIOUSNESS

In the last chapter I argued against Chalmers' zombie argument by attempting to show that his claims for the logical possibility of zombies and for the irrelevance of *a posteriori* considerations to his zombie argument were questionable. In this chapter I intend to dispute his zombie argument against a materialist view of consciousness from another angle, exploring the concepts involved. I will query the manner in which Chalmers and certain other philosophers divide the mind into a functional aspect and a phenomenal (and, therefore, non-functional) aspect. I will examine Chalmers' notion of a zombie and argue that the bizarre consequences of such a notion suggest that it is ultimately not a coherent notion. I also try to outline possible functions for phenomenal consciousness, questioning whether we could function as we do without it.

In the beginning of this chapter, I try to clarify what Chalmers means by phenomenal consciousness, comparing his notion with some other definitions. In the first section I examine his division of mental life into phenomenal and psychological aspects, based on qualitative feel and functionality respectively. I argue that placing all functionality of mind in the psychological aspect already prejudices from the start the question of a functional role for the phenomenal. Of course, the basis of Chalmers' dualist argument is that the *essence* of phenomenal consciousness is not functional but rather qualitative. However, I will argue that there may be some plausible functional roles for phenomenal consciousness which could to some extent form the basis of an explanation of experience, even if it is only a partial explanation. By analysing some of his descriptions of psychological consciousness, I try to show that it is not that obvious that the phenomenal aspect can be so easily separated from any functional aspects of mind, as experience still seems to be central to some largely functional notions.

In the next section I examine Chalmers' notion of awareness, again trying to show that experience is in reality central to the notion and that awareness without

experience is something different. Block (1995) proposes a division between the functional and phenomenal aspects of mind similar to Chalmers' proposal. In particular Block's notion of access-consciousness seems to be very close to Chalmers' notion of awareness, as Chalmers himself mentions. I discuss some of Block's examples of the presence of phenomenal consciousness without access consciousness, arguing that these really show the opposite of what Block intended. Thus dividing up consciousness into functional and phenomenal aspects, as Chalmers also does, is by no means so straightforward or obvious and is in fact perhaps misleading.

In the third section, I investigate various notions of a zombie in philosophy of mind, identifying Chalmers' particular notion. If one accepts his notion and then reasons further about it, one seems to end up with some bizarre, if not outright contradictory consequences, which, I argue, all point to a fundamental incoherence at the base of this notion of a zombie. In the last section I explore a number of possible functions for phenomenal consciousness, some of which are mentioned by Chalmers. In the later sections of his book, Chalmers wants to establish a coherence between experience and cognition (functional aspects of mind), however he argues for a natural link, not a logical one. I suggest that the qualitative aspect of our conscious life, phenomenal consciousness, could be essential to functioning in the way we do and in that sense may be to some extent at least functionally analysable and therefore more tightly connected to cognition than Chalmers' proposed "natural law-like" correlation allows.

To begin with, however, I wish to clarify what Chalmers means by phenomenal consciousness and how it is generally regarded in relation to mental functioning in philosophy of mind. Chalmers describes phenomenal consciousness as "the subjective quality of experience", saying, à la Nagel¹⁶, that "there is something it feels like to be a cognitive agent" (p. 4). As explained in chapter one of this thesis, Chalmers associates

¹⁶ Thomas Nagel (1974), "What is it like to be a bat?".

phenomenal consciousness with the “feel” of what it is like to be an agent. Other philosophers too describe phenomenal consciousness in such a way. For example, Block (1995), of whose paper Chalmers is aware, similarly relates phenomenal consciousness, or P-consciousness as he calls it for short, to experience: “P-conscious properties are experiential ones” (Block, p. 230). This is a fairly normal usage in philosophy of mind where the phenomenal is generally used to refer to subjective experience or the “what-it-is-like” aspect of consciousness. However, perhaps more controversial in philosophy are the activities with which phenomenal consciousness is to be associated. Most philosophers would probably agree that there is an experiential or phenomenal aspect to sensations and perceptions, but Chalmers explicitly goes further than this to also include thoughts, desires and emotions:

Some of the things we think and believe do not have any particular qualitative feel associated with them, but many do. This applies particularly to explicit, occurrent thoughts that one thinks to oneself, and to various thoughts that affect one’s stream of consciousness. It is often hard to pin down just what the qualitative feel of an occurrent thought is, but it is certainly there. There is *something* it is like to be having such thoughts (pp. 9-10).

Similarly Block says: “P-conscious properties include the experiential properties of sensations, feelings, and perceptions, but I would also include thoughts, desires, and emotions” (Block, p. 230).

Not all philosophers would agree that thoughts *per se* have an experiential quality and some would restrict such “qualia” to sensations, but I would certainly agree with both Chalmers and Block about this aspect of subjective experience. Janet Levin, in the Routledge Encyclopedia of Philosophy describes qualia as “most commonly used to characterize what may be called the qualitative, phenomenal or ‘felt’ properties of our mental states” and says “a mental state is considered to have qualitative properties just in case there is something it is like to be in it” (Levin, p. 863). This seems to be a similar description to Chalmers’ one, although, because the only examples of qualia mentioned

in the piece are sensations, it leaves it open as to whether thoughts, desires, etc. would have such a phenomenal feel. While it is usually more obvious that there is something it is like to experience a taste, touch, sound, etc., nevertheless, from one's own experience, it certainly does seem that there is also something it is like to be thinking, puzzling, remembering. In fact I would suggest that the whole time we are awake, or even dreaming, we are phenomenally conscious insofar as we are constantly subjectively experiencing the world and/or ourselves and our thoughts (although when we dream it is just our thoughts/dreams that we subjectively experience). Whether deeply engrossed in thought, say of a mathematical problem, or captivated by a particular sensation, say the warmth of a fire after being out in the cold, in both cases there is *something it is like* to be undergoing such an experience: thinking or feeling. A computer can solve a mathematical problem or sense temperature, but arguably we do not normally expect there to be something it is like to be the computer solving the problem or sensing the temperature (at least, not with current run-of-the-mill computers). So it does seem reasonable to associate phenomenal consciousness, as Chalmers does (and indeed many other philosophers do), with more than just sensory perceptions, but also with thoughts, emotions, (occurrent) beliefs, etc. – in the sense that there is something it is like to be thinking, believing, etc.

PHENOMENAL VS PSYCHOLOGICAL CONSCIOUSNESS

As explained in chapter one, Chalmers distinguishes two aspects of our mental life: the psychological and the phenomenal. Phenomenal consciousness as we saw in the previous paragraphs, is essentially qualitative and is distinguishable by its “feel”. Psychological consciousness on the other hand, Chalmers defines in terms of function and fulfilling a causal role. Although he does not want to beg questions about whether or not these two aspects will turn out to be the same or not (or at least inseparably

intertwined), as noted in chapter one Chalmers nevertheless does conclude that both these aspects are “real and distinct aspects of mind” (p. 16). Thus he makes a very definite distinction between the phenomenal and the psychological and goes on to rely heavily on this division in his arguments against materialism. The underlying assumption of his arguments against a materialist view of phenomenal consciousness is roughly that, because the phenomenal aspect of our mental life does not play any structural or functional role and indeed essentially does not seem to be of the kind that *could* play such a role, we have no means of analysing it in such terms. Therefore, it is not of the kind that can be reductively explained or thus explained in physical terms.

This is not to say that Chalmers denies *any* function of phenomenal consciousness (although he does seem to allocate all mental functioning to the psychological aspect of mind) but rather his claim is that the *essence* of phenomenal consciousness is qualitative and thus not explicable in terms of function. As he puts it:

Although [phenomenally] conscious states may play various causal roles, they are not *defined* by their causal roles. Rather, what makes them conscious is that they have a certain phenomenal feel, and this feel is not something that can be functionally defined away (p. 105).

However, Chalmers’ very division of consciousness along the lines of function, despite his claims, arguably already begs the question as to whether phenomenal consciousness could have a function or not. Of course one can identify a certain “feel” to one’s actions, thoughts, perceptions, etc. – what we call subjective experience – but it is questionable to what extent such a feel can be separated, even conceptually, from mental functioning. Arguably this phenomenal feel plays more of a functional and causal role than Chalmers seems to allow in his rather strict division of the phenomenal and psychological (functional) aspects of mind. Indeed I would suggest that this phenomenal feel may be necessary to our *kind* of functioning. It seems quite plausible that if we did not have subjective experience we would function rather differently than

we do. Despite his admission that phenomenal states “may play various causal roles” Chalmers nevertheless does not identify any of these potential roles and indeed seems to place *all* functionality firmly on the psychological side of mind. I would dispute such attempts and argue that the qualitative aspect of our mental life is also important to our functioning. Of course this does not necessarily prove that the *essence* of phenomenal consciousness is functional rather than qualitative, but rather it shows that quality could be important for function, in particular in ways that go beyond the mere law-like correlation that Chalmers speculates is the case in the final part of his book. In this sense, the essence of phenomenal consciousness may be to *some* extent explainable in terms of function contrary to Chalmers’ claims, though such analysis may still not completely explain it. I believe that divorcing the phenomenal from the functional as Chalmers appears to be doing, warps the issue because it seems to ignore the affect of any qualitative aspects of mind on its functioning and so takes us further away from understanding how our (integrated) mental life works.

As explained in chapter one, Chalmers describes the psychological aspect of mind as “characterized by an associated role in causation and/or explanation of behavior” (p. 12). However, most people tend to think of their subjective experience (the phenomenal aspect) as at least part of the explanation of their behaviour: for example, when one decides not to do something because one does not *feel* like it. We often say or do things because we are angry, happy, etc. - because of our subjective state or “feel”. That is, we explain our functioning in terms of our subjective experience. Of course many of our actions are based on reason too but, again, this is often not so easily separated from our subjective feel. We choose to do things we may not want to do, like go to work perhaps, because of longer term reasoning that we need the money to survive, however, coupled with this reasoning is more than likely a fear or desire not to be penniless – in other words, a phenomenal feel about what we do or do not want in our

lives. So it is by no means initially obvious that we can separate our subjective experience from the causal role and explanation of our behaviour.

Such subjective experiences often form strong *motivations* for actions and choices we make in our lives, so arguably they most certainly do play a functional role in our mental lives. Indeed often the qualitative feel appears to be *central* to our functioning. Thus it seems as if Chalmers is already making assumptions about the role (or lack thereof) of phenomenal consciousness in his very definitions of the two aspects of consciousness, phenomenal and psychological. Such definitions, if one accepts them, appear to already prejudice the question of whether subjective experience could play a central role in our functioning before any discussion even begins. Arguably quality might be more strongly tied to function than Chalmers seems to allow. So I will examine this division in more detail in the following paragraphs.

In making this distinction between the two aspects of mind, Chalmers briefly examines a number of notions often associated with the term “consciousness”, which he claims are psychological (functional) notions. It is interesting to look at some of these notions and Chalmers’ take on them to reach a clearer understanding of his proposed division between the psychological and phenomenal aspects of mind. One such notion is what he calls “awakeness”: that is, time when we are not asleep. Chalmers argues that because we have experiences when we are asleep (dreaming) then awakeness “clearly does not coincide with phenomenal consciousness” and “can plausibly be analyzed in functional terms – perhaps, at first approximation, in terms of an ability to process information about the world and deal with it in a rational fashion” (p. 26). While the latter analysis, in functional terms, seems applicable to the state of being awake, so also however does phenomenal consciousness. Just because we can analyse “awakeness” in functional terms does not at all show that the notion “clearly does not coincide with phenomenal consciousness”. In fact, arguably to most people subjective experience

would be most strongly associated with the state of being awake – we expect people to be subjectively experiencing the world around them, themselves and their own thoughts when awake. We do not expect this of them when they are asleep. (Dreaming is of course an exception to this for we know that people do have subjective experiences when they dream, but only of their own thoughts and not of the outside world to which they are more or less oblivious.) So in fact, I would suggest that in general phenomenal consciousness most certainly *does* coincide with the state of being awake (with the one exception of dreaming).

Indeed subjective experience would seem to be not just an important but rather a central aspect of being awake. In fact, when phenomenal consciousness is apparently not present (at least, if we have reason to think that the person is not phenomenally conscious of their surroundings because they act like an automaton) yet part of the functional aspect of awokeness is – for example, when someone is sleep-walking – then we are not inclined to say that that person is awake. The sleep-walker can to some extent “process information about the world and deal with it in a rational fashion”, because they can often walk and avoid obstacles (although their functioning usually is quite impaired). However they do not appear to be normally subjectively experiencing the world as they would when awake and thus we count them as still asleep. Indeed it is not clear what one could mean by the term “awake” *without* phenomenal consciousness – we do not refer to (current) computers or machines, things which obviously lack consciousness, as being awake. So I would query Chalmers’ claim that “awakeness” is primarily a psychological (i.e. functional) aspect of mind. Of course it does have a functional aspect to it but, as I have tried to show, it does seem to be inextricably and essentially phenomenal in nature too.

Another variety of psychological consciousness according to Chalmers is attention. He describes in functional terms paying attention to something as when “a

significant portion of ...cognitive resources is devoted to dealing with the relevant information” and further notes that we “can be phenomenally conscious of something without attending to it, as witnessed by the fringes of a visual field” (p. 27). The implication here is that attention is a largely psychological and according to Chalmers a “largely functional” notion (ibid.). Of course attention does involve devoting cognitive resources to the relevant information, as Chalmers proposes, but one also expects and assumes that the person paying attention is *experiencing* their situation. In fact, if someone were to claim that they had no experience whatsoever of that to which they were supposedly paying attention, we would most likely seriously doubt that they had been paying attention at all. It would certainly appear that attention, at least in its everyday usages, is inextricably tied up with phenomenal consciousness: a lot of the time our subjective experience is where our cognitive resources are directed. There are of course times, when we are daydreaming for example, when our subjective experience is of our thoughts and not of the task at hand (I will say more about this in the next section). But in such a case we would not say that we were paying attention to the task at hand either – at least it would be viewed as diminished attention.

Normally it seems that our attention is where our phenomenal consciousness is: when intently watching a movie, one’s attention is focussed on the movie, and likewise one’s subjective experience is predominantly of the movie. Similarly when not paying attention to anything in the immediate surroundings, but day-dreaming, one’s attention is focussed on one’s thoughts (the daydream) and one’s subjective experience is of those thoughts. Chalmers wants to separate the notions of attention and phenomenal consciousness, but as I have tried to show they seem inextricably bound, at least it is almost impossible to think of a situation where one could be paying attention to something and yet not be experiencing it. Chalmers’ example of the opposite – phenomenal consciousness without attention, as in the fringes of a visual field – is

actually not that simple. We are only vaguely phenomenally conscious of the fringes of our visual field. Normally our main subjective experience in this regard is of the centre area of the visual field, that is, where our attention is also (mainly) directed. So *both* attention and subjective experience of the fringes of the visual field are quite attenuated – unless something “catches our eye”, in which case we generally turn to focus on the distraction (i.e. adjusting our attention) putting it centre-field in our visual field and focussing our attention on it and thus subjectively experiencing it more fully. Of course sometimes when we are occupied with one thing, we nonetheless register information about something else in the fringes of our visual field in an unconscious manner and so in this way “pay attention” to it. This is perhaps an example of some degree of attention without much phenomenal consciousness. But it is a diminished attention compared to normal because the largest portion of cognitive resources (and phenomenal consciousness) is devoted to whatever it was we were occupied with and not what was in the fringes of the visual field.

Chalmers claims that the fact that these psychological features of consciousness, such as attention, awokeness, etc., are “largely functional notions” can be seen from “how one would explain the phenomena in question” (p. 27). In the case of attention he says that if “one were to try to explain attention, one might devise a model of the cognitive processes that lead to resources being concentrated on one aspect of available information rather than another” (ibid.). He claims further that such a functional explanation of what he calls varieties of psychological consciousness, such as attention, “seems to capture what is central” (ibid.). But such an account omits a very important part of *human* attention – the fact that we *experience* the aspect of available information upon which we are focussed. One of the most telling and important aspects of our attentiveness is our subjective experience of the attended phenomenon. If we focus our attention on a movie, say, it is our subjective experience of the movie that is central to

us, and consequently forms our verdict (great, boring, etc.) of the phenomenon attended to. A computer can be programmed to switch from reading and processing one set of inputs to reading and processing another set (say, dependent on certain environmental conditions), thereby “concentrating” its resources on one “aspect of available information” over another. (For example a computer in a chemical plant that controls solvent addition based on volume readings initially but later switching to temperature readings for control purposes.) Such a process is generally not enough for us to say that the computer is *paying attention* to volume as opposed to temperature, say. Arguably this is not just an issue of language – I believe the reason we do not talk of computers “paying attention” is because we expect more than just cognitive processes with “resources being concentrated on one aspect of available information rather than another” (although admittedly such a computer would only have a rudimentary claim on the concept of *cognitive* processing). We normally expect that the attending agent is in some sense subjectively experiencing the attended phenomenon.

In response to any argument that the phenomenal aspect is a necessary part of his “largely functional” notions of attention, awakesness, etc., Chalmers admits that there is “a phenomenal and a psychological property in the vicinity of each of these concepts” (p. 27). But he says that this issue is “largely verbal” and that one may speak of “pseudo-attention” to classify his functional notion of attention in that case. Nevertheless, the question remains: is pseudo-attention (pseudo-awakesness, pseudo-awareness, etc.) the *same* as attention (awakesness, awareness, etc.), even in a purely functional sense?

Chalmers’ division of mental events into phenomenal and psychological aspects however is by no means unusual in contemporary philosophy of mind. In fact, many philosophers would agree with a conceptual distinction between the two notions, even if we cannot actually separate them. Ned Block (1995) in a similar manner describes his

notion of phenomenal consciousness as distinct from any kind of functional aspect of consciousness, the latter which he calls access-consciousness. As Nagel pointed out in his article “What is it like to be a bat?”, the subjective experience (phenomenal feel) associated with being a particular sentient, conscious creature does not seem to be the kind of phenomenon that can be captured by science. Chalmers argues that the problem for a reductive explanation of subjective experience is that this phenomenal aspect of our mental life does not appear to be definable in terms of any (functional) role in our mental lives. But arguably this is partly because Chalmers has artificially divided the notions of psychological and phenomenological in such a way that the psychological aspects of mind do all the explanatory work. If phenomenal consciousness plays some role in our functioning in the way we do – that is, if pseudo-attention is not functionally exactly the same as attention – then the phenomenal may in some measure be functionally definable. However, I will discuss the problem of trying to functionally analyse consciousness in more detail later in this chapter. First I would like to look more closely at Chalmers’ account of another “largely functional” notion, awareness, and Block’s similar notion, access-consciousness. I wish to query Block’s claims as to the presence of phenomenal consciousness in certain situations, as I believe his (to my mind) misinterpretation of these casts doubt on his (and Chalmers’) way of separating the psychological from the phenomenal aspects of mind and consequent interpretations of functional role.

AWARENESS

Even though Chalmers’ claim is that the essence of phenomenal consciousness is not functionally definable, and is not as such the claim that phenomenal consciousness has no associated functions, he nevertheless allocates all functionality to the psychological (functional) aspect of mind, thus excluding any phenomenal aspects from

mental functioning. He then curiously suggests that phenomenal consciousness itself nevertheless could have a functional property associated with it. He calls this property “awareness” and places it too under the psychological aspect of mind. He says that it is “natural to suppose that there might be a psychological [functional] property associated with experience itself, or with phenomenal consciousness” (p. 28). This seems to go a step further than his earlier distinction between psychological and phenomenal aspects of mind. With awareness, he is then suggesting that any functional aspect at all that might have been claimed for this qualitative feel (phenomenal consciousness) should also be separated out from the phenomenal aspect of mind and placed under the psychological aspect of mind. Whatever about being able at least to conceptually distinguish between a phenomenal and psychological aspect of various mental activities (perception, thinking, etc.), to claim to be able to also distinguish a functional aspect of phenomenal consciousness and yet then claim that this is a separate notion, seems itself to be begging the very question of whether phenomenal consciousness could be functional.

He calls awareness the “most general brand of psychological consciousness” and defines it roughly as “a state wherein we have access to some information, and can use that information in the control of behavior” (ibid.). However it is hard to see how one could separate awareness from phenomenal consciousness. Indeed our normal usage of the term (interchangeably with “consciousness”) I believe indicates a largely phenomenal aspect to the term. We generally expect that a person who is aware of their surroundings, say, is subjectively experiencing them. How could one be aware of something *without* subjectively experiencing it (except perhaps in a functionally impaired way, such as with blindsight or awareness of the fringes of the visual field)? It would seem that awareness entails the phenomenal and without this would not be awareness but would simply be the ability to read in “inputs” (access to information) and

determine “outputs” accordingly (control of “behaviour”). Computers can be said to have states wherein they have “access to some information” (data) and “can use that information in the control of behavior” (outputs), yet we would never say that our PC was “aware” of information coming in through the keyboard. Awareness seems to be more than this and that “more” seems to be phenomenal consciousness.

Of course there can be some confusion here since there are *two* senses in which one might be aware, as Chalmers notes. He describes the first sense of awareness as a kind of awareness *without* a corresponding experience and notes that it is “most pronounced with propositional awareness – I am aware *that* my bicycle is downstairs” (p. 222). Presumably by this he means that one is aware of the fact in the sense that one *knows* the particular fact but is not calling it to mind at the moment and therefore has no phenomenal experience of it, *at the moment*. This would be awareness of something in the sense of having knowledge about it. One can know many facts (e.g. that the Eiffel tower is in Paris, etc.) and in *this* sense be aware of them, but one is generally not phenomenally conscious of them unless, say, discussing or using them.

Chalmers *excludes* this type of awareness from his own functional notion of awareness (the second kind of awareness) because later in his own speculative theory of consciousness he wants to show that his notion of awareness always *correlates* with experience. Thus he tries to more accurately define his notion of a functional type of awareness by claiming that this second kind of awareness has “a kind of *direct* access that cases of awareness without consciousness lack” and associates it with “occurrent thoughts” (ibid.). This is indeed probably the more common sense of the term “awareness” and is the kind of awareness I referred to in the previous paragraph: a focusing of one’s attention on something, being conscious of something. Indeed, one can become “occurrently” aware of one of the many facts of knowledge from one’s “store” of knowledge, by thinking about it, becoming conscious of it. It is *this* sense of

awareness (i.e. also corresponding to Chalmers' sense of awareness) that seems to me to be inseparable from phenomenal consciousness, since to be aware of something in this way, to focus on it at this very moment, would seem to require consciously experiencing it.

So although Chalmers' notion of awareness is a psychological or functional one, he nevertheless ties it to consciousness by associating it with occurrent thoughts only. Again I would suggest that to be aware of something generally simply *is* to have an occurrent thought or experience of that thing. Chalmers however insists that awareness is "quite distinct from the concept of conscious experience" (p. 105). Obviously awareness also includes a functional aspect, as Chalmers has described. However, his notion of awareness as direct access to information which as a result is poised to make a difference in control of a system does not seem to be *enough* to be awareness in the sense in which the term is normally used. Any automatic system (automatic pilot, an oven, etc.) controls in this way – reading data directly from its environment and changing the system outputs (actions) accordingly – but we do not say the plane or oven, etc., is aware¹⁷. I would suggest that this is because such a notion does not capture an *essential* part of the nature of our concept of (sentient) awareness – that is, its phenomenal nature. So I do not find Chalmers' distinction between the concept of awareness and that of (phenomenal) consciousness completely convincing.

Yet many philosophers also propose similar, purely functional notions akin to Chalmers' notion of awareness. In his book, Chalmers refers to Block's (1995) paper and compares Block's notion of access-consciousness with his own notion of awareness. He notes a slight difference however: "access consciousness corresponds roughly to my initial definition of awareness, although my definition gives less of a role to rationality" (p. 228). Block, like Chalmers, contrasts his notion of "access-consciousness", or "A-

¹⁷However, later in his book Chalmers does argue for some kind of basic consciousness for a thermostat, but this is due to his more or less panpsychist conclusions.

consciousness”, with that of his notion of phenomenal consciousness (P-consciousness), the latter which, as mentioned earlier, is similar to Chalmers’ notion of the same. Block defines A-consciousness in terms of consciousness of states: “A perceptual state is access-conscious, roughly speaking, if its content – what is represented by the perceptual state – is processed via that information-processing function,...whereby it can be used to control reasoning and behavior” (Block, p. 229). He also describes it as follows: “A state is access-conscious (A-conscious) if, in virtue of one’s having the state, a representation of its content is (1)...poised for use as a premise in reasoning, (2) poised for rational control of action” (Block, p. 231)¹⁸. Also like Chalmers, Block’s very definitions of access-consciousness and phenomenal consciousness allocate any functional aspects of consciousness to the former rather than the latter. However, unlike Chalmers, Block admits that this is a controversial assumption on his part: “The controversial part is that I take P-consciousness properties to be distinct from any cognitive, intentional, or functional property” (Block, p. 230).

It is interesting to look at some examples Block gives of the presence of P-consciousness without A-consciousness as it may help to shed some light on the distinction between Chalmers’ very similar notions of phenomenal consciousness and awareness, and indeed on the whole project of separating the functional from the phenomenal with regard to consciousness. Chalmers actually disagrees with Block’s analysis of the various examples given because he (Chalmers) wants to show a natural lawful correlation between (his notion of) awareness and phenomenal consciousness. I also believe, but for different reasons, that Block’s examples show the opposite of what he concludes from them. One example he suggests is as follows:

suppose you are engaged in intense conversation when suddenly at noon you realize that right outside your window there is – and has been for some time – a deafening pneumatic drill digging up the street. You were aware of the noise all along, but only at noon are you *consciously* aware of it. That is, you were P-

¹⁸ He does mention a third condition, effectively “reportability”, but notes that it is not necessary to the concept of A-consciousness as he wishes to allow that non-linguistic animals have A-conscious states.

conscious of the noise all along, but at noon you are both P-conscious *and* A-conscious of it (Block, p. 234).

Blocks' conclusion here seems strange because, if anything, I would suggest that one could claim the contrary. It seems more plausible that you were (to some extent) A-conscious of the noise all along, as you may have automatically strained to hear your companion over the noise by moving your head closer to hear better, for example. Such an action would surely be the result of a *functional* aspect of consciousness, i.e. a consciousness that has information poised for use (the noise of the drill), leading to action (leaning closer to hear) – in other words, A-consciousness, according to Block's definitions. Arguably, only at noon did you become P-conscious of the noise as well in that you began to *subjectively experience* (or at least, experience more clearly) the noise.

In reality, what I think is actually the case in this example is that *both* A- and P-consciousness were to a large extent suppressed with respect to the noise because the attention was on the conversation. Thus, before noon, one's subjective experience or P-consciousness would have been *primarily* related to the (intense) conversation and any phenomenal experience of the noise outside would have been quite minimal, but plausibly still present. In other words, before noon, you would have a much more detailed and vivid experience of the conversation than of hearing the noise. Similarly with A-consciousness: before noon it would be for the most part related to the conversation, for example, having the information from the conversation poised for use in reasoning (answering your companion's question) or action (your responses). Again there could be some minimal A-consciousness of the noise outside, which could show itself by the fact that you automatically (without explicitly deciding to do so) move your head closer to your companion in order to hear better over the noise or speak louder. Such actions, at least according to Block's own definition of the terms, would seem to be typically motivated by A-conscious states, rather than P-conscious ones. Therefore I do

not think Block succeeds with this example in showing the existence of either A- or P-consciousness without the other. Indeed Chalmers also argues this point, although from a slightly different angle than I do. He claims that it is “plausible that relevant information about the drill was *available* the whole time; it simply was not *accessed*” (p. 228).

One can also use a very similar but more common illustration of such automatic behaviour that apparently does not involve P-consciousness. For example, often while doing one (routine) task you can be “wrapped up in your thoughts” and completely unaware of the task at hand. Typically when walking or perhaps driving a familiar route you can arrive at your destination and realise that because you have been thinking so hard about something else, you have no memory of having made the journey (or at least part of it). So it is not that you are not P-conscious at all, but are rather that you are not P-conscious of a particular activity. Again, in such a case it seems that A-consciousness is working with respect to the routine task, albeit in a diminished way, otherwise you would trip, get run over if walking or maybe crash if driving. Information taken in, arguably P-unconsciously, during the walk/drive is still “poised for use as a premise in reasoning” (albeit routine reasoning, e.g. red light = stop) and also “poised for rational control of action” – all activities the very definition of which constitute A-consciousness, according to Block. Because the tasks involved (stopping at red lights, avoiding obstacles) are so routine and the route so familiar one does not need to think about them at all, but A-consciousness is still active and involved in executing the task. Block does allow that A-consciousness would be involved when driving in this “unconscious” way, as he discusses Searle’s example of driving “on automatic pilot”. He says the “sense of ‘conscious’ in which the car would crash if the driver weren’t conscious is *A-consciousness*, not P-consciousness” (Block, p. 241).

Block argues nonetheless that in such cases of inattention to the execution of

routine tasks it is primarily A-consciousness that is absent rather than P-consciousness: “Inattentiveness just *is* lack of A-consciousness (though it will have effects on P-consciousness)” (Block, p. 241). He claims that when the inattentive driver stops at a red light “there is presumably something it is like for him to see the red light” and that because he is thinking of something else, “he may not be using this information very much in his reasoning nor is he using it to control his speech or action in any sophisticated way – that is, perhaps his A-consciousness of what he sees is diminished” (Block, p. 241). But again it would seem to be the opposite that occurs: the driver *is* using the information in his reasoning (albeit for a routine task) as he *does* stop when he sees the red light which is an example of A-consciousness at work. However he has no (or little) subjective experience of seeing the red light (and making a decision to stop) because his subjective experience at that moment was of his thoughts – whatever it was in which he was “wrapped up”. This illustrates again, as I mentioned in the previous section, that it would certainly seem reasonable to tie attention to P-consciousness, as one tends to have to attend to something, to become aware of it, in order to *have* a subjective experience of it. One can still have some vague subjective experience of peripheral events/objects, but one’s *strongest* subjective experience is of that to which one is actively paying attention. What is going on in the periphery is noticed less (less attention) and therefore one has diminished experience of it, that is diminished P-consciousness of peripheral affairs.

This is not to say that A-consciousness is unrelated to attention – in order to carry out a cognitive task properly one generally has to take in information and use it in control of behaviour (A-consciousness activities), particularly if the task is new or difficult. Thus A-consciousness would presumably decrease with lack of attention, as Block indicates. However, these examples seem to show that we are able to carry out routine tasks without (P-)consciously thinking about them, while we are, for example,

thinking about (and thus experiencing) something else entirely. Performing any task surely requires A-consciousness since, according to Block's definition, A-consciousness is the cognitive, functional aspect of consciousness, whereby information is poised for use in reasoning and rational control. Thus when we perform such (routine) tasks with our thoughts elsewhere, contrary to what Block claims we must nevertheless still be to some extent A-conscious in performing the task since otherwise we could do nothing. As he says himself about his example of the epileptic who can continue to drive home during a seizure: "P-consciousness *all by itself* wouldn't keep the car from crashing – the P-conscious contents have to be put to use in rationally controlling the car, *which is an aspect of A-consciousness*" (Block, p. 241). By Block's very definitions of A- and P-consciousness, all functionality is allocated to A-consciousness and therefore *some* degree of A-consciousness *must always* be present when one is performing any task or action that requires cognitive, rational abilities and/or rational control. P-consciousness on the other hand, by his own definition, would not appear to be necessary to carrying out such tasks as he claims that it has no functional role.

I think that Block's examples just discussed show the difficulty of trying to sift out the functional from the phenomenal in relation to consciousness – this is by no means as obvious or straightforward as sometimes portrayed. So it is debatable whether Chalmers' purely functional sense of awareness is enough to cover our "everyday" sense of the term normally applied to humans (or sentient creatures), where subjective experience seems to be pivotal to the notion. The fact that it is so easy to misinterpret the role of phenomenal consciousness in our mental life, as I have tried to show Block (1995) has done in his paper (and indeed Chalmers argues this way as well), I suggest makes it clear that this role is not as easily dismissed as some might suppose. Through examining Chalmers' and Block's interpretations of phenomenal consciousness in this section, I have tried to indicate that phenomenal consciousness and its role in our mental

life is not so straightforwardly ascertained and hence perhaps not so easily isolated from functional aspects of mind (whether they are later associated back together or not), as Chalmers has proposed.

The “awareness” (in Chalmers’ functional sense) of even a quite sophisticated robot does not (currently) seem to be the same as the awareness of a sentient creature. This is because it seems that one can expect more in functional terms from a system that is aware in the phenomenal sense. However, it can be argued that this is because such a robot would not have the same degree of sophistication in functionality as a sentient creature, but my point is that it is quite possible that that very sentience (phenomenal consciousness) is needed for such a degree of sophistication in functioning (i.e. in the way we do). Of course, this may turn out *not* be the case, but I believe, at this point in time, such a conclusion is by no means obvious, as Chalmers appears to assume. While he may not *deny* phenomenal consciousness a functional role, he certainly makes no attempt to give it one, on the contrary allocating all functionality to the psychological aspect of mind. But it is hard to know how we could function as we do without the experiential aspect of mental life – arguably this aspect provides a functionality that (currently) does not appear to be duplicable in any other way.

The fact that both Chalmers and Block start off with a similar interpretation of phenomenal consciousness as a non-functional aspect of mental life, and yet end up with conflicting assertions about whether phenomenal consciousness is present or not in various scenarios, shows the difficulty of placing phenomenal consciousness and its role in our mental life. Arguably the same kind of difficulties plague the exercise of trying to discern whether zombies, as Chalmers describes them, are genuinely conceivable. I will delve further into the problem of the role of phenomenal consciousness in the following section by examining in more detail the notion of a zombie as proposed by Chalmers, and the coherence of this notion.

ZOMBIES

In his article “Varieties of Zombiehood”, Güven Güzeldere (1995) proposes three types of philosophical zombie: “I will distinguish three kinds of zombies on the basis of three different aspects in which they are equivalent to, or indistinguishable from, human beings: behaviour, functional/computational specification and physical constituency” (Güzeldere, p. 327-8). He describes a behavioural zombie as behaviourally indistinguishable from humans but notes that such a zombie could go “through the bodily movements that we take to be sophisticated human behaviour by a miracle, an act of Deity, and hence those movements shouldn’t be construed as anything beyond ‘as-if-behaviour’” (Güzeldere, p. 328). He adds that the behavioural zombie “has no internal structure or mechanism that would support a functional description of its psychology, and it would also immediately fail the physical indistinguishability test at the anatomy table” (ibid.). The functional zombie, according to Güzeldere, is “not only indistinguishable from human beings in behaviour, but can also be attributed a psychology that holds true of it at the right level of functional characterization” (ibid.), however it is made up of “entirely different kinds of matter” than its human counterpart. The third type of zombie, the physiological zombie, not only shares behaviour and psychological functionality with humans, but is also “identical in its physiology down to the minutest component” (ibid.). Obviously from these descriptions, as Güzeldere points out, a functional zombie encompasses a behavioural zombie and a physiological zombie also encompasses both of these kinds of zombie.

Such zombies are philosophically interesting, he contends, because they each supposedly lack something that humans have. The behavioural zombie, for example, could be accused of lacking “crucial aspects of the psychology of a human being” (Güzeldere, p. 329), since such a zombie need have no mind at all and could be merely

cleverly programmed to perfectly mimic human behaviour. Both the functional and physiological zombies, according to Güzeldere, could be said to lack “qualia-laden mental states”. He proposes that a functionalist would not accept that a functional zombie would lack qualia, and that physicalists would need “minimally to claim that the ‘physiological-zombies’ lack nothing at all” (Güzeldere, p. 328). As far as the physicalist is concerned, if the zombie is an exact physical replica of a human then automatically it has the kind of subjective experiences a human has because these are entailed by the physical. I think it is clear from Chalmers’ descriptions of a zombie that it is physiological zombies with which he is concerned in his zombie argument against a materialist view of consciousness. He describes his “zombie twin” as a creature who is “molecule for molecule identical to me, and identical in all the low-level properties postulated by a complete physics, but he lacks conscious experience entirely” (p. 94).

So unlike the physicalist or functionalist who would maintain that such an exact physiological duplicate does (must) indeed have experience, Chalmers asserts (as indeed some other philosophers do) that such a zombie could nevertheless conceivably *lack* experience. Of course, as Robert Stalnaker (2002) points out, in order to conceive of a physiological zombie that lacks experience we have to *assume* some form of dualism – that is, we must assume that there are properties (such as consciousness) over and above physical properties which are not reproduced in a world that is physically exactly like ours (zombie world). As a consequence Stalnaker notes, “if this is the only sense in which zombies are conceivable, their conceivability will provide no argument against materialism, since we must assume that materialism is false to be justified in inferring that zombies are possible from the fact that they are conceivable” (Stalnaker 2002, p. 399). My own intuition would be that, contrary to Chalmers’ claims, a physiological zombie *would* have experience. However, presumably one has to assume materialism to make *this* claim! So one’s intuition on zombies seems to provide little argument for or

against materialism.

However, I think it may be interesting to suspend belief temporarily as it were and critically examine Chalmers' notion of a zombie as he presents it – i.e. a physiological duplicate yet lacking experience. Thus I will not assume either materialism or dualism with regard to consciousness (if this is possible!) and from this “agnostic” standpoint try to evaluate Chalmers' notion of a zombie in itself in order to see if it is a coherent notion. I suspect that even if we remain open to the dualist claim that such a physiological zombie could nevertheless lack experience (e.g. we do not yet know if such a zombie is possible or not), further reasoning about this type of zombie leads to some seemingly contradictory conclusions. And such conclusions I believe arguably cast doubt on the coherence of this dualist zombie notion itself, even accepting (or at least remaining agnostic about) the initial dualist assumption.

In describing zombies, Chalmers appears to arrive at some bizarre conclusions about his type of zombie when he claims that his “zombie twin does not have any conscious experience, but he *claims* that he does” (p. 174). Chalmers' zombie twin is of course, in addition to being a physical replica of himself, also a functional one. He claims the zombie “will certainly be identical to me *functionally*: he will be processing the same sort of information, reacting in a similar way to inputs” (p. 95). He also points out that his zombie twin will be alike in the psychological (functional) aspects of consciousness discussed in the previous sections: “he will be awake, able to report the contents of his internal states, able to focus attention in various places, and so on” (ibid.). But he makes clear that none of this is accompanied by phenomenal consciousness: “none of this functioning will be accompanied by any real conscious experience... There is nothing it is like to be a zombie” (ibid.). So, accepting Chalmers' notion of a zombie at face value, it seems strange then that such a zombie would nevertheless claim to *have* subjective experiences. Why would a zombie claim to have

experiences when it does not have them, and, furthermore, most likely could not understand the very concept of subjective experience or what it would be like to undergo them?

In fact this notion, that a (Chalmers') zombie would claim to have subjective experiences even when it by stipulation does not, is a consequence of the very definition of a zombie as behaviourally and functionally exactly like humans. Todd Moody (1994) coins verbs for zombies - believe^z, understand^z, etc. – such that the words are without their phenomenally-laden¹⁹ aspect and only retain the functional aspect of their meaning. He says:

when a zombie uses the word 'understand' we must understand that he or she is not making any reference to any sort of conscious experience. To distinguish zombie-English words of this ilk from their English counterparts, I shall use the superscript^z. Thus, we can say that zombies understand^z many of the same things that we understand (Moody, p. 197).

Such words can presumably safely be used with respect to this type of zombie as the words are meant to convey a purely functional notion of the normal human mental activity mentioned without any experiential part, since these zombies by definition do not have experiences. However, despite the creation of such a zombie language, presumably many such terms – i.e. those equivalent to human terms dealing with the phenomenal - must, by definition, be empty. Moody to some extent acknowledges this emptiness of many zombie-English words, pointing out, for example, that for zombies “the word ‘experience^z’ can have at most a behavioural meaning” (Moody, p. 198).

Simply creating a zombie language equivalent to a human one does not shed light on how such zombies could be equivalent to humans, despite definitions. Stipulations can harbour incoherencies. Indeed, the fact that the zombie terms (with superscript z) would have different (non-phenomenal) meanings to ours for Chalmers' type of zombie, already appears to introduce a *functional* difference between such

¹⁹ Maintaining an initially “agnostic” view as to the possibility of Chalmers' type of zombie, I think that we can accept Moody's z-language as just a way of referring to the same actions but without any of the normal assumptions of an experiential aspect to them.

creatures and humans. Even if both use a similar-sounding language, the terms have different meanings and therefore cannot function in exactly the same way. For example, when Chalmers' zombies talk about dreaming^z, they must mean something different to what we mean when we talk of dreaming, for their dreaming^z is by definition empty (assuming dreaming is primarily experiential).

Owen Flanagan and Thomas Polger (1995) in their article "Zombies and the Function of Consciousness", try to fill out the zombie-equivalent of dreaming by sketching how zombies could come to dream^z, suggesting that during their "sleep" zombies recharge and that such a recharging process could have various sorts of "irregularities" which involve "continued activation of the zombies' speech centres" (Flanagan & Polger, p. 316). Thus, they claim, the zombies "come to call these reports 'dreams'" but these, of course, are "dreams^z not dreams" (ibid.). Whatever about the possibility of zombies believing^z or seeing^z, however, dreaming^z would seem to be an incoherent notion. When we dream, there is arguably no awareness, in Chalmers' functional terms, or A-consciousness, in Block's terms. We are not sensing or accessing information about our environment when we dream, nor are we using that information (or any information) for control of behaviour (unless sleepwalking perhaps, but this is an exception). The essential and most obvious aspect of dreaming is its experiential aspect. So what is left over to constitute zombie dreaming^z if they, by definition, do not have subjective experience? Dreaming^z, then, must surely be empty. So it does not seem to make sense to talk of a zombie equivalent of dreaming, contrary to attempts to explain it.

It is no surprise then that Flanagan and Polger too propose that zombies would claim they were conscious, saying "by hypothesis they would call themselves 'conscious'" (Flanagan & Polger, p. 318). So Chalmers is certainly not alone in this claim about zombies. Güzeldere also agrees that zombies can have equivalent

experiential concepts, such as dreaming^z, not, it would appear, so much because he feels that these could realistically develop, but rather because he understands that this is part of the *definition* of a zombie: “Of course there can be [such zombie equivalents]. There must be. There must be, because we postulated so in defining zombiehood” (Güzeldere, p. 330). As he points out “‘zombie-earth’ is a *stipulated construct*” (ibid.). Daniel Dennett (1995) too thinks that such zombie “mentalist vocabulary” would exist, but again not because this seems plausible (on the contrary he argues that philosophers “ought to have dropped the zombie like a hot potato” (Dennett, p. 322)), but because it is how we have defined the very notion of a zombie: “It is *not at all* unlikely or implausible that mentalistic vocabulary could evolve among zombies. That must be conceded as part of the concession that zombies are ‘behavioural’ twins of conscious beings; if it is likely that we conscious folks would develop mentalistic vocabulary, then it must be exactly as likely that zombies do” (Dennett, p. 322). (Of course materialists will also think that a physiological zombie must be conscious.)

Chalmers’ (physiological) zombie is defined to be an exact replica of a human being in every way, *except* with respect to phenomenal consciousness. So functionally speaking, if humans claim to have subjective experiences, then, again by stipulation, Chalmers’ zombies must make similar claims because they are functionally identical. But it seems nonsensical that zombies who are *stipulated* to be lacking subjective experience²⁰, should not only claim that they have it but also actually believe, according to Chalmers, that they have it. As he says: “Where I judge that I am conscious, he [zombie] judges that he is conscious” (p. 192). I think that this claim of Chalmers comes from his notion of what a functional isomorph entails:

we need also stipulate that for two systems to share their functional organization, they must **be in corresponding states at the relevant times**; although my sleeping twin might count as sharing my organization in a broad sense, he will not count in the strict sense required below. When two systems share their

²⁰ Again, trying to remain agnostic about whether this is *possible* or not, and simply examine the proposed (zombie) notion.

functional organization in this strict sense, I will say that they are *functional isomorphs* (p. 248, my bold type).

Although he is not talking directly of zombies in the above quote but rather of functional isomorphs, this definition would presumably also apply to his notion of zombie because a physiological zombie encompasses a functional zombie, and both must be functionally identical with humans. But to claim that a functional isomorph must be in a corresponding state (as its human counterpart) at the *relevant times* seems to illicitly tie the functional isomorph *forever* to its counterpart. Arguably if Chalmers' zombie twin were to be created at this very moment – functionally (and physiologically) identical and, therefore, initially claiming that it is conscious, according to Chalmers – then from this moment onwards the zombie would change and develop *differently* to its human counterpart, because (even on Chalmers' assumption) it is a very different creature if it has no subjective experience. Subjective experience frames our whole being – so much so that it would seem to be very difficult, if not impossible, to fully imagine even another type of subjective experience to the one we have (as Nagel (1974) so persuasively points out), let alone to imagine having no experience at all. So it is implausible to suggest that a zombie would continue to function exactly like a human being from the moment it was created (assuming it was created fully intact at a particular point in time) when something so fundamental to human life, i.e. experience, is by definition missing. At the very least, if the zombie is the intelligent creature it is stipulated to be (i.e. as intelligent as a human), it would notice^z that talk of experience did not match its, dare I say it, experience^z. So even though Chalmers' newly (and suddenly) created zombie twin may initially claim to have phenomenal consciousness because it is stipulated that he be behaviourally and functionally identical, he will soon discover that he does not know what he is talking about and presumably change his claims accordingly. Trying to make sense of this incoherent assertion (that zombies

would nevertheless claim to have experiences they by stipulation do not have), I suggest, merely serves to highlight the fact that Chalmers' notion of a zombie is confused and arguably fundamentally flawed.

In fact as I mentioned, paradoxically the proposal that Chalmers' zombies would claim to have subjective experience (because they must be functionally exactly like humans in every respect) even though they by stipulation do not, arguably leads ironically to the result that such zombies will thus *not* be functionally exactly like us at all. When we claim that we have experiences, this is in standard cases a veridical claim that there is something truly going on within us (assuming we are not lying of course). So, functionally-speaking, this claim gives whoever is listening correct information about us and thus the wherewithal to better predict/judge our behaviour. When a zombie does this, however, the claim is by definition (of a zombie) false and therefore misleading because incorrect information is being given. So in this case there is something functionally *different* to the human case going on. Thus ironically the attempt to force the zombie as stipulated into being *exactly* (functionally and behaviourally) like human beings, leading to the proposal that zombies would systematically make false statements about themselves, appears to actually result in creating a *functional* difference between zombies and humans. Human claims about consciousness are (generally) veridical and therefore functionally useful in communication with others, whereas zombie claims are not and so *cannot play the same functional role* because of their different (non-veridical) status. Again I believe this consequence indicates a basic incoherence in the whole notion of a zombie as proposed by Chalmers (and indeed by other philosophers in philosophy of mind).

Many philosophers do not discuss how such zombies, as defined by Chalmers, could possibly come about – for example, if they could evolve – since all that is required for the thought experiment is that a zombie is conceivable (or logically possible). As

Güzeldere puts it, if zombies are defined to be exactly behaviourally the same as humans, then “evolutionary considerations are rendered orthogonal” (Güzeldere, p. 331). However I think that it is worth examining the notion of such zombies evolving to see if it is a reasonable concept, as I believe it sheds light on the coherence of the notion of such a zombie itself. Most people would not deny that a dog or a cat, and often even simpler entities, have some kind of subjective experience. So the notion that a being could evolve to the complexity of a human – with their much greater mental abilities – without any subjective experience at all, would appear completely incongruous and grossly at odds with our conception of nature as a whole. One wonders how a zombie could evolve so far *without* developing consciousness especially when such consciousness shows itself so much earlier on in the evolutionary chain. Nature does not usually create such huge anomalies in its systems.

In addition, the proposal that zombies would claim to have experiences when they do not, just adds insult to injury, so to speak, to nature. Not only would there have naturally developed a being so grossly out of step with everything else in nature (in that it could evolve so very far without subjective experience), but that being would consistently be out of step with itself, as it were. The zombie would, it is proposed, believe of itself that it had subjective experiences even though it does not have them. So, such a creature would be a huge incongruity in not one, but two ways. Gross anomalies like this are rare in nature and it seems unlikely that any such creature could ever evolve – much more likely is that the zombie would evolve with consciousness, as most other animals have.

Even if it did somehow evolve without subjective experience, then there is absolutely no reason to think that it would nevertheless develop talk of such experience or “mentalistic vocabulary”, as Dennett calls it. For one thing, as I pointed out, such a development would be a major incoherence among the (zombie) species itself: that its

individuals be systematically wrong about many, if not most, of the things they report about *themselves*. Also, why would zombies develop talk of experience when they have never experienced anything? If they are not (phenomenally) conscious then they would never come upon the concept by themselves. One of the main reasons humans conclude that animals (and other humans) are conscious is based on our own personal experience. Arguably zombies could not even conceive of something so alien to anything they have ever undergone and, therefore, also would not be able to grasp the concept of consciousness if humans tried to explain it to them. This conclusion is of course at odds with Chalmers' proposed definition of a zombie which entails that the zombie should both lack experience and yet believe itself to have experience. Certainly if humans came across such zombies they would probably fairly quickly be able to ascertain that the zombies did not have subjective experiences, simply by talking to them. Chalmers claims that his zombie twin "often comments on the pleasure he gets from certain sensory qualia, professing a particular love for deep greens and purples", even though the zombie has "no conscious experience at all" (p. 180). In reality, if such creatures had somehow managed to evolve, then, contrary to Chalmers' zombie stipulation, we have absolutely no reason to think that they would talk of experiences and therefore every reason to believe that we could tell from talking to them that they were zombies.

Chalmers however argues that evolutionary considerations cannot be used to make a case against the notion of a zombie since they cannot provide an explanation of consciousness. As he puts it, "the process of natural selection cannot distinguish between me and my zombie twin....[he] performs all the functions that I perform just as well as I do" (p. 120). But the discussion in the previous paragraphs is not meant to argue that evolution *had* to produce phenomenally conscious intelligent creatures, but rather *given* existing life as it is - that it has taken that direction - then zombies would be a complete anomaly and, from that standpoint, probably could not evolve without

consciousness. I am also not trying to prove categorically that zombies could not evolve. (The thought experiment never addresses the thorny question of the origins of the zombie, claiming mere logical possibility of the zombie. But, as I have already argued in the previous chapter, logical possibility may well require a fit with our conceptual schemes and therefore an explanation of the evolution of the zombie.) I wish to argue in a general way for the implausibility and incoherence of this notion of a zombie in philosophy of mind. Of course philosophers like Chalmers will come back to the stipulation that a zombie is behaviourally and functionally exactly like a human being and thus there can be no differences in these aspects, but my point is that this notion does not make any sense partly because we cannot explain how the zombie could possibly have got that way. How a zombie could come to exist in the first place is rarely discussed (or, if it is, is consequently dismissed as “orthogonal” to matters since the definition of a zombie sets the limits, regardless of how the zombie came to be) because there are no plausible explanations of how zombies, as defined, *could* come into existence. Arguably it is such incoherence at the basis of the notion of a zombie that casts doubt on the logical possibility of zombies. As I have argued in the previous chapter, logical possibility has to take account of our conceptual frameworks and, where these are sketchy, one cannot be assured that a proposed scenario is at all logically possible.

Moody (1995) claims that we “*can* readily conceive of intelligence, in all its informational complexity, apart from consciousness” and says, in relation to Dennett’s claim that this is a “bogus feat of imagination”, that it is “far from clear what makes it bogus” (Moody, p. 370). But, in the same way that it seems highly implausible that (phenomenally) unconscious zombies would nevertheless not only talk of experience, but also believe themselves to have experiences when they do not, it seems equally implausible that creatures *with* consciousness could conceptualise something so alien as

functioning exactly like they do, but without any subjective experience. If we know nothing else, if our world is always, from birth to death, seeped in subjective experience, how can we really conceptualise functioning as we do without that experience?

Philosophers who believe in Chalmers' kind of zombie, arguably talk about such zombies believing^z, seeing^z, dreaming^z, etc., because they cannot truly conceptualise a being functioning as humans do but without subjective experience. So they inadvertently re-introduce experience into their zombies by using the same language (mentalist vocabulary) to talk about zombies as humans, however with the *z* superscript rendering that language *supposedly* devoid of all phenomenal aspects. How does one conceive of a zombie in Chalmers' sense? Presumably we start with what we know, ourselves, and then try to "subtract" subjective experience, but the latter is so ingrained and so much a part of the fabric of our existence that it is highly questionable whether we can genuinely perform this feat of conception. In fact our tendency is much stronger in the other direction, that is, to anthropomorphise inanimate systems, such as robots in popular science fiction stories that show anything near the functionality of humans, like Data in Star Trek or R2D2 in Star Wars.

So it appears that even when we accept Chalmers' proposal of a zombie (a physical duplicate of a human yet lacking phenomenal consciousness) and reason further about it, the proposal seems to be full of tensions. We get beings stipulated to be functionally the same as us, yet that plausibly would be functionally different to us because any of their claims about experience, unlike ours, are false since by stipulation they have no experiences. Perhaps the assertion that beings who systematically make false claims about themselves (zombies) would end up *functionally* different to similar beings who do not such false claims (humans) rests upon a materialist assumption. However, even the notion that such zombies would systematically make false claims about themselves at all arguably shows the incongruity of the notion of this type of

zombie. It seems that the two stipulations required by Chalmers' definition of a zombie (that it be a physiological/functional duplicate of a human *and* that it have no phenomenal consciousness) are mutually exclusive. At any rate, as I have tried to show, further reasoning about the combination of these two stipulations certainly appears to lead to incoherencies. Also, the question of how such zombies could ever come about in the first place leads to problems for this type of zombie. The argument of course is that these are stipulated constructs and therefore evolutionary concerns are rendered "orthogonal". But as I have argued in the last chapter, anyone proposing the logical possibility of such a creature arguably does owe us at least an indication of how they could fit in with our current conceptual frameworks.

Another thought experiment, similar to Chalmers' zombie one, but with an account of how the zombie "evolved", is described by John Searle (1992) about silicon brains. He suggests imagining that, to save your brain from an inexplicable deterioration, doctors progressively replace more and more of your brain with silicon chips – which to "your amazement and theirs" seem to restore your normal functioning – until "in the end, we imagine that your brain is entirely replaced by silicon chips" (Searle, p. 66). According to Searle there are a number of possibilities in relation to this scenario: one logical possibility he says is that your mental life, experiences, etc., remain the same as before; the other is that you find your conscious experience progressively (as more and more of your brain becomes silicon) dwindling to nothing, even though your behaviour remains exactly the same. This latter possibility is of course another way of approaching the zombie experiment (perhaps considered easier to conceptualise and, also, dealing more plausibly with the origin of the zombie), with the similar end result that we have a being who is supposed to be functionally and behaviourally (although not physically, so it is not a physiological zombie) identical to a human but without conscious experience.

However, in this case it is not *stipulated* that the being have no phenomenal experience, that is merely claimed to be one of a number of logical possibilities with regard to the given scenario. Searle, however, definitely does not think that the first logical possibility (that your experience, etc., would remain the same as before) is even “remotely empirically possible” (Searle, p. 66). On the contrary, he thinks that it is “empirically absurd to suppose that we could duplicate the causal powers of neurons entirely in silicon” (ibid.). The only reason he seems to give for this belief is that he says that in this (first) case “we are imagining that the silicon chips have the power not only to duplicate your input-output functions, but also to duplicate the mental phenomena, conscious and otherwise, that are normally responsible for your input-output functions” (ibid.). But within this reasoning there is already the built-in assumption that duplicating the input-output functions of each neuron is not enough – Searle apparently assumes that “mental phenomena” are an “extra” that must also be duplicated. As with the zombie thought experiment, Searle seems to assume that all functionality and behaviour is duplicated by the silicon chips yet not the phenomenal, so that a similar bizarre result is reached: you will slowly lose phenomenal experience yet will not be able to communicate this to others and will continue to function as before, as if you still had phenomenal experience.

Selmer Bringsjord (1999) goes a step further and claims that such a scenario is *physically* possible. He discusses Searle’s “silicon brain” thought experiment above and, with regard to the second possibility of losing phenomenal experience yet remaining functionally and behaviourally the same, says that there would appear to be no reason why this “ought not to be regarded [as] *physically* possible” (Bringsjord, p. 62). He suggests that a “neuroscience-schooled Kafka [could] write us a detailed, compelling account..., replete with wonderfully fine-grained revelations about brain surgery and ‘neurochips’” (ibid.). He claims that this “thought-experiment is to be

devised to preserve consistency with neuroscience and neurosurgery specifically, and biology and physics generally” (Bringsjord, pp. 63). It seems implausible that Bringsjord should argue for the physical possibility of this scenario since such a claim would normally be considered a much stronger one than that of “mere” logical possibility. He seems to think that with his proposed detailed, scientific “Kafkan” account of this procedure, described so as to “avoid any inconsistency with neuroscience, neurosurgery, etc.”, that it “should be easy enough to convince someone” (Bringsjord, p. 63) that it is physically possible. But it is irrelevant how good the “Kafkan” story is, *convincing* someone that something is physically possible is not the same as *showing* that it really *is* physically possible. In fact, Bringsjord is ignoring the possibility that when this Kafka-like writer were to be able to flesh out the details of the proposed brain-replacement thought experiment, that there *could* turn out to be inconsistencies with our current knowledge of the laws of nature. But perhaps Bringsjord has an unusual notion of what “physically possible” means: he claims that Chalmers too “considers it *obvious* that zombies are both logically and physically possible – though he doesn’t think zombies are *naturally* possible” (Bringsjord, note³⁴, p. 63). He does not say from where in Chalmers’ (1996) book he deduces this sentiment. In fact, as far as I can discover, Chalmers only talks of natural possibility, which seems to mean physical possibility for him – certainly he defines it in a similar way, as I examined in the last chapter.

Bringsjord also tries to use Chalmers’ tactic to argue against those who would deny his claim of physical possibility for the silicon brain scenario. Chalmers argues, with regard to the logical possibility of zombies, that “a certain burden of proof lies on those who claim that a given description is logically *impossible*” (p. 96). Similarly Bringsjord argues, with regard to the silicon brain thought experiment, “the same principle would presumably hold with respect to *physical* possibility:...the burden of

proof is on those who would resist affirming [that a particular proposal is physically possible]...to indicate where physical laws are contravened” (Bringsjord, p. 63).

However, arguably our current knowledge does not tell us whether such a procedure could be carried out in principle or not. This is not just because such a task is currently technically impossible, but rather we do not know enough at this stage to say whether it is physically possible at all. Certainly the person who wishes to claim that such a feat is physically possible (as opposed to the weaker claim of logical possibility) would need to at least indicate how such a procedure could be consistent with the laws of nature.

Arguably the most Bringsjord can claim is that there does not *seem* to be any reason to think that any physical law must be contravened in the proposal that a person’s brain could be gradually replaced with silicon chips while maintaining functionality and behaviour. But he in fact claims more than this: he argues that when the person’s brain is replaced with silicon chips that they will also have lost phenomenal consciousness. Whatever about the physical possibility of replacing someone’s brain with silicon chips which still preserve all functionality, the notion that a person would thereby lose subjective experience is by no means convincingly physically possible. In fact, it seems more natural to suppose all mental life would remain more or less the same. Here again there is a dualist supposition built into the thought experiment, only this time it is taken one step further with the claim that it is physically (as opposed to just logically) possible. Basically, brain replacement by silicon, while conserving all functionality, may well be physically possible – certainly it seems plausible enough – however, the *further* assumption that this would result in loss of phenomenal consciousness (which could not even be communicated.) does not have the same ring of plausibility about it. At the very least, it is equally as plausible (more so, to some people) that phenomenal consciousness would remain intact. So the claim of physical possibility cannot yet be determined for *either* scenario and so carries no weight to prove it one way or another. Bringsjord’s

contentions, I feel, merely highlight how seriously misled we can be by persuasive possibility claims.

There is an inbuilt dualist assumption in the notion that a physical duplicate of a human could lack phenomenal consciousness, or that replacement of a person's brain by silicon while retaining exactly the same functionality could result loss of phenomenal consciousness. The assumption is that replicating the exact physical or functional structures and dynamic processes of a (phenomenally) conscious creature would not necessarily replicate the phenomenal consciousness. However, as I noted earlier, to argue that such duplication *would* necessarily replicate phenomenal consciousness presumably involves a materialist assumption: that such consciousness is nothing more than the underlying physical/functional structures and processes. So by the discussions in this section I have tried to demonstrate that even accepting (or remaining agnostic about) the dualist assumption in Chalmers' proposal for a zombie, the notion itself has fundamental difficulties, and that this in itself is a good reason for rejecting the possibility of such a zombie. The incongruous conclusion that such creatures would systematically hold false beliefs about themselves; the fact that such false beliefs would plausibly create a functional difference with humans when there is stipulated that there should be no such difference; and the lack of any remotely plausible explanation of how such zombies could evolve or develop "mentalistic" vocabulary, together I would suggest indicate a fundamental incoherence at the base of the zombie definition itself. Of course these are not hard proofs that the notion of such a zombie is incoherent but such basic tensions in the notion when we reason thus further about it, arguably are good grounds for questioning the logical possibility of the notion. To use Chalmers' (2002) terms, I am suggesting that his notion of a zombie is perhaps "prima facie conceivable" but that further rational reflection about the notion exposes fundamental tensions in it, rendering it, I would argue, not in his sense "ideally conceivable".

Arguably it is very difficult and perhaps even impossible, for us to genuinely conceive of beings exactly like ourselves but without any subjective experience because each person's world is so completely determined by experience. We constantly explain our behaviour and functioning in terms of our subjective experience, so arguably it seems more natural to conclude that experience does play a role in our functioning, and that, without experience, we would behave quite differently. This is what I will look at in the next section.

FUNCTIONS FOR PHENOMENAL CONSCIOUSNESS?

The basic premise of the zombie argument against a materialist view of consciousness is that phenomenal consciousness is not explicable in terms of a functional role in our mental life and therefore duplication of the physical basis and functionality of our mental life does not automatically mean duplication of our subjective experience. As Chalmers puts it, the way to a reductive explanation of a phenomenon is via analysis: "With such an analysis in hand, all we need to do is to show how certain lower-level physical mechanisms allow the analysis to be satisfied, and an explanation will result" (p. 44). However, as he also contends (and many other philosophers do), phenomenal consciousness is not functionally analysable because its *essence* is qualitative - "experiences range from vivid colour sensations to experiences of the faintest background aromas;...All these have a distinct experienced quality" (p. 4). Hence, the reasoning goes, a functional or even structural analysis does not capture this essence.

Chalmers argues for this view by first conceptually separating out the functional aspects of mental life from the experiential (as discussed earlier in this chapter) and then asking with regard to any given cognition, "why should this be accompanied by an experience?" (p. 5). Further he asks "why is seeing red like *this*, rather than like *that*?"

(ibid.). But one could nonetheless query this division. If the essential characteristic of experience is its quality, this does not exclude an essentially functional aspect on the part of experience either: that is, a function of this qualitative feel. The very fact that the essence of experience is qualitative may indeed allow it a particular kind of functioning that could not have been achieved in any other way. Qualitative variances are arguably informational (they are differentiable) and, in this sense, this qualitative essence of experience could supply the experiencing organism with information, affecting its behaviour and functioning as a result. If this qualitative essence of phenomenal consciousness does play a functional role in our mental life then a physical duplicate (of a human) lacking experience could not be a functional duplicate as well. Conversely, a functional duplicate of a human could not lack experience. So even though Chalmers' point is that the *essence* of phenomenal consciousness is not definable in functional terms, it seems that he must actually deny phenomenal consciousness a functional role at all in order to support his zombie argument.

This is exactly what Chalmers effectively does. Despite allowing that phenomenal consciousness *may* play a role in mental functioning, through his conceptual distinction between the phenomenal and psychological aspects of mental life, his analysis places *all* functionality with the latter aspect. Later in his book, however, Chalmers does allow that the phenomenal may be *lawfully correlated* with the psychological (functional) when he proposes his own theory of consciousness, which he terms "naturalistic dualism". He argues in the first half of his book against any logical or *a priori* entailment (logical supervenience with respect to primary intensions) of the phenomenal by the physical, but later suggests that there is nevertheless a *natural* nomological connection between the two. He notes that there "appears to be a systematic dependence of conscious experience on physical structure in the cases with which we are familiar" and claims that this is "natural supervenience without logical

supervenience” (p. 124).

He describes his naturalistic dualism as a sort of property dualism insofar as “conscious experience involves properties of an individual that are not entailed by the physical properties of that individual, although they may depend lawfully on those properties” (p. 125). Such properties of conscious experience are “ontologically independent of physical properties” so that “consciousness arises from a physical substrate in virtue of certain *contingent* laws of nature, which are not themselves implied by physical laws” (ibid., my italics). In other words he is proposing that our “ontology must be expanded by new fundamental properties (phenomenal or proto-phenomenal properties) and new fundamental psychophysical laws” (Chalmers 1999b, p. 436). Thus in the latter half of his book, Chalmers argues for principles of *coherence* between the phenomenal and the physical, proposing that “consciousness and awareness covary, as do their structures” (Chalmers 1999b, p. 437). In this manner he believes he has created a theory which takes phenomenal consciousness seriously, properly accounting for experience in a way that he feels materialist theories do not because, he has argued, any functional/structural descriptions of conscious experience do not capture its essence. He also feels that his theory is a naturalistic one, as he proposes that (proto)phenomenal properties will just be part of nature in an expanded ontology.

On the presumption that materialism is false, Chalmers’ naturalistic dualism is arguably a plausible theory of consciousness. Although there may be concerns about the apparently epiphenomenalist²¹ and panpsychist²² consequences of his theory, Chalmers argues that such consequences are counterintuitive but not fatal. However it is questionable whether we need to go as far as such a naturalistic dualism in order to explain consciousness. Of course Chalmers believes that he has proven materialism to

²¹ Chalmers says that he does not describe his view as epiphenomenalism because he claims that the “question of the causal relevance of experience remains open”. However, he does admit that his view “implies at least a weak form of epiphenomenalism, and it may end up leading to a stronger sort” (p. 160).

²² Similarly with panpsychism, Chalmers says that while his view may not necessarily result in panpsychism, it is “probably fair to say that the view is a variety of panpsychism” (p. 299).

be false, but I have argued in chapter two of this thesis that his zombie argument against materialism is unsound insofar as the claim of logical possibility he makes and *a priori* entailment cannot do the work he requires of them. (I have concentrated on Chalmers' zombie argument against materialism but of course he does provide a number of other arguments, claiming that all these arguments together defeat materialism. However the foundation for his version of such arguments is the same as for his zombie argument and lies in his proposed two-dimensional intensional framework and appeal to the *a priori* and the resulting logical possibility to do the relevant work – all of which I have argued against in chapter two.)

I believe Chalmers' strict opening partition of the functional aspects of mind from the phenomenal leads to an artificial polarisation of the issue and consequently brings him down the road of dualism of one sort or another. His argument that the essence of phenomenal consciousness is not captured by any functional or structural description is the basis of all his (and indeed most) contentions against a materialist view of experience. This is indeed a powerful argument, but separating out all functional aspects of mind from the phenomenal merely serves to reinforce the original intuition, instead of looking at what functions (if any) may possibly belong to phenomenal consciousness. This is not to say that if a function could be found for experience that it would fully explain it, but one should at least not prejudice the issue from the start, as I believe Chalmers does by his analysis of our mental life which leaves the phenomenal aspect completely devoid of any kind of function.

Interestingly, in the later part of his book when he is trying to argue that phenomenal consciousness is *naturally* (though not logically) always tied to psychological functioning, Chalmers appears to indirectly allow for some functions for phenomenal consciousness. Of course he would allocate any *function* to the psychological rather than the phenomenal aspect of mind, as his theory requires that the

psychological aspect of mind do the functional work. (He allows that the phenomenal may play a role in causation but apparently a kind of superfluous one, as he contends that it “remains the case, for example, that one can imagine removing the phenomenal properties, with the pattern of causal flux remaining the same” (p. 154).) Nevertheless in trying to show a correlation between phenomenal consciousness and his functional notion of awareness, he is of course showing a correlation between the phenomenal and function.

For example, with regard to our belief that animals also have subjective experience, he suggests that we “are generally prepared to attribute perceptual experience of a stimulus to mammals in cases where the direction of behaviour can be made to depend on that stimulus” (p. 226). If we commonly attribute subjective experience to individuals on the basis of behaviour (in response to stimuli) then obviously we expect such experience to have an effect on one’s *behaviour*. I agree that this is something we regularly do, basing judgements about other people and animals on our own subjective experience and resulting behaviour, and thus assuming that they experience in a similar way. So this would seem to indicate that phenomenal consciousness does affect behaviour (at least, we expect it to) and in this way does have a function. Although Chalmers is trying to show that phenomenal experience is merely lawfully *correlated* with functioning, it seems a more straightforward explanation to simply say that experience is playing the functional role of affecting behaviour in particular ways. Even if such a functional role for phenomenal consciousness would not adequately capture its qualitative essence, as Chalmers has argued, it is a start. Arguably taking this route in the first place and not polarising the issue by separating the functional from the phenomenal, would not leave one in the position of having to embrace dualism.

Again, in this later section of his book where Chalmers is trying to show a

correlation between consciousness and cognition, he seems to once more advocate a function of phenomenal experience. He proposes that “Siewert (1994) makes a compelling case that experience by its nature is informative about the state of the world: a visual experience, for example, is something that is assessable for accuracy (it can represent the world correctly and incorrectly), and indeed is assessable in virtue of its very nature as a visual experience” (note 10, p. 233). Chalmers himself is inclined to think that “experiences and associated registrations could both have content autonomously, without there being a strange, coincidental overdetermination whereby the same content is constituted twice over” (ibid.). However, this indicates support for the notion, from a materialist viewpoint, that there is a function for phenomenal experience – i.e. representing information to a system – in particular the proposal that experiential content would not be overdetermined, and thus rendered functionless. Chalmers gives further weight to this interpretation by later discussing “phenomenally realized information”, claiming that: “Physical realization is the most common way to think about information embedded in the world, but it is not the only way information can be found. We can also find information realized in our *phenomenology*” (pp. 283-4). Here again, Chalmers is presumably not arguing that phenomenal experience has a function (supplying the organism with information), however it is certainly open to the materialist to make this conclusion. Why contend that experience is merely lawfully correlated with a particular kind of function (information supply) as Chalmers suggests, instead of just saying that supply of information is a functional role of phenomenal experience? Returning to the notion of a zombie, if phenomenal experience does realise information then zombies cannot be functionally (or probably even behaviourally) exactly like humans because they will of course lack such (experiential) information and this will affect their functionality and behaviour.

As I have mentioned Chalmers is arguing here for a merely natural, observed

correlation between phenomenal experience and any kind of functionality, which leads him to his speculative ontology. However in his earlier analysis of phenomenal consciousness, which seemed to leave all functioning with the psychological aspect of mind, Chalmers plays our intuitions one way; and yet later, to show a connection between the phenomenal and the functional, albeit a mere correlation, he is now playing our intuitions the opposite way. At any rate, as I have tried to show in the last chapter, logical connections and entailment between natural phenomena and their underlying basis are generally not developed purely *a priori*, before we know anything about a given phenomenon, but rather start to become apparent through the process of discovering the underlying nature of the phenomenon in question. Therefore the observation of possible functions of experience, I would argue, is the beginnings of the exploration of possible entailment of the experiential by the physical.

Undoubtedly it seems hard to describe, let alone capture, the qualitative essence of subjective experience in terms of function, or even structure or material. Quality in general is often seen as ineffable: it is difficult even to pinpoint the qualitative essence of, say, a painting or a beautiful scene. However, there is a perhaps more commonplace, “day-to-day” sense of quality whereby the quality of an object, for example, often more or less amounts to a combination of factors such as the strength of the materials from which the object is made, the adequacy of its composition and so on. Naturally this may be a far cry from the type of “quality” involved in phenomenal experience, but it is arguably related. Quality is a highly complex phenomenon and it is presumably complicated by the fact that it is almost inevitably tied to consciousness, with sentient beings usually the perceivers of and judges of quality. But such qualitative aspects may be more accessible to description in terms of function/structure than Chalmers and some other philosophers appear to allow.

It certainly seems difficult to imagine any being as sophisticated as a human,

functioning in the same way without experience. For example, a neural network system for colour recognition might be able to distinguish between, say, red and blue, by the different wavelengths of light reflected off a given object. This is obviously (since difference in wavelength is the basis for difference in colour) also the way the human visual system differentiates between colours. But this is not how we *know* the difference: we recognise blue or red because of the way each colour looks to us – subjective experience is our human organism’s (and presumably some animals’) means of distinguishing one colour from another. A zombie’s visual system would of course also differentiate colours by their wavelength, but if there is nothing it is like for the zombie to see red or blue, etc., then how does the organism itself (the zombie) *know*, at least in the same way we immediately and directly recognise through our subjective experience, that something is red or blue?

Chalmers acknowledges this to some extent, saying that “the semantic *content* of my phenomenal beliefs is partly constituted in subtle ways by conscious experience itself (for example, red sensations may play a role in constituting the content of certain beliefs about red sensations)” (p. 174). Thus he allows that if this is so, then “some of the zombie’s judgements will have contents that are not as rich as my corresponding belief contents” but maintains nevertheless that “they will...*function* in the same way in directing behaviour as mine” (ibid.). Thus he sees a difference in the presence or absence of subjective experience as merely a difference in “richness” of belief/judgement content and not at all as having any functional effect. However, even by his own examples just mentioned in previous paragraphs, it seems that a difference of richness between the zombies’ belief content and human belief content is also a difference in information between the two contents. That is, the zombies’ less “rich” belief content is arguably missing some information – otherwise, what makes it less “rich”? It seems natural to suppose that the increased “richness” of our belief contents

gives more information. For example, someone who has blindsight²³ is plausibly processing some information about objects in their “blind” field (since they can make much better than average “guesses” at what is there) but they do not *know* what they are seeing there, they are not phenomenally conscious of it, so *for them* it is only a guess and not knowledge²⁴. If the contents of zombies’ beliefs are not as “rich” as ours, then presumably there is *less* in their belief contents, which are thus different to ours, and accordingly it seems reasonable to expect that there would be differences in function and/or behaviour between such zombies and humans.

The way Chalmers has set up his distinction between what he terms as the (functional) psychological and phenomenal aspects of mind, is such that he claims that according to “the psychological concept, it matters little whether a mental state has a conscious quality or not” (p. 11). However, this seems to go completely against the grain when one thinks of human behaviour. Certainly a common-sense notion would be that our behaviour is very much dictated by the subjective “feel” of experience and, indeed, we generally interpret a person’s behaviour based on what we estimate their current subjective experience to be. Chalmers even tries to separate out a functional role of happiness: “whenever one has the conscious experience of happiness, the functional role associated with happiness is generally being played by some internal state” (p. 22). He contends that it is merely an empirical fact that the experience and the causation go together, musing that it is perhaps “logically possible that one could have the experience without the causation” (ibid.). However, happiness *is* the experience – we *feel* happy. If we were zombies, thus without any subjective experience, then we would not feel

²³ Chalmers describes blindsight as a “deficit arising from damage to the visual cortex” (p. 226) whereby subjects believe that they can see nothing in certain areas of their visual field. However when forced to “guess” what is in this blind part of their visual field they are correct much more often than not. So it appears that information from the blind area is taken in by the patients but they do not subjectively experience this information and therefore do not realise they have it.

²⁴ Chalmers claims in this regard that “it is not *obvious* that there is no experience in these cases” (p. 227). However, if the patients insist that they do not experience anything in their blind field then this gives us very good reason to believe that there is no experience in these cases (or at least negligible experience) – we currently have no other way of discovering otherwise.

anything in this sense and there would be no happiness. Emotions are inextricably tied to subjective experience and completely colour our lives, behaviour and, therefore, the way we function as human beings. Any particular emotion is in essence very much the subjective feeling or experience it produces in us and this *feeling* in return results in various actions. In fact it does not even make sense to speak of the functional aspect of the emotion (presumably the resulting behaviour and mental states) without speaking of the emotion, that is the feel or subjective experience.

So it is hard to see how zombies could be said to have the resulting behaviour of a given emotion without the subjective experience – the essence of the emotion itself. It just does not make sense to talk of beings without subjective experience being happy or sad, etc. But broader than just emotions, all our subjective experience colours and deeply influences our lives. If there were no nasty subjective feel to touching something that burns your hand, then why would you immediately withdraw your hand? When asked to explain your action you would probably say that you withdrew your hand because it hurt. If chocolate did not taste nice, why would we bother to eat it? Of course, we have to eat to stay alive but, if taste was not a reason for our choice of food, we could choose purely according to the nutritional value of a given food and our own nutritional requirements, rather than (often to our own detriment) according to its taste. We very often give experiential feeling as our reason for action so it would seem that subjective experience is a strong motivator for those organisms that have it.

Of course organisms we do not suspect of having experience, such as plants or bacteria, react in a reflexive manner: flowers turn their head to follow the sun, single-celled organisms may move toward or away from certain stimuli such as sunlight. This is presumably carried out purely by means of chemical reactions or some such mechanism without the organism experiencing anything. From an evolutionary viewpoint one could wonder why experience developed at all, could not every reaction

be just reflexive? Possibly once organisms became more sophisticated and, in particular, able to move freely around, the choices of action to take became increasingly complex and perhaps pure reflex reaction would not have been enough to enable the organism to survive in such an environment. Presumably an animal with the ability to move needs a strong warning and motivational system to stay away from dangerous situations for example, and pain certainly provides this. The unpleasantness of the experience of pain seems to be a very direct way for the organism to immediately know its (dangerous) situation and react, and also to be motivated to avoid such unpleasant pain-producing situations in the future.

Arguably, subjective experience provides a very strong learning tool for a sophisticated organism in a changing environment. Chalmers' functional account of learning (roughly "appropriate changes in behavioral capacity in response to various kinds of environmental stimulation" (p. 46)) leaves out this most important aspect: motivation. How one learns depends hugely on one's interest and motivation. So experiences like pain and pleasure provide the required interest or motivation in learning that humans and many animals have. One learns very quickly to stay away from hot stoves if one gets burnt. A robot could also be programmed to retract its limb if the surface it touches is over a particular temperature, but where would be the motivational factor for it to stay away from such surfaces (and perhaps consequently learn to recognise them speedily) if it did not feel any pain touching something too hot?

In fact, if a person loses their sense of touch it can be a potentially dangerous occurrence for them as they no longer have the means to feel if there is damage being done to (that part of) their body. Antonio Damasio (1996) gives an example of such a condition, whereby individuals are born with a "congenital absence of pain" (Damasio, p. 264). Such people he says "do not acquire normal behavior strategies" (ibid.) and, for example, do not withdraw their hands from a hot plate or sharp blade even though it may

be damaging them. Thus many of them “seem to be eternally giggly and pleased, in spite of the fact that their condition leads to damage in their joints..., severe burns, cuts” (ibid.). So it seems that without being able to subjectively experience pain, such people do not know when any part of their body is in pain, and therefore in danger, and so remove the part from danger. This would seem to indicate a role for phenomenal experience – an organism’s means of directly knowing what is going on with itself in relation to its environment. Damasio suggests that pain could be “a lever for the proper deployment of drives and instincts, and for the development of related decision-making strategies” (ibid.).

Returning to Chalmers’ notion of the functional aspect of consciousness: if awareness is having access to information which can be used for control of behaviour, then perhaps phenomenal consciousness is the system’s means of *actually* accessing that information. Similarly, if access-consciousness, according to Block, is having information poised for use in reason and control of action, maybe phenomenal consciousness is actively accessing that information. Certainly the phenomenon of blindsight lends credence to this proposal, as patients *do* seem to receive the information about objects in their blind field (as they “guess” correctly much more than average) but, as they do not subjectively experience the information, they do not seem to *know* that they have this information – i.e. they cannot access the information in the normal way.

Of course, all these sketches of possible functions for phenomenal experience are largely speculative. Currently we do not know if experience is *necessary* for sophisticated sentient functioning – it undoubtedly seems that way to us, as we so often give our experience or feel as a reason for various actions. However, it may be possible that entities functionally identical to us, yet without any subjective experience, could be created in the future. But also equally plausible, is that to attain our degree of sophistication in functioning, *conscious* entities would have to be developed. At any

rate, as I have tried to show in this section, there is still plenty of room for exploring possible functions of subjective experience in sentient creatures. Also it is plausible that the very qualitative essence of experience is informational and therefore plausibly could affect our functioning and behaviour. If this is so, then in this sense experience can play a functional role and be functionally analysable. Certainly one does not have to accept an analysis like Chalmers' (and indeed Block's) which appears to prejudice the question from the start by seemingly leaving phenomenal consciousness without any functional role. Even if it appears that a functional explanation does not capture what is essential to phenomenal experience, this arguably does not preclude experience from playing a functional role, nor from being *partially* explainable in functional terms.

CONCLUSIONS

In conclusion I would like to review some of the salient points I have made in relation to Chalmers' zombie argument against a materialist view of phenomenal consciousness. I have approached Chalmers' argument from two different angles: firstly concentrating on the soundness of his argument; and then examining the notion of a zombie and of phenomenal consciousness in general.

I have tried to show that his basis for the zombie argument²⁵ – logical possibility with regard to *a priori* aspects – does not provide the support needed for the argument. With regard to possibility claims, I differentiated between two senses of logical possibility: the *framework* sense which involves judging logical possibility against current conceptual frameworks; and the *formal* sense which involves ascertaining minimally that the proposed scenario is not logically incoherent. On these grounds I argued that Chalmers' claim for the logical possibility of the existence of zombies is a claim of logical possibility in the *formal* sense. Since we do not currently have a developed framework for dealing with consciousness, the most he can try to do is to claim minimally that there is no (logical) contradiction in the notion of a zombie as he proposes it.

However, even with regard to logical possibility in the *formal* sense, in order to determine that there is no contradiction in the proposed situation we must employ the relevant conceptual frameworks to show this. As Quine (1976) says: "We have to work within some conceptual scheme or other; we can switch schemes, but we cannot stand apart from all of them" (Quine 1976, p. 65). But as our current conceptual frameworks are not as yet adequate to explain consciousness, I have argued that we cannot show one way or another whether there is a logical contradiction in this notion of a zombie or not. Arguably logical possibility claims only carry as much weight as the strength of the background frameworks against which they were made. In this sense one could perhaps

²⁵ And indeed the basis for his other arguments against materialism.

view such claims as a sort of continuum whereby logical possibility claims with regard to phenomena that can more or less be accounted for by current conceptual schemes are more “secure” and therefore carry more weight than claims with regard to phenomena at the “edge” of our knowledge, or outside these schemes.

Thus I concluded that Chalmers’ claim of the logical possibility of the existence of a zombie world is really a claim of mere epistemic, or apparent, possibility based on the current lack of knowledge in the area of consciousness. Therefore, I argued, it seems doubtful that such a claim could carry the burden of proof for the argument that materialism is false.

I also challenged Chalmers’ proposal that it is the primary intensions or *a priori* aspects of the concepts involved in arguments against materialism that are the most relevant. As I have tried to show, logical entailment of a phenomenon by the physical is generally only developed with the help of *a posteriori* discoveries. We did not always think that biological facts were *logically* entailed by physical ones, as Chalmers claims they are. This would have been a startling idea many centuries ago. Presumably once we started discovering various physical underpinnings to biological life, we started to slowly adjust our (what Chalmers terms) *a priori* concept of what the biological is and started defining it in functional terms. So I argued that when we currently do not know what the underlying nature of consciousness is, why should we expect that there be any logical entailment in this regard at this stage? Thus I suggested that Chalmers is demanding too much of a materialist theory by looking for such entailment at this point.

Having queried a number of the foundations of Chalmers’ zombie argument, I then explored some of the concepts involved in the argument itself. I examined Chalmers’ notion of a zombie and tried to demonstrate that even if we accept this notion (as presumably materialists would not accept it) and reason further about it, the notion seems to result in incoherencies. Chalmers stipulates that his zombie be functionally

(and physically) identical to a human, and yet that it have no subjective experience. The bizarre consequence that such a creature who by definition has no experience, would nevertheless believe itself to have experience, I argued, indicates that these two stipulations in the definition of a zombie are at odds with each other.

In addition, I pointed out, this tension in Chalmers' definition of a zombie paradoxically appears to result in a *difference* in functioning between these zombies and their human counterparts because many of the former's reports would be systematically false, unlike the latter's. Of course Chalmers might respond that the fact that a zombie's claims about its experiences are non-veridical does not affect its functioning which will remain identical to human functioning – indeed *must* remain identical to human functioning for this is stipulated in his definition of a zombie. But it seems that a creature which is systematically wrong about itself and hence regularly produces untrue statements about its own states will not function in exactly the same way (at least in a social context) as one which does not suffer from this defect, since the former will thereby give incorrect information about itself and the latter will not. However perhaps whether one thinks this will affect the zombie's functioning depends on one's dualist or materialist leanings, as indeed does conceivability of a zombie itself. But arguably these tensions in Chalmers' zombie stipulation show that the notion itself is fundamentally incoherent and thus, I proposed, cast doubt on the logical possibility of such a zombie.

At any rate the debate over whether one can truly conceive of such a creature exactly like a human in every way but lacking phenomenal consciousness, relies on the assertion that phenomenal consciousness does not seem to be of a kind that can be explained in functional or structural (and therefore, the argument goes, physical) terms. So I have also argued that Chalmers' initial division of mind into psychological and phenomenal aspects leads to an artificially polarised view of mind whereby phenomenal experience does not seem to be necessary for any mental functioning. Although he does

not deny that phenomenal consciousness could have a function, Chalmers' subsequent analysis nevertheless leaves it completely devoid of function. Even though the phenomenal may appear not to be of a kind that could be explained by any functional role, I have argued that this does not mean that it does not play any such role in our mental life; nor that it could not at least be *partially* definable in a functional way. Arguably investigating any such potential roles for phenomenal consciousness is the beginnings of a better understanding of it. As I discussed, Chalmers himself even hints at such functional roles for experience such as an informational role, although in line with his naturalistic dualism he presumably interprets this as mere correlation of experience with a function.

However, as I pointed out, experiences certainly do *seem* to play a (functional) role in guiding our behaviour and indeed may be essential for functioning in the way that sentient beings do. In the same way that we have a constant proprioceptive mapping of the position of our body in three-dimensional space so that we know where each limb is in relation to the rest of the body, it seems plausible that subjective experience provides us with a similar sort of mapping of the *state* of our organism as a whole. Antonio Damasio (1996) suggests something like this when he says that “the essence of a feeling²⁶ may not be an elusive mental quality attached to an object, but rather the direct perception of a specific landscape: that of the body” (Damasio, p.xvi). He argues that pain and pleasure “are the levers the organism requires for instinctual and acquired strategies to operate efficiently” (Damasio, p. 262).

Such lines of thought, whereby mental functioning and consciousness are more involved with the body than had been previously realised, are currently being pursued in cognitive science areas. Researchers have also moved on from only studying what

²⁶ Damasio's notion of “feeling” seems to be quite similar to (although perhaps not exactly the same as) the notion of subjective experience. He describes feeling as “[t]hat process of continuous monitoring, that experience of what your body is doing *while* thoughts about specific contents roll by” (Damsio, p. 145) and as a “sense of being” (Damasio, p. 150).

Chalmers calls the psychological aspect of mind (learning, memory, etc.) to looking for an explanation of experience itself. The fact that there are as yet no real answers in this area is often used as fuel for the anti-materialist position in debates on consciousness. However, Damasio contends that “Cartesian-based neglect of the mind in Western biology and medicine” has had the consequence that the “effort to understand the mind in general biological terms has been retarded by several decades, and it is fair to say that it has barely begun” (Damasio, p. 256). Arguably it is the kind of artificial polarisation of the issue in which I have suggested Chalmers’, among others, association of apparently all mental functioning with a proposed non-phenomenal (psychological) aspect of mind rather than with the phenomenal aspect of mind results, that can hamper such development.

However, in his defence Chalmers argues that we need some sort of functional or structural analysis of our *a priori*²⁷ notion of a concept in order to be able to explain it in physical terms. This of course is the crux of the whole issue and undoubtedly does currently present a serious problem for materialist theories. But since I have argued that Chalmers’ argument against materialism is unsound and so does not prove that materialism is false, I believe that the avenue of interpreting any observations of “correlations” between causal roles and experience as *potential* functions of experience, should not be closed off at this early stage. As I have mentioned Chalmers himself does seem to allow that the qualitative essence of experience could be informational. Of course, following his naturalistic dualist theory, he does not interpret this in terms of a function of phenomenal consciousness but merely as a lawful correlation. But, as I have pointed out, it is open to the materialist to interpret any informational role as a function of the qualitative essence of experience. And consequently it may be that this qualitative essence of experience is to an extent greater than some would allow, explainable in terms of structure or function.

²⁷ In Chalmers’ terms: that is, relative *a priori* – before we know what a phenomenon is *a posteriori*.

Perhaps our conceptual frameworks regarding quality in general may have to be revised to be able to view, and thus describe, quality in terms of the physical (if this is indeed possible) before we can solve the “mind-mind” problem. This would possibly be in a fashion similar to the way we presumably had to change our conception of life and begin to see it in a functional and structural way in order to be able to explain it in purely physical terms, as Chalmers asserts we can.

In conclusion, through querying a number of the claims and proposals that form the basis of his arguments and through investigating some of the concepts involved, in this thesis I have endeavoured to show that Chalmers’ case against a materialist view of phenomenal consciousness ultimately does not succeed. Chalmers certainly presents an intuitively powerful argument against materialism and, I believe, helps clarify many of the issues involved. However, though his contentions undoubtedly constitute major problems to be addressed by the materialist, I have tried to show that they nonetheless do not, as he suggests, demonstrate that the materialist project is a failure.

BIBLIOGRAPHY

- Block, Ned. (1995). 'On a confusion about a function of consciousness.' *Behavioral and Brain Sciences*, 18, 227-87.
- Bringsjord, Selmer. (1999). 'The Zombie Attack on the Computational Conception of Mind.' *Philosophy and Phenomenological Research*, 59(1), 41-69.
- Chalmers, David J. (1996). *The Conscious Mind*. New York: Oxford University Press.
- _____. (1999a). 'Materialism and the Metaphysics of Modality.' *Philosophy and Phenomenological Research*, vol. 59, no. 2, 473-96.
- _____. (1999b). 'Précis of *The Conscious Mind*.' *Philosophy and Phenomenological Research*, vol. 59, no. 2, 435-8.
- _____. (2002). "Does Conceivability Entail Possibility?" In Tamar Szabó Gendler and John Hawthorne, eds., *Conceivability and Possibility* (Oxford: Clarendon Press), 145-200.
- Crane, Tim and Mellor, D.H. (1990). 'There is No Question of Physicalism.' *Mind*, vol. 99, no. 394, 187-206.
- Damasio, Antonio R. (1996). *Descartes' Error*. London: Papermac.
- Dennett, Daniel, C. (1995). 'The Unimagined Preposterousness of Zombies.' *Journal of Consciousness Studies*, 2(4), 322-26.
- DeRose, Keith. (1991). 'Epistemic Possibilities.' *The Philosophical Review*, vol. C, no. 4, 581-605.
- Flanagan, Owen and Polger, Thomas. (1995). 'Zombies and the Function of Consciousness.' *Journal of Consciousness Studies*, 2(4), 313-21.
- Fodor, J.A. (1974). 'Special Sciences (or: the Disunity of Science as a Working Hypothesis).' *Synthese*, 28, 97-115.
- Güzeldere, Güven. (1995). 'Varieties of Zombiehood.' *Journal of Consciousness Studies*, 2(4), 326-33.
- Jackson, Frank. (1982). 'Epiphenomenal qualia.' *Philosophical Quarterly* 32, 127-36.
- Kripke, Saul A. (1980). *Naming and Necessity*. Oxford: Blackwell Publisher Ltd. (Originally published 1972).
- Levin, Janet. (1998). "Qualia." In Edward Craig, gen. ed., *Routledge Encyclopedia of Philosophy* (London: Routledge), 863-7.
- Lewis, David. (1986). *On the Plurality of Worlds*. Oxford: Blackwell Publisher Ltd.
- Mayes, G. Randolph. (1990). 'Ross and Scotus on the Existence of God: Two Proofs from Possibility.' *Thomist*, 54(1), 97-114.

- Moody, Todd. (1994). 'Conversations with Zombies.' *Journal of Consciousness Studies*, 1(2), 196-200.
- _____. (1995). 'Why Zombies won't Stay Dead.' *Journal of Consciousness Studies*, 2(4), 365-72.
- Nagel, Thomas. (1974). 'What is it like to be a bat?' *The Philosophical Review*, 83, 435-50.
- Putnam, Hilary. (1975). "The meaning of 'meaning'." In *Mind, Language and Reality; Philosophical Papers Volume 2* (New York: Cambridge University Press), 215-71.
- _____. (1994). "Rethinking Mathematical Necessity." In James Conant, ed., *Words and Life* (Cambridge Mass.: Harvard University Press), 245-63.
- Quine, W.V. (1976). "The Limits of Knowledge." In *The Ways of Paradox and Other Essays* (Cambridge, Mass.: Harvard University Press), 59-67.
- _____. (1980). "Two Dogmas of Empiricism." In *From a Logical Point of View*, (Cambridge Mass.: Harvard University Press), 20-46. (Original article published in 1951).
- Reinhardt, Lloyd. (1978). 'Metaphysical Possibility.' *Mind*, 87, 210-29.
- Searle, John. (1992). *The Rediscovery of the Mind*. Cambridge, Mass.: MIT Press.
- Seddon, George. (1972). 'Logical Possibility.' *Mind*, vol. 81, no. 324, 481-94.
- Sidelle, Alan. (2002). "On the Metaphysical Contingency of Laws of Nature." In Tamar Szabó Gendler and John Hawthorne, eds., *Conceivability and Possibility* (Oxford: Clarendon Press), 309-36.
- Siewert, C. (1994). *Understanding Consciousness*. PhD diss. University of California, Berkeley.
- Smart, J.J.C. (1995). "Sensations and Brain Processes." In Paul K. Moser and J.D. Trout eds., *Contemporary Materialism: A Reader* (London: Routledge), 93-106. (Original article published in 1959).
- Stalnaker, Robert. (2002). "What is it Like to be a Zombie?" In Tamar Szabó Gendler and John Hawthorne, eds., *Conceivability and Possibility* (Oxford: Clarendon Press), 385-400.
- Yablo, Stephen. (1993). 'Is Conceivability a Guide to Possibility?' *Philosophy and Phenomenological Research*, 53, 1-42.