

AN EMPIRICAL CASE AGAINST MATERIALISM

Andrew Clifton

Abstract

Empirical arguments for materialism are highly circumstantial—based, as they are, upon inductions from our knowledge of the physical and upon the fact that mental phenomena have physical correlates, causes and effects. However, the qualitative characteristics of first-person conscious experience are empirically distinct from uncontroversially physical phenomena in being—at least on our present knowledge—thoroughly resistant to the kind of abstract, formal description to which the latter are always, to some degree, readily amenable. The *prima facie* inference that phenomenal qualities are, most probably, *non-physical* may be resisted either by denying their existence altogether or by proposing that they are properties of some peculiar sort of mysterious physical complexity, located, for example, within the functioning of the brain. It is argued here, however, that the first, eliminative hypothesis is empirically absurd—while the second is extravagant, vague, *ad hoc* and (for various additional reasons) profoundly implausible. Taken together, these considerations provide a compelling empirical case against materialism—yet its converse, mentalism, is usually regarded as subject to serious difficulties of its own. I conclude by suggesting empirical and theoretical desiderata, respectively, for the vindication of materialism and alternatively, for the development and defense of a potentially robust and viable mentalist theory of consciousness.

1. INTRODUCTION

In the context of the philosophy of mind, materialism is the view that mental entities—if they exist at all—are not metaphysically primitive. Mental entities may be understood here to include the various subjective phenomena of conscious experience, such as beliefs, feelings, desires, sensations, thoughts and ideas. Superficially at least, these familiar denizens of the mental realm are strikingly distinct from the external objects of our experience; a *thought* or a *sensation* seems altogether different in kind from, let's say, a rock, a tree, a waterfall, a hunk of cheese or a chemical reaction. For the materialist, however, this observation is nothing other than a beguiling 'Cartesian intuition'. The fundamental constituents of the world are exclusively 'physical'—a notion which remains notoriously hard to define,¹ but which is generally considered, at least at the level of primitive ontology, to exclude all things mental. It follows that the subjective phenomena of consciousness—if we concede their existence at all—must be assumed to consist in complex, *polyadic* or *relational* physical properties; which is to say that they are structural and/or dynamic relationships, or patterns, which obtain between instances or combinations of fundamental physical entities.²

It seems appropriate, then, to call the converse of materialism '*mentalism*'. On this view, at least *some* mental phenomena—whose existence, of course, the mentalist must needs defend—are *sui generis*, fundamental constituents of the world; they are, in other words, *monadic* or *intrinsic* properties of conscious experience itself—and hence, they cannot be reduced to a manifold of non-mental components.³

¹ Difficult, that is, without circularity or *a priori* reference to mental concepts.

² I understand this definition of materialism to encompass all of its various technical varieties—including analytical functionalism, type and token identity theories and 'non-reductive' or mysterian versions of materialism (or at some authors prefer, physicalism).

³ Mentalism, on this definition, is not necessarily restricted to traditional *substance dualism*, which posits the existence of two kinds of stuff, mental and physical; for it also encompasses monistic theories such as idealism

It is widely believed that of these two opposing views, materialism enjoys overwhelming empirical support—and mentalism, none at all. This popular consensus arises, I suggest, from the confluence of four great currents of conventional wisdom—which I will call, respectively, the arguments from *disenchantment*, from *correlation*, from *computation* and from *causal closure*. Let us consider these each in turn.

First, a vast array of phenomena *other than those of conscious experience* can be readily explained in terms of formal, spatio-temporal relationships of one sort or another, which obtain among fundamental physical entities—without any need for an assumption that these primitive constituents possess mental properties. Many such natural phenomena were once commonly regarded as manifestations of mind, in the form of animistic gods, demons or spirits; yet the scientific descriptions now available to us seem to have no need of such romantic hypotheses. On the contrary, in each particular case, we have direct empirical evidence to suggest that the phenomenon in question is exclusively physical. This is held to provide, at least, a *prima facie* basis for the expectation that consciousness itself might also, some day, be so understood.

Second, neuroscientists have discovered that a great many subjective mental phenomena are closely correlated with physical events in the brain; thus, when certain groups of nerve cells fire, certain kinds of experience are reported—and specific injuries to certain parts of the brain, or abnormalities in the metabolism of neurotransmitters etc., can cause particular kinds of mental dysfunction, including changes of subjective affect. The evident *causal dependency* of subjective experience upon neurological events has led many contemporary philosophers and scientists to suppose that they are, more than likely, one and the same.

Third, the parallel development of cognitive psychology and computer science and has provided much evidence to the effect that artificial machines are capable of carrying out what seem to be more-or-less the same sort of *mental work*, that is, ‘information processing’, as we perform when we consciously think, perceive and reason. Such cybernetic devices, while still rather limited in some respects, have nevertheless made astonishing advances over recent decades. This impressive technical track record is widely considered to suggest that *all* mental phenomena are the product of some form of *computation*—which may be readily implemented within a suitably programmed (and of course, wholly physical) calculating machine.⁴ On this view, since nothing that we normally *do* with our mental faculties seems to require the involvement of a ‘ghost in the machine’, we can dispense with that hypothesis.

Forth, it is commonly believed that science has persuasively demonstrated what is known as the ‘completeness of physics’. In other words, our studies of causal interactions between uncontroversially physical entities strongly suggest that all such physical events are wholly explicable as consequences only of other physical events, in accordance with strict physical laws which permit no exceptions. If we conclude, on these grounds, that the physical world is altogether ‘causally closed’, then a hypothetical non-physical mental entity, such as an immaterial spirit or soul, can have no influence upon it. Common-sense experience tells us, however, that our first-person thoughts and feelings have definite causal affects upon our physical behavior; hence it seems clear that such mental phenomena, if they exist at all, must be physical too.⁵

For many contemporary philosophers and scientists, these arguments are sufficient to establish materialism as the *sine qua non* of modern metaphysics. There is no conceivable, scientifically credible, intellectually comfortable alternative. Whatever difficulties it may have in accounting for certain embarrassingly recalcitrant mental phenomena; however strange and counterintuitive its implications may seem—*materialism must be true*.

and panpsychism. The former denies the existence of physical entities altogether, whereas the latter asserts that the intrinsic nature of fundamental physical entities is irreducibly mental.

⁴ I criticise this view, commonly known as ‘Strong AI’, more directly in Clifton (2004) [b], [c] and [d].

⁵ I address the issue of the supposed completeness of physics in further detail in Clifton (2004) [a].

I disagree. Materialism is by no means immune either to doubt or criticism; most obviously, because the four major arguments in its favor are so thoroughly *circumstantial*. In each case, the materialist's conclusions regarding the fundamental nature of mental entities are based almost entirely upon our knowledge and understanding of *non-mental* entities; indeed, the only consideration of mental phenomena which plays a part in these arguments consists in the observation that they seem to have *physical* correlates, causes and effects. Properties of mental items *as such* are not mentioned at all—and this raises an obvious question: if mental phenomena are indeed—just like all other natural objects, properties or processes—exclusively physical, why, in this particular case alone, do we not have *direct evidence* to this effect?

It is surely open to the mentalist to reply with an argument to suggest that this is because mental phenomena possess distinctive properties which are altogether incompatible with a materialist explanation; hence, we *lack* the direct evidence of physicality which *ought to be discernible* if materialism were true. If such a rebuttal of the argument from disenchantment can be reasonably defended on empirical grounds, it will establish, at least, a significant *probability* that mental phenomena are *non-physical*—and hence, will to that degree, weigh against the empirical force of each of the remaining three circumstantial arguments.

Amongst a variety of mental phenomena which are sometimes thought to challenge materialism in this way, the most striking, in my view, are the *qualitative* characteristics of conscious experience.⁶ In recent years, there has been much debate as to whether or not such subjectively familiar properties—variously known to philosophers as ‘phenomenal qualities’, ‘sense data’, or ‘qualia’—can ever be accounted for within a strictly materialist theory of the mind. Unfortunately, few discussions of these worries are very clear about what the foregoing expressions mean. A typical definition of ‘qualia’, for example, gives us only: ‘those properties of mental states or events, in particular of sensations or perceptual states, which determine “what it is like” to have them.’ (Shoemaker 1990 p. 666). The scare quotes seem to hint at an important difference between ‘what it is like’ to have an experience and what it is like, for instance, to be a rock;⁷ yet the definition falls short of conveying any sense of what that difference might be. Some philosophers, no doubt, would point to various controversial, theoretical properties of qualitative experience such as *irreducibility* or *intrinsicity*—whilst others, skeptical of these alleged characteristics, are inclined to deny the existence of ‘qualia’ altogether (see e.g. Rorty 1970 1979; Churchland 1983; Dennett 1988; Harman 1990; Rey 1996, 1997).

What, then, do we really mean when we talk about ‘qualia’ or ‘phenomenal qualities’—and what, if anything, is special, or distinctive, about them? A great deal of attention has been paid, at least, to the second of these questions, yet there is remarkably little agreement as to how it should be answered. Part of the problem, lies, I think, in the fact that much of the literature on this topic deals with the validity, or otherwise, of widespread intuitions about the epistemic implications or logical possibility of various *imaginary situations*. ‘Knowledge arguments’, for example, imagine the consequences of our obtaining complete knowledge of the physical details of some aspect of our own, or another's, neurological activity (see Nagel 1974; Jackson 1982); ‘conceivability arguments’ on the other hand, assert (or deny) that we can ‘coherently imagine’ the existence of an exact duplicate of a normal human being, identical in every detail except that qualia are either *absent* (Block and Fodor 1972; Kirk 1974; Squires 1974; Robinson 1976; Block 1980) or *inverted* on various qualitative dimensions (Lycan 1973; Block 1990; Levine 1991; Johnsen 1993). While materialists typically claim that the intuitions provoked by these arguments are unfounded, or at least misleading, anti-

⁶ Another, of course is intentionality—a detailed treatment of which is beyond the scope of the present paper. see, however, Clifton (2004) [a].

⁷ Assuming, for the moment, that rocks lack conscious awareness, and that to be a rock is, quite simply, to have rock-like physical properties. See Sloman 1999.

materialist critics tend to take them at face value, challenging the materialist to supply *proof* to the contrary. Materialists, however, generally have strong intuitions of their own, inclining them to deny the need to do so; there just has to be something wrong with such troublesome dualist intuitions, and that's that. In any case, it seems impossible to decisively establish *anything* about the workings of empirically inaccessible, imaginary realms in which philosophical intuitions compete for our allegiance. The result, it seems to me, is a frustrating stalemate. In this essay, therefore, I wish to canvass an alternative approach, which I believe to be more fruitful.

My starting point is to consider what we can definitely say, empirically,⁸ about the qualitative characteristics of experience, without invoking metaphysical assumptions or relying upon intuitions prompted by bizarre and impracticable thought experiments.

In §2 I will introduce an empirical definition of phenomenal qualities. This theoretically neutral definition does not *presume* that phenomenal qualities are non-physical, but it does allow us to develop a compelling argument to this effect. The 'description argument', set out in §3, rests upon the observation that physical phenomena are generally amenable to a strictly physical description (that is, a non-ostensive formal description, solely in terms of structural and/or dynamic relationships obtaining amongst physical entities)—at least to *some* degree of approximation; indeed, this is precisely why we consider them to be physical. On our present knowledge, however, phenomenal qualities cannot be described in this way at all, not even to the slightest degree of approximation; hence, *prima facie*, they are probably not physical.

In the absence of any proven formal descriptions of phenomenal qualities, it seems to me that the materialist has two possible means of resisting this conclusion: first, to flatly *deny* the existence of phenomenal qualities; second, to propose that phenomenal qualities are complex physical phenomena of some peculiar and unusual kind, whose special relational properties *somehow* determine their subjective character, while also, *somehow*, concealing from us all evidence to this effect. In §§4-5 I argue that these claims are, respectively, empirically absurd and (very implausibly) *ad hoc*.

In §6 I conclude that on our present knowledge, we cannot justify the confident assumption that materialism is true; nor, alas, can we just casually assume that its converse is true. As possible routes out of this perplexing dilemma, I outline, on the one hand, what I consider to be the minimal empirical requirements for the vindication of materialism—and on the other hand, the formidable demands we must make of any mentalist theory of consciousness that would seek to take its place.

2. PHENOMENAL QUALITIES

2.1 An empirical definition

In order to adequately define 'phenomenal qualities', it will be helpful, I suggest, to clarify two issues: namely, what it is that makes them 'phenomenal' and what it is that makes them 'qualities'. The first requirement seems easy enough: a phenomenal *Q* is a *Q* considered from the perspective of first-person conscious experience. On this basis, for example, Andrew Bailey defines '*qualia*' quite straightforwardly, as "properties considered from the first-person perspective." (Bailey 1998, 2003). This strikes me as a definite improvement upon the '*something it is like*' approach, but it still seems a little too broad. It would include, for example, the phenomenal 'roundness' which features within my first-person awareness of viewing a spherical object; yet this sort of subjective impression does not seem to be *qualitative*, in the usual sense of the word. Ned Block, I think, does slightly better, defining a 'quale' as "...an

⁸ that is, in this context, based upon rational consideration and interpretation of sensory experience.

aspect of subjective experience that cannot be shown by *a priori* or other armchair means to be intentional, functional or cognitive” (Block 2003). Paul Churchland takes a similar approach, defining ‘qualia’ as “...those intrinsic or monadic properties of our sensations discriminated in introspection” and notes that “The quale of a sensation is typically contrasted with its causal, relational, or functional features.” (Churchland 1989 p. 23). Yet both definitions fail, I think, to specify the respect in which qualia *seem*, at least *a priori*, to be ‘intrinsic’, ‘monadic’ and distinct from the categories mentioned.

That respect, I suggest, is *qualitativity*—a notion we may clarify, by means of a distinction between qualitative and ‘non-qualitative’ or *formally describable* properties. The latter may be defined as properties that can be identified and recognized by means of *exclusively abstract* description or definition, such as spatio-temporal relationships, numerical and logical properties etc. Qualitative properties, on the other hand, are those that we find—at least in our own case and to the full extent of our experience—*impossible* to formally describe; and which, therefore, we can only identify ostensively (that is, by directly or indirectly pointing them out). The paradigmatic examples of such qualitative properties are, of course, familiar phenomenal impressions such as colours, tastes, smells, pains and the like. We may say, therefore, that a particular, introspectively individuated sensation exhibits a particular ‘phenomenal quality’ when its qualitative character is (or seems to be) *homogenous* (that is, invariant) throughout its phenomenal extent (that is, its duration and extent in phenomenal space, as when a patch of color occupies a particular area within the visual field). We may summarise these considerations as follows:

A ‘phenomenal quality’ is a qualitative aspect, homogenous in its phenomenal extent, of any subjectively individuated feature of conscious experience. It is distinct from other aspects of experience in manifesting a particular character which is, at least, *very difficult* to formally describe.

This definition has, for me, the particular merit of being theoretically neutral. It does not presume that phenomenal qualities are *fundamentally* irreducible or non-physical in nature, or that we could never, somehow, *learn* to formally describe them; nor does it stipulate that they are necessarily private, that our knowledge of them is incorrigible, that they are mental ‘objects’ as opposed to (e.g.) processes or ‘adverbial’ modes of perceiving; or even that they are exclusively ‘mental’. We might, perhaps, reserve the rather more ethereal term ‘qualia’ for phenomenal qualities *considered as having such controversial properties*. It should be clear, however, that the term ‘phenomenal qualities’ does not, on the present definition, refer to theoretical abstractions or hypothetical entities whose existence is proposed in order to *explain* the way things seem,⁹ but rather, to salient features of conscious seemings; as abundantly evidenced—for the benefit of those philosophers who distrust the first person perspective—in the explicit reports and comments we make concerning them.

There remains, however, one important consideration which needs to be clarified. Many of the familiar sensations which we would recognize as phenomenal qualities, in the foregoing sense, can, in fact, be *partially* described—but only as combinations or mixtures of qualitative components, such as ‘reddish-brown’, or ‘sweet-and-sour’.¹⁰ These are, in other words,

⁹ Such as, for instance, our conceptions of physical entities!

¹⁰ Some people would argue that all of the phenomenal qualities we normally experience are like this, because, as William James pointed out, “no one ever had a simple sensation by itself” (James, 1890 p.224); so we never encounter a ‘pure’, *simple* quality in isolation from others. Be that as it may, some phenomenal qualities (such as phenomenal *red*) seem very much simpler, in this respect, than others (for example, the unique aroma of Great Aunt Hildegard’s gumbo sauce). Within the former, *subjectively simple* phenomenal qualities, we cannot discriminate a *complex* mixture of qualitative components—if indeed, they are mixtures at all. It would be absurd, in any case, to suggest that *all* phenomenal qualities (including the discriminable components of complex

qualitative as opposed to strictly formal, descriptions. We can, of course, tease out from them a purely formal account, in terms of the number and relative proportions of the qualitative components; but since this tells us nothing about their qualitative nature as such, it does not provide even the smallest step towards a formal description of the qualitative character of the sensation as a whole. In so far as the discernible ingredients which constitute such qualitative mixtures are thoroughly hard-to-describe, the *mélange* is, *a fortiori*, hard-to-describe.

It is most important, however, not to confuse a phenomenal quality *per se*, be it simple or ‘mixed’, with what I will call a ‘qualitative manifold’; that is, a spatio-temporally structured array of multiple qualitative elements, arranged in some sort of pattern. This violates the principle, included in the foregoing definition, that a *particular* phenomenal quality must be *homogenous* in character throughout its phenomenal extent. Mixing flavorsome ingredients within a marinade gives rise to a composite sensation which may still, legitimately, be called a *flavor*; but my entire gustatory experience of a six-course banquet is no such thing. Likewise, my overall impression of hearing a Bach cantata is not a timbre and my visual experience of viewing an oil painting is not a hue. There are some philosophers and cognitive scientists who like to treat such highly structured, complex, information-rich episodes as ‘qualia’ (or at least, to present them as typical examples of experiences which we tend to regard as ‘ineffable’).¹¹ There is, of course, ‘something it is like’ to have them and they certainly do include multiple qualitative components; but unlike these, they cannot reasonably be construed as qualitative *types*. It would be misleading, furthermore, to conflate their constituent qualitative aspects with their rich and complex features of structure, dynamics and pattern.

2.2 *The valuative dimension*

I take the foregoing definition to include the *valuative* features of conscious experience, that is, the sense of personal value or disvalue that we associate with particular episodes, thoughts or sensation. As Daniel Dennett has observed—notwithstanding his unequivocal position as a strict reductionist—consciousness certainly *seems* to be ‘a source of mattering’ (Dennett 1991, p. 33). That is to say, we find it hard to doubt that feelings, emotions, joys and pains are important, and in varying degrees, ‘good’ or ‘bad’ *in themselves*—regardless of any secondary, functional roles they might have in terms of survival, reproduction and so on.

Familiar impressions of this kind have prompted some philosophers to argue that personal intrinsic value is a property *only* of conscious experience and cannot be assigned to contingent events or conditions such as illness or health, death or survival, reproduction or extinction—because none of these things can be identified as either ‘good’ or ‘bad’ in themselves. On this view, such eventualities only matter to us by virtue of the contributions they can make to the quantity and quality of our conscious experience.¹² Furthermore, just like sensory qualities, the value that we find within conscious experiences seems impervious to analysis; it appears that in this case at least, as G. E. Moore declared, “‘good’ is a simple notion just as the color

quantitative mixtures) are *complex* quantitative mixtures, as this would entail an infinite regress in which ordinary, straightforward sensory impressions become infinitely complex.

¹¹ The Bach Cantata is Dennett’s example (See Dennett 1991, p. 387). Gerald Edelman and Giulio Tononi claim that ‘... *each differentiable conscious experience* represents a different quale, whether it is primarily a sensation, an image, a thought, or even a mood and whether, in retrospect, it appears simple or composite.’ (Edelman and Tononi 2000, p.158). They go on to suggest that: “A pure perception of red [that is, a state in which the entire visual field is perceived as uniformly red] is as informative as a perception of a busy city street because they both rule out a more or less equal number of conscious states.” (p. 168). But what about the redness of a London bus, or the blueness of a child’s baseball cap? Clearly, these relatively simple perceptual elements cannot be equal in informative complexity to the entire city-street vista of which they are each, merely, a part.

¹² See Lemos 1994, for a thorough discussion. Amongst philosophers who accept the notion, there is some controversy as to whether or not ‘intrinsic value’ may belong only to conscious experiences; but widespread agreement that at least some, if not all, conscious experiences possess it.

yellow is a simple notion.¹³ (Moore 1903/1993 p. 59). Certainly, no formal description of phenomenal ‘goodness’ is currently available, by means of which, for example, we might expect to recognize that property in a particular pattern of neural activity.

Thus, phenomenal qualities are not only mysterious and hard to describe: they may also be profoundly *important* and *valuable*—and not indirectly, by virtue of their extrinsic functions, causes and effects, but simply and immediately, by virtue of their very nature; that is, of the way they seem.

2.3 Functional roles and phenomenal content

The difficulty of describing phenomenal qualities does not prevent us from mapping out their similarities and differences, cataloguing their physical correlates, causes and effects—and then considering what, if anything, they *do*. The last of these questions is somewhat controversial, since some philosophers maintain that phenomenal qualities are *epiphenomena*, having no functionality at all.¹⁴ It can, indeed, be plausibly argued that phenomenal qualities *as such* are ‘functionally redundant’, in so far as they do not appear to be *needed* for the functional roles which they appear to serve; yet it does not follow that in fact, they *do nothing at all*. A thing can exhibit properties which are, so to speak, *surplus to requirements* for a given functional role and yet still remain functionally effective. I do not, strictly, *need* to use my twenty pound sledgehammer to crack a nut, but it works every time. Likewise I do not, necessarily, *need* an experience of phenomenal *blue* to be informed that light rays of wavelength $\cong 4 \times 10^{-5}$ cm are striking my retinas; but the sensation serves that purpose, well enough.

Both common-sense experience and systematic, experimental investigations in psychology and psychophysics provide us with a fairly clear conception of the functional roles which phenomenal qualities appear to serve within conscious cognition. They seem, first of all, to flag the occurrence of a wide range of sensory discriminations—whose subjective intensity is typically related, in moderately complex, context-dependent, yet scientifically intelligible ways, to measurable characteristics of particular physical stimuli (see, e.g., Land 1976, 1986; Block 1995; Gregory 1996; Gescheider 1997). In so far as particular phenomenal qualities possess a distinctly valuative dimension, they also seem to stimulate or inhibit voluntary behavior (see, e.g., Spencer 1880; Troland 1928; Young 1959; Bindra 1969; Bozarth 1994; Wall 1999; Hardcastle 1999). In short, they appear to function, in general, as *signs, signals and switches*.¹⁵ Each phenomenal quality has, accordingly, a particular individuating character which distinguishes it from others—and which may, furthermore, locate it within a multi-dimensional functional space of more-or-less similar items (see, e.g., Bartoshuk 1978; Hardin 1993; Clark 1993). Our impression of this character may vary in one or more quantitative dimensions (such as pitch, intensity) while still remaining recognizably the same. In addition, phenomenal qualities often come, so to speak, with *metadata* attached; they might be, for example, located in sensory space, either within or outside our bodies (Clark 2000).¹⁶

We may consistently allow, then, that phenomenal qualities appear to have certain discernible functions; that certain formal relations obtain between instances of their various types and also that they are typically locatable in phenomenal space. These considerations,

¹³ I do not wish to endorse, here, Moore’s ‘open question’ argument for the irreducibility of the good, but rather to point out that, like the color yellow, the phenomenal ‘goodness’ of a joyful experience is something we find it hard—perhaps, impossible—to describe or define.

¹⁴ The notorious difficulties associated with this position need not detain us here. For a more detailed critique of the epiphenomenalist view, see Clifton (2004) [a].

¹⁵ ‘Controls’ would perhaps serve as a better (but still imperfect) metaphor here than switches; for our behavioural responses to such motivational influences are not always binary (that is, like a light switch, ON or OFF) but may be mitigated, resisted or suppressed altogether by a wide variety of contextual factors.

¹⁶ We may by all means speculate that phenomenal qualities might have, in addition, various other more mysterious functions, besides those outlined here—but at present, we have no hard evidence to support such a view.

however, tell us nothing about the qualitative nature of phenomenal qualities *as such*—and we therefore have no reason to suppose, with some of the aforementioned authors, that phenomenal qualities may be readily and straightforwardly *identified* with these functions and/or with the associated formal relationships (Hardin 1993; Clark 1993, 2000). Were this so, formal descriptions of these relational properties would serve as valid descriptions of the associated phenomenal qualities—and this is not the case. Those who say that there is *nothing further to be explained* beyond the functional relations mentioned above are, in effect, denying the existence of phenomenal qualities, as presently defined.¹⁷

2.4 Conclusions

In summary, the present empirical definition of phenomenal qualities:

- (1) is theoretically neutral, i.e., it is independent of metaphysical assumptions, philosophical and scientific theories of the nature of perception etc.;
- (2) distinguishes phenomenal qualities (that is, subjectively homogenous sensory impressions such as tastes, colors etc.) from *qualitative manifolds*—which include multiple distinct phenomenal qualities, arranged into complex, yet formally describable patterns.
- (3) includes the subjectively ‘valuative’ features of experience, which, at least, *seem* to serve as a source of ‘mattering’;
- (4) is compatible with (and is by no means *eliminated* by) evidence indicating the apparent *functions* of phenomenal qualities, that is, their cognitive and motivational roles as ‘signs, signals and switches’.

The *existence* of phenomenal qualities—on the present definition—may, I think, be reasonably accepted on a combination of first-person and third-person evidence, independently of metaphysical suppositions with regard to the nature of consciousness. Having taken its full implications into account, however, we are surely at liberty to consider which of the various metaphysical systems available to us can most plausibly accommodate them.

3. THE DESCRIPTION ARGUMENT

Consider once again the materialist proposition that the world is wholly constituted by numerous instances of a limited number of fundamental physical entities—and that all other objects, events or processes are nothing more than structures or patterns that obtain among collections thereof. It seems reasonable to ask, what empirical considerations lead us to think that this is so? Indeed, what evidence suggests to us that any *particular* phenomenon is a relational, physical one?

I suggest that the relevant evidence—indeed, the only possible positive evidence—is provided by our ability to *describe* observable objects, events and processes in abstract, spatio-temporal terms—fundamental entities, of course, being known to us only by virtue of their relational properties, that is, their spatio-temporal interactions with one another (and hence, with ourselves). While many phenomena are far too complex to be exhaustively described in this way, a partial description is generally sufficient to suggest that a maximally thorough account may be possible in principle (within the limits, of course, of quantum uncertainty). One important caveat must be mentioned here: an *ostensive reference* does not, of course, count as a description—not even a partial one. For example, given a suitably accurate map reference, it is possible to pick out St Paul’s Cathedral from other famous London landmarks; however, these coordinates do not qualify as a *description* of the building’s architecture. Likewise,

¹⁷ A further argument would, of course, be needed to justify such radically eliminative claims. I will discuss eliminative arguments in more detail in §4.

comparison and analogy may not feature in a formal description; thus, for ‘dodecahedron’, “sort of like an octahedron, but even more like an icosahedron” just won’t do—whereas, “convex regular polyhedron with twelve pentagonal faces” passes muster.

I would suggest that a non-ostensive, formal description of any phenomenon, however rudimentary or imprecise it may be, can be considered *epistemically valid* (and therefore, empirically *useful*) when, without invoking properties that themselves lack valid formal description, it allows us to construct a conceptual model of the phenomenon in question and thus to recognize and *identify* it, with significant reliability, when we encounter it ourselves, on the basis of the description alone. If a crude, yet valid, formal description of a phenomenon exists, then it seem reasonable to suppose that a process of refinement and elaboration of the description might continue, in principle, until the phenomenon is fully understood in terms of structural, logical and/or dynamic relationships between instances of a limited set of fundamental entities.

It seem to me that such primitive-but-promising formal descriptions exist in abundance, or at least, are clearly possible, for all those phenomena that are uncontroversially recognized as physical—even when they are complex, poorly understood, and difficult to describe in great detail. Living organisms, for example, are enormously complicated, both in structure and behavior. Nevertheless (with the possible exception of conscious experience), all of the familiar *biological* properties such as reproduction and growth etc., which characterize living things, can be clearly described or defined, in formal, abstract, unmythical terms—at least, to a useful degree of approximation. Therefore, we have no need to posit some strange, non-physical, vital force to account for these phenomena. Of course, we are by no means omniscient about the intricate workings of living organisms—and there is, doubtless, far more still to be learned than we currently know. None the less, our ability to provide, at least, rudimentary formal descriptions of all the general biological phenomena which characterize life serves as very strong evidence to suggest that a full understanding of these phenomena, in purely physical terms, is possible in principle.

For materialist philosophers, the example of vitalism provides a favorite cautionary tale. It is often argued, by analogy, that just as an earlier generation of philosophers and scientists were wrong to assume the existence of some ‘fifth element’, or *élan vital*, to account for the special properties of living tissue—or, similarly, to propose a strange substance, caloric fluid, to explain the phenomenon of heat—so dualists today are wrong to suggest that consciousness cannot possibly be explained in purely physical terms (see, e.g., Dennett 1996; Churchland, P. M. 1996 Churchland, P. S. 1996). However, it seems to me that in the case of qualitative conscious experience, these analogies fail: there is a crucial and obvious difference, which the materialists have persistently ignored.

Objective, physical phenomena such as heat, biological life and countless others are always amenable, at least, to *rudimentary* formal description. It is therefore inappropriate, in these cases, to dismiss the possibility of a reductive explanation—because there is positive evidence to suggest that we might find one, if we looked. When, on the other hand, we consider some simple phenomenal quality, such as that which is evoked when we perceive or visualize a particular primary color, and try to construct even a crude, rough and ready, yet valid formal description of it—without comparative reference to phenomenal qualities of any kind—we invariably fail. I base this generalization not only upon my own experience and the reported experience of others, but also upon the fact that, to my knowledge, no one has ever published a convincing account of success. If, by careful introspection, we could find within the most basic phenomenal qualities some underlying, non-qualitative, logical structure which defines their nature—something that might be expressible, perhaps, as a number, a vector or an equation—then we might indeed have the makings of a valid formal description; yet it seems to be a universal feature of human experience that we can do no such thing. No mathematical formula is known, for example, which recognizably expresses, to any degree of

approximation, the phenomenal character of the color blue, or the intrinsic disvalue of a pain. Thus, phenomenal qualities are sharply distinct from uncontroversially physical phenomena inasmuch as they present the materialist with what might be called a ‘*description problem*’: that is, we appear to lack *positive evidence* to suggest that they are, in general or in any particular case, relational physical properties; or to put this differently, they appear to possess special properties which are incompatible with such an analysis.

This problem is particularly striking when we consider specific identity claims, in which it is suggested that a particular sort of phenomenal quality is identical with the occurrence of some well-defined (or for that matter, *hypothetically* well-defined) physical phenomenon.

Now, materialist philosophers sometimes argue (by analogy, once again) that a single phenomenon may manifest itself, or may be known or understood, in various different ways which seem quite distinct—giving the misleading impression that there are two or more distinct phenomena instead of just one. Thus, the morning star and the evening star appear distinct, when in fact they are merely different *appearances* of a single object; likewise it is possible for someone to know a set of facts about Mark Twain and another set of facts about Samuel Clemens, remaining ignorant of the fact that they are one and the same person. Similarly, it is argued, we have two ways of knowing about the physical phenomenon we call a pain; via the first-person experience of having one and via a third-person, scientific description of its alleged physical basis (such as, for the sake of argument, the firing of nociceptive-specific neurons in the parietal cortex. See Papineau 1993, 1995, 2002). These distinctive appearances, or ways of knowing, may seem quite different—but this need not worry us; it is merely a trick of perspective.

The weakness of this argument should now be obvious. In the case of the cited examples, our conception of each distinct appearance, or named individual, has many points in common with its counterpart—and no anomalous features which seem inconsistent with an identity claim. Phosphorus and Hesperus may both be described as a bright object of a certain size, appearing in the sky. Twain and Clemens may each be described as a male human being, a journalist and author, who was born in Missouri in 1835 and died in Connecticut in 1910. In neither case is the suggestion that the two appearances correspond to a single individual particularly shocking or implausible. In the case at issue, however, we are presented with, on the one hand, a subjectively qualitative sensation, which we *cannot formally describe* to any extent whatsoever—and on the other hand, a neurological process which we *can* formally describe in arbitrary detail, yet without giving any hint that the qualitative sensation exists *at all*, much less, any clue as to its character. Here, the two manifest phenomena have, as far as we can tell, *nothing in common*, other than the fact that they seem to be closely correlated and causally interdependent.¹⁸

Furthermore, any conventional notion we have of the construction of a particular aspect, perspective view or manifestation of an uncontroversially physical phenomenon involves some sort of mathematical operation which simply converts one structural spatio-temporal form into another. The output of such a transformation may look very distinct from the original form, yet it always presents us with some sort of structure and it is always amenable to formal description. There are no known mathematical functions whose output is some peculiar sort of qualitative content, which altogether *lacks* structure and dynamics and which therefore defies even crude and rudimentary formal description. Indeed, the conjecture that this is *generally*

¹⁸ This disjunction, I suggest, is a major source of the intuition that ‘absent qualia’ are possible. Were it alone in this respect, then this supposed possibility could certainly not be used as a *premise* in an argument to show that the disjunction exists, that is, that phenomenal qualities have special properties which physical phenomena lack. There are other, independent grounds upon which the possibility of ‘absent qualia’ can be plausibly defended (see §5.3, n. 29). However, it seems to me that arguments based upon such notions need to be handled with great care in order to avoid paradoxical consequences (see Clifton 2004 [e]).

true seems at least as inductively sound as the hypothesis that the sum of two natural numbers is always another number—and not, lets say, a fried egg.

It therefore seems highly probable, *a priori*, that observable or experienced phenomena that consist entirely in structural and/or dynamic relationships amongst physical entities will be describable in terms of structure and dynamics; the fact that this actually applies, in practice, to all known (uncontroversially) physical phenomena strongly suggests that any detectable physical property will probably be amenable, at some level, to such a description. Thus, if phenomenal qualities are physical in nature (as materialists would have us believe), then it seems that we ought to be able to formally describe them in much the same way, at least in crude outline. We evidently cannot do so—and this surely implies that they not relational physical properties at all. The future discovery of a valid formal description of at least one instance of a phenomenal quality represents the only hope for positive evidence to the contrary—yet this hope seems unlikely ever to be fulfilled. The long-standing interest of philosophers, artists, psychologists, neuroscientists—indeed, of humanity in general—in the nature of conscious experience has surely afforded considerable opportunity for such a discovery to be made—yet no such description has every been discovered, published and recognized as genuine. In the systematic absence of such evidence, it seems reasonable to conclude, at least, temporarily, that phenomenal qualities are probably not relational physical properties; that materialism is, therefore, probably false—and hence, that mentalism is probably true. The argument may be summarized as follows:

- (1) Materialism, with respect to the nature of consciousness, is the view that subjective mental phenomena are extrinsic physical phenomena, that is, structures, relationships or spatio-temporal patterns that obtain amongst instances or collections of fundamental physical entities (whose essential nature, if any, is presumed to be non-mental).
- (2) All known extrinsic physical phenomena, including those which are poorly understood, admit to an *empirically valid*, non-ostensive, non-qualitative, *formal description*—at least to *some* degree of approximation—such that they can be identified, with significant reliability, on the basis of that description alone.
- (3) To our knowledge, all known mathematical translations or transformations of a formally describable spatio-temporal form into some particular aspect or manifestation thereof invariably yield a formally describable spatio-temporal form.
- (4) From (2) and (3), it is overwhelmingly likely that any physical phenomenon which, as materialists claim, *constitutes* the content of a particular conscious experience will be amenable, at least, to a *rudimentary* yet empirically valid formal description.
- (5) There is no evidence to suggest that an empirically valid formal description, to any degree of approximation, of any phenomenal quality has ever been discovered.
- (6) Therefore, materialism is probably false.

Our conclusion must surely be that subjectively *simple* phenomenal qualities (or perhaps, the simplest, or ‘purest’, qualitative components thereof) are metaphysically primitive. That each of these fundamental elements of conscious experience has a certain intrinsic nature must, on this view, be accepted as a brute fact; for this character cannot, even in principle, be analysed or broken down into anything else.

4. COUNTERARGUMENTS I: DENIAL AND EVASION

4.1 *Openly eliminative arguments*

One possible way to resist the conclusion of the description argument would be to claim that phenomenal qualities do not exist. Of course, as I mentioned earlier, a number of materialist philosophers have, in some sense or other, argued just that (see e.g. Rorty 1970, 1979;

Churchland 1983; Dennett 1988; Harman 1990; Rey 1996, 1997). The most plausible (or at least, charitable) interpretation of such arguments, however, is simply as rejections of any axiomatic *assumption* to the effect that experience is composed of non-physical, irreducible, ineffable entities. Certainly, these authors would be right to point out that we cannot automatically accept, *a priori*, the existence of ‘qualia’ *in this strong, mentalist sense*—but this does not, of course, justify the general dismissal of arguments which make no such initial suppositions. I have been careful, therefore, to define ‘phenomenal qualities’ *neutrally*, as salient, qualitative features of conscious experience—and it seems impossible to reasonably doubt their existence, in this straightforward sense.

In the case of valuative qualities, for example, the eliminative hypothesis flatly denies that some experiences are, as a matter of fact, in their own right, just *nicer* or *better* to have than others; that is, that the quality of our experience *matters*. Our belief that this is so is just another foolish ‘Cartesian intuition’; in reality, absolutely nothing fundamentally *matters*, for reality is exclusively physical and ‘mattering’ is a notion which cannot be defined in primitive physical terms. On this view, ‘values’—if, indeed, such dubious, ghostly language is strictly permissible—may only be regarded as instrumental, expressing functional contributions to some ultimately arbitrary goal; thus the vivid impressions I have that, *ceteris paribus*, joy is, (in some sense, if only *for me*) intrinsically ‘good’ and misery ‘bad’, are merely illusions.

Yet it is, at least, *difficult* to conceive how such impressions could be illusory. Indeed, the idea seems *empirically absurd*—in a sense which becomes clear if we consider the question: what is it like *to experience the illusion* that one is aware of a phenomenal quality, or in particular, a qualitative sensation which seems inherently *nasty* or *pleasant*? This gives rise to an argument we may call the Illusion *Reductio*. Suppose that an item Φ is defined as a *Q-like subjective impression* (where *Q-like* may be taken to stand for either a particular sort of qualitative character, or for the general property of being thoroughly hard-to-describe). The eliminative hypothesis ‘all Φ are illusory’ says that in place of a Φ , I experience an illusion Ψ , that is, a compelling subjective impression *that* I am experiencing a Φ . But in order to fool me, Ψ must itself be convincingly *Q-like* (that is, exactly like experiencing some special kind of conscious content, which I cannot formally describe); hence, Ψ is necessarily a *Q-like subjective impression*—an instance of Φ . Thus, the illusion itself falls under our definition of a phenomenal quality—and the eliminative hypothesis fails.

Wittgenstein’s celebrated argument against ‘private languages’ (Wittgenstein 1958 part 1 §§ 258-256) cannot, I think, come to the materialist’s rescue here and defeat the foregoing illusion *reductio*. Wittgenstein claims that I have no grounds for supposing that I can consistently apply a particular phenomenal term (such as ‘blue’) to a *particular* private sensation; hence I have no basis for claiming that any such ‘beast in the box’ exists. The premise is in my view, unjustified;¹⁹ but suppose we accept it, for the sake of argument. I do not have to be able to consistently recognize and apply the correct *name* to a particular sensation (such as is evoked, for example when I look up at sky on a sunny, cloudless day) in order to determine that I find it inordinately hard to describe—and therefore to affirm (whether or not it be the same sensation that I called ‘blue’ yesterday) that it is, without doubt, a phenomenal quality.

¹⁹ It seems unlikely that my impression of having clear, consistent memories of particular qualitative states, of the names I habitually apply to them and the circumstances in which they typically occur is *utterly* unfounded—given that my memory of commonplace publicly observable events, their names and typical antecedents, is, verifiably, fairly reliable. Furthermore, the correlation between my public expression of a phenomenal name and a set of objective circumstances in which (I claim) the associated private impression typically arises is publicly verifiable; and it seems unlikely that such a correspondence would obtain if my memory of phenomenal states were hopelessly inadequate—or indeed, if they did not exist at all. For further criticism of the anti-private language argument, see e.g. Ayer 1968; Walker 1989 and Robinson 1994, ch. 4.

There are certain followers of Wilfred Sellars (e.g., Brandom 1994; McDowell 1994; Rorty 1997; Williams 1999) who are disposed—unlike Sellars himself²⁰—to interpret his critique of ‘givenness’ (1956/1997) in radically eliminative terms. Sellars did not, however, deny the existence of qualitative sensation *per se*, but rather, rejected the possibility of *non-inferential knowledge*—and as I have argued, the knowledge that I am experiencing something which falls under the definition of a phenomenal quality is established *inferentially*, by means of an argument. Such knowledge is not ‘given’, but it is *self-justifying*—in much the same way that “This sentence has five words” may be verified without reference to anything other than the sentence itself.

To put this point somewhat differently, if I attend to some particular feature of my experience, such as a patch of yellow within my visual field, then attempt to formally describe the way it seems and find this task impossible, I will naturally affirm, from my stated definition, that ‘*this experience* is a phenomenal quality.’²¹ If I then consider the proposition: ‘phenomenal qualities do not exist’, I am faced with two alternative implications. Either, ‘I am not really experiencing *this experience*’, or ‘I really can formally describe this experience, even though I can’t’. Both are self-contradictory, therefore, the proposition is false.²²

4.2 *Implicitly eliminative arguments*

Perhaps for the very reason that it is vulnerable to such compelling objections, eliminativism is less often explicitly defended than it is surreptitiously implied.

There are, for example, a number of influential philosophical theories of perception which seeks to deny that there are such things as ‘mental objects’, which we somehow inwardly perceive; hence, they are widely considered to challenge the traditional notion of ‘sense data’ or qualia. They are not, for this reason, necessarily incompatible with the existence of *phenomenal qualities*, as currently defined; however, they are often presented in a way which carries this eliminative implication.

Externalism (or ‘*representationalism*’, or ‘*intentionalism*’), with respect to phenomenal qualities, is a modern variant of the view which used to be known as naive, or direct realism (see, e.g., Dretske 1995, 1996; Tye 1996; Byrne and Hilbert 1997 Lycan 2001). For the externalist, phenomenal qualities are, so to speak, *what they are usually taken to represent*; that is, dispositional physical properties either of external objects, or of intentional objects whose existence we imagine or hallucinate. *Functionalism* is the view that mental attributes such as phenomenal qualities may identified solely with some functional set of their causes and effects; in short, they are *what they do* (see e.g., Rorty 1972; Fodor 1975, 1981, 1990; Lewis 1980; Rey 1980; Dennett 1991; Harman 1999). *Adverbialism*, on the other hand, holds that they are *how* they do what they do. On this view, phenomenal qualities should be understood as ‘ways of perceiving’ (See e.g., Ducasse 1952; Chisholm 1957; 1975; Kraut 1982; Tye 1984a, 1984b; Douglas 1998; Caruso 1999). When we have what we take to be a ‘phenomenally red’ experience, we are not observing a red ‘sense-datum’; on the contrary, we are merely ‘sensing *redly*’.

There is more than a hint here of a common desire to avoid the fact that phenomenal qualities, being *seemings* (as indicated by the word ‘phenomenal’) are, by definition, *how they seem*—that is, amongst other things, qualitative and hard-to-describe. Not all supporters of these views would be prepared to explicitly deny that this is so; however, in so far as they assert that

²⁰ See Sellars 1962, 1965, 1971. For a well-argued critique of Sellars’ anti-foundationalism, See Robinson (unpublished ms).

²¹ The expression ‘*this experience*’ substitutes here for an attended sensation, which I take to be *part of the content* of my thought concerning it.

²² On the minimal assumption, of course, that I understand the concept ‘phenomenal quality’ (as specified in my definition) and that I am able to apply it correctly.

qualitative experience may be straightforwardly identified with processes or properties which are scientifically well-understood and which, therefore, we *know* how to describe in formal, abstract terms—and in so far, furthermore, as these objective descriptions bear *no resemblance whatsoever* to the subjective qualities to which they supposedly apply—these theories may only be interpreted either as patently and trivially false, or as radically eliminative of phenomenal qualities as currently defined; which amounts, as I argued in §4.1, to the same thing.²³

According to Fred Dretske, for example, if we want to know what it is like, for a dogfish, to sense a particular kind of electric field, then all we need do is *describe the field*: “...there is no more to *experiencing* an electric field of type T than there is to *being* an electric field of type T since T is exactly what makes this electric field.” (Dretske 1995 p. 85; emphasis added). If this seems highly implausible even in cases of experiences we do not share, it is *a fortiori* empirically absurd in more familiar cases, such as that of color. The dispositional physical properties which cause an object to look blue under normal conditions are well-known and well-understood; yet their formal description provides no hint whatsoever as to the qualitative character of phenomenal blue.

Recall now our earlier discussion of the apparent *functionality* of phenomenal qualities—as signs, signals and switches. Functionalism, then, may be readily interpreted as identifying phenomenal qualities with these roles—and if an adverbialist theory (as is often the case) construes ‘redly’ in functional terms, for example ‘the manner of conscious perceiving which makes us able to report our perceptual discrimination of red objects’, then it does just the same.²⁴ However, the cybernetic function of signaling the detection of some physical stimulus (such as light of a certain wavelength) can be easily explained, yet such a formal description tells us nothing about the manifest character of that signal; it gives us no reason to expect that the signal will be subjectively qualitative and hard to describe—and no indication as to the particular kind of qualitative character which is, in fact, associated with this function.

In each case, to ignore the stark discrepancies to which I have drawn attention is, in effect, to deny the existence of phenomenal qualities outright.

The same may be said of a further implicitly eliminative ruse, which consists in the claim that phenomenal qualities may be identified with an *ability*—acquired upon attaining, for the first time, a particular sort of brain-state—to subsequently *recognize* its occurrence and also to ‘re-create’ or simulate it, introspectively; that is, through an act of imagination, *cause* a similar brain-state to occur (Lewis 1988; Nemirow 1990).²⁵ David Papineau follows Brian Loar (1990) in extending this idea to include the notion of a ‘phenomenal concept’; that is, the *idea* one acquires of having such an ability (Papineau 1993, 1995, 2002).

²³ All three of these theories can, in principle, be formulated in explicitly non-eliminative terms; in which case, they are vulnerable to a number of objections, to be considered in §5. All three, furthermore, are subject to a variety of difficulties which need not detain us here. See, for example, on externalism, Bailey 1998, 2003 and Robinson 1994, ch 5.; on functionalism (from various perspectives), Rorty 1972, Bealer 1978, 1985; Block 1978; Seager 1983; Schiffer 1986 and Pereboom 1991; on adverbialism, Jackson 1995, 1996, 1997.

²⁴ Non-eliminative adverbialists have been known to interpret the manner of a perception, such as ‘redly’, in explicitly phenomenal terms. As Ducasse explains: “‘blue,’ ‘bitter,’ ‘sweet,’ etc., are names not of objects of experience nor of species of objects of experience but of species of experience itself” (Ducasse 1942, 232). Now, it seems to me that this stipulative distinction is by no means necessary; for phenomenal ‘red’ might conceivably be understood as something which is both a species of experience and a kind of object thereof; that is, that the constituents of experience may also be considered, in some sense, its objects. Indeed, the phenomenal adverbialist’s strict separation of ‘species’ and objects of experience is open to a number of objections which I will not explore here; but for our present purposes, it will suffice to observe that this non-eliminative form of adverbialism is fully compatible with the present definition of phenomenal qualities.

²⁵ For more extensive criticism of the ability hypothesis and its variants, from various perspectives, see Alter 1998, 2000; Conee 1994, Gertler 1999; Lycan 1975, 1976.

The rhetorical advantage of this approach is that it allows its advocate to claim to be a ‘phenomenal realist’ or ‘inflationist’; that is, to assert that there is “something it is like” to have a phenomenal experience (that is, *like* having an ability or ability-concept), knowledge of which cannot be obtained by considering a third-person description of brain state, but only by actually *realizing* that state in one’s own case. Yet we should not be too easily deceived. This ‘ability’ account neither takes into account nor offers any explanation for the fact that phenomenal qualities are thoroughly hard to describe. Thus, in his recent book-length exposition, Papineau does not mention this problem at all; indeed, he *never once* uses the terms ‘phenomenal quality’, quale or qualia (Papineau 2002).²⁶ Instead, he focuses on the fact that contemplation of a third person, scientific description of a particular neural state does not elicit an occurrence of that state in his brain and hence, does not evoke the corresponding first-person experience. He then suggests that this *alone* explains what he calls ‘the illusion of distinctness’; that is, our fallacious ‘intuition’ that first person experiences have some special properties which their neural correlates do not possess.²⁷

For Papineau there is, accordingly, no mystery in the identification of specific phenomenal concepts with well-defined physical events. We already *know* (on the overwhelming strength of the arguments from correlation and causal closure) that phenomenal pain *just is* the firing of nociceptive-specific neurons in the parietal cortex; or something very similar—and “*genuine identities need no explaining*”:

If ‘two’ entities are one, then the one doesn’t ‘accompany’ or ‘give rise’ to the other—it *is* the other. And if this is so there is nothing to explain. It is possible to explain why one thing ‘accompanies’ or ‘gives rise’ to *another* thing. But you can’t explain why one thing is itself. (Papineau 2002 p. 144)

This is, of course, trivial—provided only that a genuine identity is established beyond doubt. It seems to me, however, that Papineau simply assumes this to be the case as an excuse for ignoring contradictory evidence. Had he acknowledged the fact that phenomenal qualities, unlike familiar physical phenomena (as conventionally conceived), have definite character which is thoroughly hard to describe, the illegitimacy of his argument would be abundantly clear. If an identity claim is made on the basis of circumstantial evidence, *but seems to be contradicted* by some evident inconsistency (that is, by the presence of contradictory evidence or by a conspicuous, unexplained absence of corroboration or both), then the genuineness of the identity is in doubt. It is the responsibility of the prosecution, so to speak, to account for the defendant’s *apparent alibi*—and explain how it is that two individuals which appear to have wholly contradictory properties can be the same. It simply won’t do to declare that on the basis of circumstantial evidence, the defendant *just is* the murderer—and then use the slogan “*genuine identities need no explaining*” as grounds for dismissing inconvenient facts which seem wholly *inconsistent* with this identity claim; such as, let’s say, that the victim was evidently kicked to death, whereas the defendant is a paraplegic.

Yet another example of a covert eliminative strategy would be provided by a theory which identifies the difficulty of describing phenomenal qualities with the difficulty of explaining

²⁶ The term ‘qualia’ appears in the titles of articles cited as references, but nowhere in the text of the book itself.

²⁷ This is, of course, false if we acknowledge the existence of phenomenal qualities, since their resistance to formal description provides a far better reason for distinguishing them from physical concepts; furthermore, it is implausible if we *deny* their existence, since this reason is then removed. Suppose, *per impossibile*, that there are no phenomenal qualities and that I have instead a ‘phenomenal concept’ of some sort of *functionally definable* discriminative state *F* such that consideration of the concept elicits an (attenuated) instance of *F*. I am then assured that *F* is identical with (or is instantiated by) a type of physical state *M*. Am I likely to find this *inconceivable*? Of course, contemplation of a third-person description of *M* does not elicit either an *F*, nor an *M*; likewise, when I think of an explosion, I (fortunately!) do not cause one. However, *F* and *M* have much in common; they are both formally describable, and indeed, at least in functional terms, their descriptions are essentially the same.

the *origin* of an intuition or hunch—whose introspectible content is non-qualitative and easily expressible as a proposition.

We may by all means acknowledge that, quite apart from the difficulty of describing their manifest content, phenomenal qualities are mysterious in another sense: we do not know, introspectively, how they arise within our minds. The experience of a phenomenal quality reveals nothing about the mechanism which gives rise to its occurrence: it *signals* a discrimination of some sort, but does not explain it. The same may be said, of course, of intuitions, hunches and other mental events in which we come to know (or believe) something, without knowing *how* we know. In recent years, cognitive scientists of the connectionist school have made much of the fact that computer simulated ‘neural networks’ seem to have this very property: for instance, the output of a pattern-recognition system which has been trained to recognize a particular sort of input may be highly reliable, even though it is impossible to extract from a system an account of the mechanism or algorithm it has used. It just *knows something*, so to speak, without knowing *how* it knows. It therefore seems plausible that the sensory discriminations and cognitive processes which give rise to phenomenal qualities, and also to intuitions or hunches, work in this sort of way.

It would be a mistake, however, to conflate *indescribability of content* with *ignorance of origin*—and deeply misleading suggest that the former may be readily explained in terms of the latter. The propositional content of an intuition (such as ‘someone is staring at me!’) may be perfectly easy to describe, however mysterious its origin. If our inability to introspect into the causal origins of mental states somehow gave rise to qualitative, hard-to-describe content, this would not be the case; *all* intuitions and hunches would be qualitative in character. Phenomenal qualities are not at all the same as intuitions or hunches—and to deny the distinction is, in effect, to deny that phenomenal *qualities*, as such, exist.

I am inclined to doubt that the foregoing survey covers all of the possible implicitly eliminative strategies to which the materialist might turn; however, I feel that I have sufficiently demonstrated, by example, how any such arguments may be diagnosed and refuted. In general, I think, on the basis of the arguments set out in §4.1, we may safely conclude that the radically eliminative approach—whether or not it be openly professed—provides no remotely plausible refuge from the implications of the description argument.

5. COUNTERARGUMENTS II: *AD HOC* SPECULATION

5.1 *The cryptic complexity hypothesis*

Those materialists who are prepared to acknowledge the absurdity of radically eliminative claims may take solace in the fact that even if we accept the existence of phenomenal qualities, the description argument does not categorically *exclude* the possibility that materialism may still be true. Its conclusion could, perhaps, be avoided by means of a sophisticated hypothesis which clearly *explained* how it is that phenomenal qualities, while wholly physical, are so thoroughly resistant to description in physical terms. In the absence, however, of a full explanation based on the discovery and validation of formal descriptions of phenomenal qualities, the best that the materialist can manage will necessarily be limited to something like the following:

Phenomenal qualities (*qua* physical phenomena) are remarkably complex and subtle in nature; sufficiently so, indeed, to possess certain deeply mysterious and highly unusual (yet still wholly physical) properties—which endow them, *somehow*, with their familiar subjective characteristics whilst also, *somehow*, thoroughly obscuring from us all trace of their fundamental structure and dynamics.

Let us call this suggestion the Cryptic Complexity Hypothesis (or in short, ‘CRYCH’). Some contemporary materialists seem to be remarkably confident not only that it is, quite obviously, true, but also that the various impenetrable mysteries embedded within it may

eventually be explained, reductively, in terms of the micro-physical facts. Others—the ‘non-reductive’ or ‘mysterian’ materialists—either deny, or are inclined to doubt, that this will ever be possible. Either way, if we accept the premises of the description argument yet maintain that materialism is, none the less true, then something akin to CRYCH must also be true. As a scientific proposition, however, by all conventional standards, it is patently unacceptable—since it not only posits an unspecified extent of utterly unknown and mysterious complexity, but also predicts, without clearly explaining, a complete absence of supporting evidence. Yet without such empirical validation, the only satisfactory grounds we could have for taking CRYCH seriously would be some independent argument to show that materialism is *necessarily* true—and this, of course, we do not have. To give credit where credit is due, there are at least *some* materialist philosophers who seem to be acutely aware of this problem. George Rey, for example:

“Of course,” declares the thoroughly modern materialist, “we are machines but”—and here he waves his hands towards properties as fabulously mysterious as any conceived by the dualist—“we are very, very complex ones.” I call this view “facile materialism.” (Rey 1997 p. 462)²⁸

In other words, in order to defeat the description argument, the materialist may not merely assume the veracity of CRYCH, but must also defend its plausibility.

5.2 Argumentum ad ignorantiam?

There are some, however, who would deny that this is strictly necessary. For all we know, CRYCH *might* be true—and this mere possibility allows us to accuse the description argument of committing the classic Aristotelian informal fallacy, ‘*argumentum ad ignorantiam*’ (See e.g. Churchland, P. S. 1996, 1998). On this view, we have no right or reason to infer, from our lack of insight into the nature of phenomenal qualities or from our failure to imagine their identity with neurological processes, that they are not, in fact, relation physical properties at all. Our knowledge and understanding of the functioning of the brain, in all its astonishing complexity, is still rudimentary; there are surely many discoveries to be made that will surprise and astonish us. Perhaps CRYCH will be verified by such a discovery. The fact that such revelation has yet to occur does not prove that it never will.

Now, it is certainly true that in some circumstances, arguments based upon the absence of a particular sort of evidence are unsound. Consider, for example:

- (1) We have no hard evidence for the existence of intelligent life outside this solar system; therefore, there is probably no such life outside this solar system.

The conclusion, in this instance, seems premature and altogether unjustified. It is not the case, however, that all arguments of this general form are lacking in empirical force. Consider:

- (2) We have no hard evidence for the existence of intelligent life on Mars; therefore, there is probably no such life on Mars.

Given the present extent of our knowledge of the red planet, the fact that we are not omniscient on this subject does not seem to make this argument fallacious or unreasonable. What, then, distinguishes a bad argument, of this burden-of-proof type, from a good one?

A trivial answer would be provided by a case in which, contrary to the argument’s premise, evidence of the relevant sort is actually available. Let us assume, however, that for both of the above examples (as in the case of the description argument), the ‘absence of

²⁸ Rey, alas, prefers a radically eliminative approach—which, I am inclined to call the *other* sort of facile materialism.

evidence' premise is correct. In this situation, the difference lies, I suggest, in what might be called the '*knowledge context*'—that is, the nature and extent of our knowledge of the domain in which evidence has failed to emerge—and in how this knowledge bears upon the plausibility of the claim that, *were the contested proposition true*, the corroborative evidence in question would almost inevitably be available to us. In the case of argument (1), this is far from plausible, since we have scarcely begun to search for such evidence—and the territory that remains to be explored, in anything approaching sufficient detail, is vast. In example (2), however, the domain in question has already been quite extensively studied, explored and investigated—and the consensus of informed scientific opinion seems to be that, were there any intelligent life on Mars, certain kinds of evidence would almost certainly have become apparent, by now. These considerations suggest a general model for an empirically valid argument from absence of evidence:

- (a). If p, a certain type (or set of types) of evidence E would (probably) be apparent, given sufficient search
- (b). There has been sufficient search for E
- (c). E is not apparent
- (d). Therefore, p is (probably) false.

The description argument can, in part, be interpreted as an instance of this model—but this alone does not, of course, render it's conclusion secure; even if we accept that in this case, (c) is true. To begin with, if we had reasonable grounds for rejecting (a), then the argument would fail.

We are not, however, entitled to dismiss (a) by merely assuming the veracity of an *ad hoc* hypothesis. It would be insufficient, for example, for a believer in the existence of intelligent Martians to respond to (2) with the suggestion that:

- (3) The Martians—motivated by their fiendish plans for invasion—are able, somehow, to telepathically manipulate our Earthling astronomers—forcing them to suppress all the relevant evidence.

The absence of evidence argument summarized in (2) does not altogether eliminate the possibility that (3) is true. However, this sort of hypothesis—unsupported by further evidence of argument—presents no credible threat to the anti-Martian conclusion. It introduces a vast extravagance both of unknown complexity and unsolved mystery (such as how, exactly, does Martian telepathic manipulation work?); whilst remaining, for all practical purposes, immune to refutation. Without additional evidence, the only grounds we could have for taking (3) seriously would be an *a priori* assumption that intelligent Martians exist. Hence, any attempt to defend this position simply by appealing to the possibility of (3) would be begging the question. As we saw in §5.1, CRYCH is also a conspicuously vague and extravagant *ad hoc* hypothesis—indeed, it seems *at least* as suspect, in this respect, as (3)—so an accusation of *argumentum ad ignorantiam* based solely upon CRYCH is itself an instance of *petitio principii*—and the accusation cannot stand.

The materialist's only remaining option, then, is to challenge (b), and suggest that there has been *insufficient search* for some special sort of evidence which might vindicate CRYCH—or indeed, provide (at least) some rudimentary, yet empirically valid, formal descriptions of phenomenal qualities. It seems to me, however, that this tactic runs the risk of reducing a charge of *argumentum ad ignorantiam* to an instance of that very mistake; for example:

- (4) There is no conclusive proof that there is no intelligent life on Mars; and we have yet to look under every rock. Furthermore, we cannot deny the possibility that the malevolent telepathic-censorship hypothesis is true. Therefore, there probably really is intelligent life on Mars!

Or alternatively:

- (5) Our knowledge and understanding of the neurological activity of the brain is incomplete, and thus, we have no conclusive proof that CRYCH is false. Therefore, CRYCH (and hence, materialism) is probably true.

As it stands, this simply will not do. This point is particularly clear when we have an additional argument against the *ad hoc* hypothesis, based upon inductions from evidence we already abundantly possess. The general form of such an argument can be set out as follows:

- (e). If p, there is an X with an anomalous property Y
- (f). To the best of our theoretical knowledge and understanding of the possible properties of X, all X *necessarily* lack property Y
- (g). As far as we can tell, all known X *lack* property Y
- (h). Therefore it is (at least) highly *probable* that all X lack property Y
- (i). Therefore, p is (probably) false.

In the case of the description argument, as summarised at the end of §3, (f) and (g) correspond to proposition (3) and (h) to proposition (4); thus, X is any formally describable, physically instantiated structure, process or pattern and Y is the property of instantiating—some particular perspective or formal transformation, a phenomenon of distinctive and vivid character, which altogether *lacks* all discernible trace of structure or dynamics, geometry or form.

In order to sustain a charge of insufficient search, the materialist must not only identify a domain (such as some aspect of the functioning of the brain) of which our knowledge is profoundly limited, but must also provide, in reply to *both* strands of the description argument, some plausible grounds for supposing that the decisive evidence we seek (that is, peculiar, descriptively opaque complexity, plausibly identifiable with phenomenal qualities) is likely to be found there.

5.3 A critique of cryptic complexity

A number of philosophers and cognitive scientists have drawn attention to physically complex phenomena which might, *prima facie*, be considered relevant to a defense of CRYCH.²⁹ Daniel Dennett, for example, has argued that the brain's reportable discriminative states are far more complex than we think; they do not correspond in some neat and simple way to fundamental physical features of the environment but instead to complex features which, in evolutionary terms, it is advantageous for us to detect—and they are also further complicated by a host of 'reactive dispositions', which are difficult, perhaps impossible, to dissociate from the basic information state itself (Dennett 1991, ch. 12). The requisite complexity, in other words, is out there (or perhaps, in here). Be that as it may, any theory which seeks to *identify* such candidate complexities with phenomenal qualities will need to overcome a series of formidable objections. Let us consider these each in turn.

- (1) *insufficient corroboration*: The complexities of the candidate domain shows no promise of yielding, under any interpretation or analysis, even a *rudimentary* formal description of the phenomenal qualities with which they are supposedly identified.

Perhaps, to be charitable, we should let this one pass—as this is the very objection which CRYCH is designed to avoid. Maybe we can allow that, even without such corroboration, a

²⁹

I shall exclude from consideration here those discussions which appeal to the complexity of what I have called 'qualitative manifolds', that is, complex, multifaceted episodes or experience which include a host of structural elements, in addition to then many different qualitative components. These are not phenomenal qualities, as defined in §2.1

tentative degree of support for CRYCH could be provided by the discovery of some physical phenomenon whose complex properties are profoundly *unusual* or unprecedented in some mysterious way; such that we may, at least, *speculate* that such a phenomenon could somehow be endowed with the peculiar properties which CRYCH requires. Failing even that, the materialist needs to demonstrate, at the very least, the existence of a physical domain which is largely unexplored and deeply mysterious—when such strange monsters might conceivably lurk. Otherwise, we are faced with:

- (2) *insufficient mystery*: the candidate domain has been extensively studied and explored; and while there may yet be plentiful fine detail to be discovered, there is no evidence to suggest that its complex properties are profoundly *unusual*, in the way required by CRYCH; or that some altogether unexplored, ‘virgin territory’ within that domain might reveal such bizarre and uncanny complexity.

Now it seems to me that, given the current extent of our relevant scientific knowledge, this objection applies, with compelling force, to any domain the materialist might care to name—from ‘upstream’ phenomena (external physical stimuli or neural processes) which give rise the occurrence of phenomenal qualities, to their proximal neural correlates (such as *neural* signs and signals associated therewith), or even ‘downstream’ consequences (that is, neural, physiological or behavioral reactions and side-effects of their occurrence). We could, therefore, I think, quite legitimately stop here—and declare CRYCH (and hence, materialism) empirically indefensible. Let us imagine, however, for the sake of argument, that some ingenious sort of CRYCH-based theory could somehow survive this difficulty. The materialist must then proceed to *identify* the peculiarly complex properties of the candidate domain with phenomenal qualities—and this requirement gives rise to what I will call the *functional trilemma*. If the chosen domain is (so to speak) ‘upstream’ of the proximal neural correlates of phenomenal qualities, then the first horn presents itself:

- (3) *irrelevance*: in the case of subjectively *simple* phenomenal qualities, there is no reason why the complexity of ‘upstream’ *causes* of the associated discriminative state should have any bearing upon the complexity of the sign or signal which flags its occurrence.

Recall that phenomenal qualities have no discernible *structural* information content, while subjectively *simple* phenomenal qualities, moreover, lack even *qualitative* complexity; that is, we are unable to detect within them a *complex mixture* of simpler qualitative components.³⁰ Such primitive impressions do not, therefore, seem to embody any highly complex information.

This makes good functional sense, because the characteristic appearance of a signal S which announces, *in simple terms*, the occurrence (and perhaps, the quantitative level) of a stimulus event E is altogether functionally independent of the complexity of E and the means by which E is detected—and indeed, of any further consequences which E or the activation of S might provoke. Thus, such upstream causes and downstream effects, therefore, have no necessary influence upon the character of *simple* phenomenal qualities.³¹

The materialist’s first reply to this objection might well be, *so what?* The fact that there is no strictly *necessary* connection here does not exclude the possibility that, for some unknown reason, the individuating character of particular neural signs and signals might be a complex function of their complex external or internal causes; alternatively, if there is no such connection, these neural correlates of phenomenal qualities might be complex for some other

³⁰ See my comments on subjectively simple phenomenal qualities in §2.1, n. 8.

³¹ In the case of qualitatively complex sensations, (such as the taste of Great Aunt Hildegard’s gumbo source) the *number* and *relative proportions* of the qualitative components we are able to discern may correspond, roughly, to the physical composition of the stimulus; however, a description of the latter tells us nothing about the character of these components.

reason. Either way, the mysterious complexity which instantiates subjectively simple phenomenal qualities might consist, precisely, in such neural signs and signals. This response, however, serves only to impale the materialist upon the second horn of our trilemma:

- (4) *Functional redundancy*: The functional roles served by simple phenomenal qualities do not *require* great complexity—and *functionally redundant* complexity could not have arisen through natural selection by virtue of adaptive merit.

The functional requirements which are satisfied by subjectively simple phenomenal signs are, as far as we can tell, quite straightforward. Each such sign must possess a particular, unique character which reliably and consistently distinguishes it from others; it may also be required to manifest a magnitude or level of intensity. That is all. There is, therefore, no evident functional *need* for these information states to be characterized by impenetrably complex, subjectively ineffable tokens.³² The materialist may still speculate that phenomenal signs are, nevertheless, *for some unknown reason*, highly complex. The difficulty is that such excessive, *functionally redundant* complexity would serve no adaptive purpose; it would not be naturally selected for in the evolution of the brain—and it therefore seems all the more unlikely that it exists at all.

The materialist next move may be to avoid this objection by proposing that phenomenal qualities have arisen as an unadaptive *side-effect* of the evolution of cognitive activity in the brain. Let us call this suggestion the Cryptic *Unadaptive* Complexity Hypothesis (CRUNCH). It is true, indeed, that natural selection frequently gives rise to biological features which have not, apparently, been selected for in themselves—but which seem to have arisen accidentally, as a consequence of the development of other, functionally valuable adaptations. In such cases, however, evolutionists are generally able to identify the primary adaptations in question and plausibly explain how the ‘side-effects’ came about.³³ Without such a theoretical defense, CRUNCH is no more than an *ad hoc* elaboration of CRYCH and is therefore, *a fortiori*, empirically untenable. Worse still, it is also vulnerable to the third horn of our trilemma:

- (5) *Inefficiency*: functionally redundant, excess complexity within signs and signals would constitute a profligate waste of bandwidth—and would therefore be expected to present a considerable evolutionary disadvantage.

Cognitive signs, signals and switches—which is what phenomenal qualities seem to be—surely need to function quickly and efficiently. Any greater complexity than the minimum required, in accordance with information theory, to reliably convey the relevant information would be a waste of bandwidth, reducing efficiency and compromising the performance of the system as a whole. Thus, the hypothetical, mysteriously complex physical phenomena which supposedly instantiate phenomenal qualities seem not only entirely unnecessary for the performance of ordinary, practical, cognitive functions—but inefficient and counterproductive.

It might be suggested, of course, that this mysterious complexity actually does play some yet-to-be-discovered, distinctively useful functional role in cognitive activity; somehow—in spite of the disadvantage of greater bandwidth-consumption—making the system work better than it would with a simple, formally-describable representational language. This mysterious contribution to efficient information processing could therefore, in principle, have evolved by natural selection.

³² This point should not be confused with a traditional absent qualia conceivability argument, in which it is standardly assumed that the physical facts remain unchanged. Here, on the contrary, we note that particular functional roles could be served by different *physical* phenomena: either (a) by simple neural signs and signals, devoid of qualitative content or (b) more elaborately, by the hypothetical weirdly complex neural phenomena proposed by CRYCH. One could, of course, base a plausible absent qualia argument (taking care, that is, to avoid the usual, epiphenomenalist approach) on the observation that phenomenal qualities *as such* are functionally redundant; but for our present purposes, this would only serve to refute eliminative functionalism, which we have already excluded on other grounds. With a different target in mind—namely, Strong AI—I adopt this approach in Clifton (2004) [b], [c] and [d].

³³ See, e.g. Gould and Lewontin 1979.

Let us call this new proposal the Cryptic Utilitarian Complexity Hypothesis (CRUTCH). Clearly, just like CRUNCH, it compounds the *ad hoc* sins of CRYCH with still further, extravagantly vague and obscure embellishments—and it is altogether unsupported by positive evidence. There is, however, an increasingly compelling body of *contradictory* evidence. We have successfully developed sophisticated computer systems that are capable of discriminating colors, sounds and even odors with remarkable efficiency; yet their basic discrimination identifiers—that is, those internal signs and signals which indicate, for instance, the detection of particular colors—are perfectly simple. They are binary numbers, stored at particular locations in the computer’s memory. We have no reason to suppose that the computer ‘experiences’ them in a qualitative way, nor any reason to think that, if it did so, this would somehow help it do its job.

Lacking, so to speak, the support of a viable CRUTCH, the materialist must turn to the one remaining means of avoiding the functional trilemma, and look *downstream* to the side-effects and consequences of the discriminations associated with phenomenal qualities. A prominent exponent of this strategy is Daniel Dennett, who argues that those features of conscious experience which we are inclined to interpret as qualitative are not just simple neural signals, but highly complex, multi-dimensional states, comprised of the activation of various ‘reactive dispositions’—emotional responses, for example—which are generally difficult, perhaps impossible, to dissociate introspectively from the basic discriminatory state itself. ‘Phenomenal qualities’, on this view, are wholly constituted by some combination of the various reactive dispositions associated with a particular sensory discrimination (Dennett 1991, ch. 12). In other words, a phenomenal quality is, so to speak, no more than the sum of its side-effects. Now this suggestion might indeed succeed in avoiding the functional trilemma—but only, alas, to confront a further objection of its own:

- (6) *Dennett’s Dilemma*: the identification of phenomenal qualities with a complex mélange of ‘reactive dispositions’ is either implicitly eliminative or appeals to an infinite regress of qualitative complexity.

If a ‘reactive disposition’ is merely conceived as a tendency to behave in a certain readily definable way, then there is no reason to suppose that our awareness of being in such a state should be qualitative, that is, *thoroughly* hard-to-describe. The suggestion that the content a phenomenal quality is merely a mixture of behavioral propensities is implicitly eliminative: were it true, we could give a formal description of a phenomenal quality in the form: $RD_1 + RD_2 + RD_3 \dots$ where RD_n is a formally describable reactive disposition. Furthermore, many reactions of the kind Dennett mentions, such as arousal, anxiety and fight-or-flight responses etc., are characterized by phenomenal qualities of their own. Think of the mild impression of ‘warmth’, excitement or arousal which we tend to associate with the color red. Such emotional concomitants of color sensations each have their own distinctive qualitative character—which does not, in itself, (by common report) feel remotely like a hue. They recognizably occur in association with non-visual experiences and they defy formal description just as colors do. Hence it seems highly unlikely that a combination of such reactions literally *constitutes* the phenomenal character of a color sensation.³⁴ Suppose, however, for the sake of argument, that we assume a non-eliminative interpretation of Dennett’s to be generally true. In this case each ‘reactive disposition’ which serves as a qualitative ingredient within a particular, subjectively

³⁴ This point, I think, is reinforced by functional considerations. Suppose that, for evolutionary reasons, a particular smell and a particular color happen to be associated with a similar set of reactive dispositions. It seems highly probable that we would expect there to be many instances of this sort of convergence; indeed, this may partly explain why synesthetes find a sensation of one kind involuntarily evoke another (a classic example being, the color red and the sound of trumpets). It still remains functionally important to tell them apart. Thus, while synesthetes, may find associated sensations to be *similar*, are just as able as the rest of us able to tell the difference between them; a hue is still a hue and a timbre is still a timbre (See Cytowic 1989, 1993, 1995). On Dennett’s theory, we should *all* be in considerable danger of confusing any sensations whose reactive dispositions overlap.

simple phenomenal quality is *itself* a phenomenal quality; which in turn must be a complex, qualitative mixture of further reactive dispositions—and so *ad infinitum*. Thus, the character of a subjectively simple phenomenal quality, such as phenomenal yellow, must be determined by an *infinitely* complex neural state—which is surely impossible.

Once again, it seems to me that we could legitimately stop here, on the grounds that the foregoing objections are more than sufficient to show that CRYCH, in general, is empirically indefensible. Let us, however, temporarily imagine, for the sake of argument, that some yet-to-be-formulated CRYCH-based theory might somehow, *per impossibile*, overcome *all* of the foregoing difficulties. There remains one further objection which any such proposal will still need to overcome:

- (7) *insufficient ineffability*: there is no indication as to why the candidate complexity should be “cryptic” (that is, thoroughly hard-to-describe) from a first-person perspective. The mere *intricacy* of an observed or experienced phenomenon does not, of itself, account for our inability to come up with a *rudimentary* formal description, that is, a rough-but-recognizable, primitive-but-promising one.

An example, due to Dennett, may help to clarify this point. The raggedly torn edge of a strip of paper on my desk has a complexity that I cannot describe in precise detail, but I could, I think, describe a line that *approximately* matches its shape—by means of which, with luck, someone else could pick out my particular strip of paper from a pile of others which looked quite different. Thus (contrary to Dennett’s interpretation of this case; See Dennett 1991 pp. 376, 382-3), our impressions of such ordinary, everyday, intricacies are by no means ‘ineffable’.

In order to improve upon CRYCH, it is insufficient to point to some sort of structural and/or dynamic complexity and demonstrate that this phenomenon correlates closely with our awareness of phenomenal qualities. The materialist must also explain why, from the point of view of first-person experience, this physical phenomenon seems simple, homogenous and altogether devoid of structure—and yet, is somehow endowed with a particular phenomenal character.

In the light of these objections, it should now be clear that CRYCH, and its even more elaborate variants such as CRUNCH and CRUTCH, are not only illegitimate by virtue of their vague, mysterious and *ad hoc* empirical form; they are also, on various independent grounds, almost certainly false. In other words, a careful consideration of the viability of CRYCH, far from casting doubt upon the description argument, serves only to add weight to its conclusion.

6. DESIDERATA FOR A THEORY OF MIND

The empirical arguments I have presented in §§3-5 do not, of course, constitute rigorous *proofs* that materialism is false. Nevertheless, I believe they provide sufficient grounds for treating this position as exceptionally dubious.

This conclusion gives rise to a profound philosophical dilemma. On the one hand, it now seems that materialism is immensely implausible—but on the other hand, it is usually assumed that the same is true of mentalism. The least attractive option, in these circumstances, would be to throw in the towel and admit to hopeless bafflement—but if that is truly all that we can do, we must be honest about our uncertainty; we have no justification for maintaining that either position, materialism or mentalism, is more likely to be true than the other.³⁵

We may, I suggest, begin to consider materialism position credible if, and only if, we are presented with a valid—but permissibly, very rudimentary—formal description of at least one phenomenal quality. Further support for materialism would also be provided by a plausible

³⁵ Likewise, there is nothing to be gained by disguising our bafflement under the cloak of a radical new theory which we claim is neither materialist nor mentalist—but which in fact either falls entirely under one or other definition, or inconsistently combines elements of the two.

defense of some version of the cryptic complexity hypothesis—preferably, accompanied by positive supporting evidence.

A mentalist theory, on the other hand, must also overcome a number of formidable hurdles if we wish it to be taken more seriously. It must be scientific in spirit and as parsimonious as possible; the fewer fundamental entities and psychophysical laws such a theory proposes, the better. It must avoid the trap of eternally separating mental subjects from objects and thereby implying an infinite regress of inner observers. It must propose a coherent, theoretical account of the causal relationship, if any, between non-physical mental phenomena and physical events in our brains and bodies—consistently on the one hand, with evidence suggesting the reality of mental causation and on the other hand, with evidence which is held to suggest that the physical world is causally closed. A truly respectable mentalist theory should give rise to predictions that are testable, at least in principle. Ideally, it should also offer some account of mental phenomena other than phenomenal qualities, such as intentionality, the unity of consciousness, personal identity and mental causation.

For a defense of a mentalist theory of consciousness which sets out to satisfy these criteria, see Clifton (2004) [a].

References

- Alter, T. (1998) 'A Limited Defense of the Knowledge Argument.' *Philosophical Studies*, 90: 35-56
- Alter, T. (2001) 'Know-how, Ability, and the Ability Hypothesis', *Theoria* 67: 229-239.
- Ayer, A. (1968) 'Can there be a private language?' *Concept of a Person and Other Essays*, London: Macmillan.
- Bailey, A. (1998) *Phenomenal Properties: The Epistemology and Metaphysics of Qualia*. Doctoral dissertation (unpublished), Department of Philosophy, University of Calgary, Canada
- Bailey, A. (2003) *Qualia and the argument from illusion*. Unpublished MS, Department of Philosophy, University of Guelph, Canada [Email: abailey@uoguelph.ca]
- Bartoshuk, L. (1978) 'Gustatory system.' In: Masterson, R. B. (ed.) *Handbook of Behavioral Neurobiology*, Vol 1, *Sensory Integration*, New York: Plenum Press.
- Bealer, G. (1978) An inconsistency in functionalism. *Synthese* 38: 333-372
- Bealer, G. (1985) Mind and anti-mind: Why thinking has no functional definition. *Midwest Studies in Philosophy* 9: 283-328
- Bindra, D. (1969) A unified interpretation of emotion and motivation. *Annals of the New York Academy of Sciences* 159: 1071-1083
- Block, N. & Fodor, J. (1972) 'What psychological states are not.' *Philosophical Review* 81: 159-81
- Block, N. (1978) Troubles with functionalism. *Minnesota Studies in the Philosophy of Science* 9: 261-325
- Block, N. (1980) 'Are absent qualia impossible?' *Philosophical Review* 89: 257-74
- Block, N. (1990) 'Inverted earth.' *Philosophical Perspectives* 4: 53-79
- Block, N. (1995) 'On a Confusion about the Function of Consciousness.' *Behavioral and Brain Sciences*, 18: 227-247
- Block, N. (2003) 'Qualia' In Gregory, R (ed.) *Oxford Companion to the Mind*, New York: Oxford University Press (revised edition, forthcoming)
- Bozarth, M. (1994) 'Pleasure systems in the brain.' In D.M. Warburton (ed.), *Pleasure: The politics and the reality*. New York: John Wiley & Sons.
- Brandom, R. (1994) *Making it Explicit. Reasoning, Representing and Discursive Commitment*. Cambridge, MA: Harvard University Press.
- Byrne, A & Hilbert D. (1997) 'Colors and Reflectances.' In Byrne, A & Hilbert D. (ed.), *Readings on Color, Volume 1: the philosophy of color*, Cambridge, MA: MIT Press.
- Caruso, G. (1999) 'A Defense of the Adverbial Theory.' *Philosophical Writings*, 10: 51-65
- Chalmers, D. (1996) *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Chisholm, R. (1957) *Perceiving: A Philosophical Study*. Ithaca, NY: Cornell University Press.
- Churchland, P. M. (1996) 'The rediscovery of light.' *Journal of Philosophy* 93: 211-28
- Churchland, P. S. (1983) 'Consciousness: The transmutation of a concept.' *Pacific Philosophical Quarterly* 64: 80-95.
- Churchland, P. S. (1996) 'The hornswoggle problem.' *Journal of Consciousness Studies* 3: 402-8
- Churchland, P. S. (1996) 'Brainsy: Nonneural Theories of Conscious Experience' In: S. Hameroff, A. Kaszniak, A. Scott (eds.) *Toward a Science of Consciousness II: The 1996 Tucson Discussions and Debates* Cambridge MA: MIT Press.
- Clark, A. (1993) *Sensory Qualities*. New York: Oxford University Press.
- Clark, A. (2000) *A theory of sentience*. New York: Oxford University Press.
- Clifton, A. (2004) [a] 'Res cogitans.' Unpublished MS.

- Clifton, A. (2004) [b] 'The introspection game – or, does the Tin Man have a heart.' Unpublished MS.
- Clifton, A. (2004) [c] 'Blind man's bluff and the Turing test.' Unpublished MS.
- Clifton, A. (2004) [d] 'The hazards of silicon heaven.' Unpublished MS.
- Clifton, A. (2004) [e] 'The trouble with zombies.' Unpublished MS.
- Conee, E. (1994) 'Phenomenal Knowledge.' *Australasian Journal of Philosophy* 72: 136-150
- Cytowic, R. (1989) *Synesthesia: A Union of the Senses*. New York: Springer Verlag.
- Cytowic, R. (1993) *The Man Who Tasted Shapes*. New York: Putnam.
- Cytowic, R. (1995) 'Synesthesia: phenomenology and neuropsychology – a review of current knowledge.' *Psyche* 2:10 [psyche.cs.monash.edu.au/v2/psyche-2-10-cytowic.html]
- Dennett, D. (1988) 'Quining Qualia.' In A Marcel and E. Bisiach, eds., *Consciousness in Contemporary Science*. Oxford: Oxford University Press
- Dennett, D. (1991) *Consciousness Explained*. New York: Penguin
- Dennett, D. (1996) 'Facing Backwards on the Problem of Consciousness.' *Journal of Consciousness Studies*, 3 (1): 4-6
- Douglas, G. (1998) 'Why Pains Are Not Mental Objects.' *Philosophical Studies*, 91(2): 127-148
- Dretske, F. (1995) *Naturalizing the Mind*. Cambridge, MA: MIT Press.
- Dretske, F. (1996) 'Phenomenal Externalism, or If Meanings Ain't in the Head, Where Are Qualia?' In Villanueva, E., ed. (1996). *Perception*. Atascadero, CA: Ridgeview.
- Ducasse, C. (1942) 'Moore's "The Refutation of Idealism."' In Schilpp P.A., ed. (1968) *The Philosophy of G.E. Moore*. Third edition. La Salle, IL: Open Court.
- Edelman, G. & Tononi, G (2000) *Consciousness: How Matter Becomes Imagination*. London: Allen Lane.
- Fodor, J. (1975) *The Language of Thought*. New York: Crowell.
- Fodor, J. (1981) 'The Mind Body Problem.' *Scientific American* 244 (1): 114-123
- Fodor, J. (1990) *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.
- Gertler, B. (1999) 'A Defense of the Knowledge Argument.' *Philosophical Studies*, 93: 317-36
- Gescheider, G. (1997) *Psychophysics: The Fundamentals*. New York: Erlbaum.
- Gould, S. and Lewontin, R. (1979) 'The spandrels of San Marco and the panglossian paradigm: a critique of the adaptationist programme.' *Proceedings of the Royal Society of London Series B.*, 205: 581-598
- Hardcastle, V. (1999) *The Myth of Pain*. Cambridge, MA: MIT Press.
- Hardin, C. (1993) *Color for Philosophers*, Indianapolis, IA.: Hackett.
- Harman, G. (1990) 'The Intrinsic Quality of Experience.' in Tomberlin, J. ed., *Action Theory and Philosophy of Mind (Philosophical Perspectives, Vol. 4)*. Atascadero, CA: Ridgeview Publishing.
- Harman, G (1999) 'Wide Functionalism' in: *Reasoning, Meaning, and Mind*. Clarendon Press.
- Jackson, F. (1975) 'On the Adverbial Analysis of Visual Experience.' *Metaphilosophy* 6: 127-135.
- Jackson, F. (1976) 'The Existence of Mental Objects.' In Rosenthal, D.M., ed., (1991) *The Nature of Mind*. New York: Oxford University Press.
- Jackson, F. (1977) *Perception: A Representative Theory*. Cambridge: Cambridge University Press.
- Jackson, F. (1982) 'Epiphenomenal qualia.' *Philosophical Quarterly* 32: 127-136
- James, W. (1890) *The Principles of Psychology*. New York: John Holt.
- Johnsen, B. (1993) 'The intelligibility of spectrum inversion.' *Canadian Journal of Philosophy* 23: 631-6
- Kirk, R. (1974) 'Zombies vs. materialists.' *Aristotelian Society Supplement* 48: 135-52
- Kraut, R. (1982) 'Sensory States and Sensory Objects' *Nous*, 16(2): 277-293
- Land, E. (1977) 'The Retinex theory of color vision.' *Scientific American*, 237: 108-128
- Land, E. (1986) 'Recent advances in Retinex theory.' *Vision Research* 26: 7-21.
- Larmer, R. (1986) 'Mind-body interaction and the conservation of energy.' *International Philosophical Quarterly*, 103: 277-285.
- Lemos, N. (1995) *Intrinsic Value: Concept and Warrant*. Cambridge University Press
- Levine, J. (1991) 'Cool red.' *Philosophical Psychology* 4: 27-40.
- Lewis, D. (1980) 'Mad Pain and Martian Pain.' in N. Block, ed., *Readings in the Philosophy of Psychology*, Vol. I. Cambridge, MA: Harvard University Press
- Loar, B. (1990) 'Phenomenal States (Revised Version)' In Block, N, Flanagan, O. and Güzeldere, G (eds.) *The Nature of Consciousness*. Cambridge, MA: MIT Press.
- Lycan, W. (1973) 'Inverted spectrum.' *Ratio* 15: 315-9.
- Lycan, W. (1995) 'A Limited Defense of Phenomenal Information', in Metzinger, T. ed., *Conscious Experience*. Tucson: University of Arizona Press.
- Lycan, W. (1996) *Consciousness and Experience*. Cambridge, MA: The MIT Press.
- Lycan, W. (2001) 'The case for phenomenal externalism.' *Philosophical Perspectives* 15: 17-35
- McDowell, J. (1994) *Mind and World*. Cambridge, MA: Harvard University Press.
- Moore, G. (1903/1993) *Principia Ethica*. Revised edition: Cambridge University Press
- Nagel, T. (1974) 'What is it like to be a bat?' *Philosophical Review*, 79: 394-403

- Nemirov, L. (1990) 'Physicalism and the Cognitive Role of Acquaintance.' in Lycan, W. G. ed., *Mind and Cognition: A Reader*. Cambridge, MIT: Blackwell.
- Papineau, D. (1993) 'Physicalism, Consciousness and the Antipathetic Fallacy.' *Australasian Journal of Philosophy*, 71: 169-183
- Papineau, D. (1995) 'The Antipathetic Fallacy and the Boundaries of Consciousness', in Metzinger, T., (ed.) *Conscious Experience*. Paderborn: Schöningh.
- Papineau, D. (2002) *Thinking about consciousness*. Oxford University Press.
- Pereboom, D. (1991) 'Why a scientific realist cannot be a functionalist.' *Synthese* 88:341-58.
- Rey, G. (1996) 'Towards a Projectivist Account of Conscious Experience.' in T. Metzinger, ed., *Conscious Experience*. Paderborn: Schöningh.
- Rey, G. (1997) 'A question about consciousness.' In Block, N, Flanagan, O. and Güzeldere, G (eds.) *The Nature of Consciousness*. Cambridge, MA, MIT Press.
- Rey, G. (1980) 'Functionalism and the Emotions.' in A. Rorty, ed., *Explaining Emotions* Berkeley: Univ. of California Press.
- Robinson, H. (1976) 'The mind-body problem in contemporary philosophy.' *Zygon* 11: 346-360
- Robinson, H. (1994) *Perception*. Routledge.
- Robinson, H. (2003) 'Empiricism and the philosophy of mind: has Sellars refuted "the given"?' Unpublished MS, Department of Philosophy, Central European University, Budapest. [Email: robinson@ceu.hu]
- Rorty, R. (1970) 'In defense of eliminative materialism.' *Review of Metaphysics* 24: 112-121.
- Rorty, R. (1972) 'Functionalism, machines and incorrigibility.' *Journal of Philosophy* 69: 203-220
- Rorty, R. (1979) *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press.
- Rorty, R. (1997) Introduction to Sellars, W (1956/1997) *Empiricism and the philosophy of mind*. Cambridge, MA: Harvard University Press.
- Schiffer, S. (1986) 'Functionalism and belief.' In Brand, M. & Harnish, R., eds. *The Representation of Knowledge and Belief*. University of Arizona Press.
- Seager, W. (1983) 'Functionalism, qualia and causation.' *Mind* 92: 174-88
- Sellars, W. (1956/1997) *Empiricism and the philosophy of mind*. Cambridge, MA: Harvard University Press.
- Sellars, W. (1962) 'Philosophy and the Scientific Image of Man.' In R. Colodny, ed., *Frontiers of Science and Philosophy*. Pittsburgh, PA: University of Pittsburgh Press.
- Sellars, W. (1965) 'The Identity Approach to the Mind-Body Problem.' *Review of Metaphysics* 18: 430-51
- Sellars, W. (1975) 'The Adverbial Theory of the Objects of Sensation.' *Metaphilosophy*, 6: 144-160
- Sellars, W. (1971) 'Science, Sense Impressions, and Sensa: A Reply to Cornman.' *Review of Metaphysics* 24: 391-447
- Shoemaker, S. (1990) in R. Audi, (ed), *The Cambridge Dictionary of Philosophy*. Cambridge: Cambridge University Press.
- Sloman, A. (1999) *What is it like to be a rock?* Unpublished ms (available online at Aaron Sloman's home page: www.cs.bham.ac.uk/~axs/misc/rock/rock.html)
- Spencer, H. (1880) *Principles of Psychology*. New York: Appleton.
- Squires, R. (1974) 'Zombies vs. materialists II.' *Aristotelian Society Supplement* 48: 153-63
- Troland, L. (1928) *The Fundamentals of Human Motivation*. New York: Van Nostrand Reinhold.
- Tye, M. (1984a) 'The Adverbial Approach to Visual Experience.' *The Philosophical Review*, 93(2): 195-225
- Tye, M. (1984b) 'Pain and the Adverbial Theory.' *American Philosophical Quarterly*, 21: 319-328
- Tye, M. (1996) *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind* MIT Press.
- Walker, R. (1989) *The Coherence Theory of Truth*. Routledge.
- Wall, P (1999) *Pain: The science of suffering*. Weidenfeld and Nicolson.
- Williams, M. (1999) *Groundless Belief*. Blackwell, Oxford.
- Wittgenstein, L. (1958) *Philosophical Investigations*. Trans. G. E. M. Anscombe. New York: Macmillan.
- Young, P. (1959) 'The role of affective processes in learning and motivation.' *Psychological Review* 66: 104-125.