Consciousness and the Introspection of 'Qualitative Simples'

Paul M. Churchland
UCSD

I. Introduction. Philosophers have long been familiar with the contrast between predicates or concepts that denote or express "qualitative simples," as opposed to predicates or concepts that denote or express "structural, relational, causal, or functional" features. The tendency has been to think of these two classes of properties as being ontologically quite different from each other. Paradigm examples of the former would be features such as the redness of a tomato, the sweetness of sugar, the low pitch of a sound, and the warmth of a hearth. These particular examples, all features of things in the objective physical world, would be joined by a further population of presumed qualitative simples, features displayed in the conscious states of a human or other cognitive creature, features such as the qualitative character of your visual sensation of a tomato, of your gustatory sensation of a sugar cube, of your auditory sensation of a sound, and of your tactile sensation of a glowing hearth. Indeed, some may want to insist that the features displayed in this *private* cognitive domain are the *only* genuinely simple qualitative features, on grounds that their external brethren all turn out to admit of a structural, relational, causal, or functional analysis of some kind after all.

Whether or not this secondary claim is correct, we shall address anon. Let me continue this opening exploration of the contrast at issue by pointing out that predicates or concepts that denote *non*-simple properties are typically supposed to be *analyzable* or *definable* in terms of sundry other features, and in terms of the characteristic configuration of relations that severally unite those other features. Thus, the property of *being explosive* can be analyzed in terms of having the disposition to burst outwards suddenly, under suitable conditions of ignition. The property of *being in motion* can be analyzed in terms of continuously changing one's spatial position relative to some background frame of reference. The property of *being a unicorn* can be analyzed in terms having a horse-like bodily configuration plus large wings and a white coat. And so forth. The great majority of our concepts are said to fall into this latter ontological category. The qualitative simples, by contrast, form a comparatively tiny elite, distinguished by their *not* being subject to any such definition or to any such decompositional analysis. We acquire these simple concepts – we learn the meaning of these simple predicates – by *ostension*, it is said, rather than by composition from concepts or predicates that we already command. This special semantic status, it is said, is further reflected in the special *epistemological* status enjoyed by the judgments that record the occurrence of these qualitative simples. We do not recognize an instance of redness, for example, by way of recognizing the peculiar configuration of some more basic features that collectively constitute a case of redness. We simply recognize a case of redness directly or immediately. We need not be *infallible* in apprehending instances of such qualitative simples, but our apprehension of them, whether in perception or in introspection, is

decidedly non-inferential and deeply inarticulable. We cannot say *how* we recognize such a simple feature. We just can.

Altogether, the apparent *ontological simplicity*, the *semantic autonomy*, and the *epistemological immediacy* of this smallish family of qualitative properties might well suggest that we are here looking at an ontologically special family of features, features that enjoy a unique status among the elements of reality. Certainly many philosophers have been inclined to claim a special ontological status for them, based on the several considerations just explored, and on various thought-experiments that are supposed to draw out their metaphysical consequences. In particular, a number of contemporary philosophers have argued that these qualitative simples, at least in their internal incarnation as features of our conscious experiences, are forever immune to the sorts of reductive/explanatory assimilations frequently displayed in the physical sciences. All parties may agree that water is H_2O, that light is electromagnetic waves, and that stars are thermonuclear furnaces. *These* 'intertheoretic reductions,' at least, are well-established parts of human knowledge, and they have successfully relocated water, light, and stars (and a great many other commonsense things) within a conception of physical reality that is broader and deeper than the everyday conceptual framework that was their original and more modest home. But a worthy minority of our profession deems it profoundly unlikely that the qualitative simples at issue above, the internal ones anyway, will ever find a similar fate. By their very nature, it is claimed, they are immune to intertheoretic reduction in terms of the properties embraced by the physical sciences.

It is not difficult, perhaps, to appreciate their position here. After all (and speaking fairly roughly), to reduce any given property to something recognized by the physical sciences is to successfully *reconstruct* the peculiar (structural, causal, relational, functional) profile displayed by the target property at issue, in terms of the conceptual resources of the particular physical theory that aspires to achieve that reduction. To use the examples already cited, the property of being water has a broad and characteristic profile of causal properties, as does the property of being light, and of being a star. And these complex causal profiles are precisely what modern chemistry, electromagnetic theory, and gravitational and nuclear physics, respectively, have successfully reconstructed in such illuminating detail. But on the face of it, at least, the target properties at issue in the preceding paragraphs *have no* such characteristic profile for the aspirant reducing theory to even try to reconstruct. They are, after all, qualitative *simples*; their unanalyzable qualitative character is what is essential to their identity; and so they present themselves as smooth-walled mystery to the reconstructive ambitions of at least the physical sciences. There is simply nothing there, apparently, for those sciences to get a reconstructive grip on.

Over the past forty years, considerations such as these have motivated a family of closely related arguments to the effect that the qualitative dimension of our internal conscious experience is forever beyond the reductive/explanatory reach of the physical sciences of the brain. Thomas Nagel was perhaps the first to pose the challenge by pointing out – quite plausibly, to judge by the paper's reception – that no matter how much one might know about the physical structure, operations, and states of the brain of a

bat, one would still not know "what it would be like" to have the experiences of a bat.[1] That is, objective information of the former kind, no matter how complete, would not suffice to specify the subjective *qualitative* character(s) of the bat's experiences.

Frank Jackson provided a similar argument focussed on an imaginary neuroscientist named Mary who was entirely colorblind or otherwise color-deprived from birth. Mary might come to know, he argued, everything there is to know about the physical operations of the brains of color-normal people, but, being colorblind herself, she would still fail to know what it is like to see the color red.[2] Only if her colorblindness were somehow reversed could she gain access to relevant qualitative character. At about the same time, James Levine published an essay that described the apparently unbridgeable "explanatory gap" between the resources of the physical sciences and the peculiar character(s) of our subjective mental "qualia."[3]

And David Chalmers subsequently produced a comprehensive book that celebrated these earlier arguments and added an argument or two of his own to underscore their collective point.[4] He, too, points to the "absence of an analysis" of any of the qualitative simples at issue, an absence he sees as diagnostic of their ontologically special nature. And he has us imagine a race of 'zombies,' creatures whose physical makeup and physical behavior is identical to ours – that is, they share with us *all* of the same physical/functional/causal/relational properties – but whose subjective *qualitative* mental life is simply absent. The fact that this scenario is at least conceivable, he argues, shows that what is *essential* to *our* internal qualitative states is something beyond what mere physical reality can hope to provide. Altogether, we have here a gathering consensus that the qualitative dimension of our conscious experience is something that the physical sciences, such as modern neuroscience, will never explain.

This conclusion, let us own at the outset, may be true. Conceivably, some form of Property Dualism will turn out to be the correct theory of mind, just as these authors severally suggest. And yet, one may want to pause here, to express amazement that such a spectacularly important factual claim should be legitimately established by arguments that arise entirely from the armchair, arguments based on preemptive and wholly *a priori* 'analyses' of the crucial properties involved, via considerations that are available to anyone who merely shares our *current* conceptual framework for comprehending conscious experience. Would that our theoretical understanding of some of the Universe's deepest mysteries were always so easily achieved.

My hesitation here, as many readers will appreciate, is not new. But in my earlier writings on this topic my impulse has always been to focus on either *1)* the conditions actually required for successful intertheoretic reduction,[5] a matter of some complexity

---

[1] Nagel, T. (1972), "What is it Like to be a Bat?" in *Philosophical Review*, 83, no. 4.
[2] Jackson, F. (1982), "Epiphenomenal Qualia," in *Philosophical Quarterly,* 32, no. 127.
[3] Levine, J. (1983), "Materialism and Qualia: The Explanatory Gap," in *Pacific Philosophical Quarterly*, 64.
[4] Chalmers, D. (1996), *The Conscious Mind*. Cambridge: Cambridge University Press.
[5] Churchland, P.M. (1985). "Reduction, Qualia, and the Direct Introspection of Brain States," in *Journal of Philosophy*, 82, no. 1.

and ongoing dispute even now, or on *2)* the genuine virtues of the emerging neuroscientific accounts of human sensory experience,[6] another unfamiliar matter of considerable complexity, or on *3)* the history of science, and the presumptive lessons that past scientific episodes provide for the issues confronting us in the present case.[7] All three approaches place serious demands on the scientifically marginal reader, and they may have been rather more opaque, to many, than I allowed myself to believe at the time. I take back nothing said in any of those papers, but on the present occasion I wish to take a more general and more philosophical approach to the anti-reductionist arguments at issue, in hopes of deflating their appeal in a more accessible manner. Those arguments not only run into trouble with the philosophy of science, with emerging neuroscience, and with the history of science generally. They lack integrity even by the standards of purely analytic philosophy. Or so I shall argue.

II. On the Determination of Essences. This undertaking is more difficult, more time-consuming, and altogether more entertaining than one might at first suppose. Let us agree that properties *have* essential features – following tradition, and without too much prejudice, we might call them 'defining' features or 'necessary' features. And let us agree also that it is a major part of the human cognitive adventure to discover the categorical structure of reality, to discover, that is, what *properties* the universe displays, and to discover what invariant or timeless relations unite and divide them. Learning about the world, after all, is not just a matter of determining which particulars happen to instantiate which properties, properties drawn from some antecedently settled population of universals. We, and all other cognitive creatures for that matter, have to *learn* the relevant properties or universals if we are ever to make contentful judgments about which particulars instantiate them. That is, we must *learn* the world's *general* features. We must learn the similarities and differences that collectively configure those features. And we must learn the many nomic or causal relations that often unite them in prototypical temporal sequences. The result of such a learning process is a *conceptual framework*, a background 'map' of the timeless structure of the universe. With such a map in place, one's sense organs can help one to locate, at some appropriate place within that structured map of categorical possibilities, whatever particulars or processes one happens to encounter. One then knows, assuming that the background map is accurate, what to expect of those local particulars as one's experience of them unfolds.

Constructing a conceptual framework that is even roughly adequate to the demands of one's practical experience is a major accomplishment, and it does not happen overnight. Humans spend years – indeed, decades (indeed, millenniums) – developing conceptual frameworks that are adequate to comprehending and navigating new domains of ever-increasing empirical complexity. Individuals at different stages of this long developmental process will display quite different conceptions of the universe's abstract structure, culturally or individually idiosyncratic conceptions that are different in the breadth, in the depth, and in the accuracy with which they portray that objective

[6] Churchland, P.M. (2005). "Chimerical Colors: Some Phenomenological Predictions from Cognitive Neuroscience," in *Philosophical Psychology* 18, no. 5. Reprinted in my (2007) *Neurophilosophy at Work* (Cambridge: Cambridge University Press).
[7] Churchland, P.M. (1996), "The Rediscovery of Light," in *Journal of Philosophy* 93, no. 5.

categorical structure. Improvement in any of these three representational dimensions constitutes *conceptual progress*. Such progress is a supremely important kind of empirical learning. Indeed, it is the most important kind of all, because it originally provides, and continually changes, the very concepts we attempt to apply in any and all of our singular judgments. Accordingly, trying to determine the essences of things is not the obscure and occasional indulgence of cloistered philosophers. Rather, it is the basic aim of the learning process in all cognitive creatures, and it is the basic aim of the empirical sciences generally.[8]

The essential character of our *conscious mental states* has become the focus of much interest and theoretical speculation in recent centuries. In a critical response to the impressive but alarming development of the "mechanical philosophy" in the Seventeenth Century, Descartes outlined a form of Substance Dualism which claimed, for *res cogitans* ("thinking stuff"), an essence or ontological status forever distinct from that of *res extensa* ("spatially extended stuff"). Notably, Descartes' argument at the time *also* used a thought-experiment, concerning what he could and could not imagine, in order to divine the presumed essence in question. Curiously, his argument was just the reverse of Chalmer's zombie exercise: Descartes thought he could successfully imagine *himself* without any of his *physical* features, and concluded that those physical features were thus not a part of his essential nature.

But by the time we reached the first half of the Twentieth Century, physicalism was once again asserting itself, and very powerfully, into the domain of the mental. For example, in a reaction against Descartes, the Logical Behaviorism of Ryle and Wittgenstein had begun to sweep through most of the philosophical profession, insisting that the essence of any given kind of mental state was simply the unique profile of observable-physical-circumstances-as-input / observable-physical-behavior-as-output relations that possession of the relevant state implied. The 'ghost within the machine' was thereby exorcised as an unnecessary ontological extravagance.

But with the ghost went all of our internal states as well, apparently, and their appropriately internal qualitative properties with them. Despite the background ideological pressure of the then-dominant Logical Positivism – with its emphasis on the epistemological and semantic primacy of *objective* observables – this blanket exorcism was persistently difficult to accept, even for us anti-dualists. Fortunately, Functionalism emerged in the 1960s, thanks to philosophers such as Putnam and Fodor, apparently to save the day. For these philosophers insisted that internal mental states were perfectly acceptable. They needed not to be exorcised from our ontology, but to be properly knit within it. They could be welcomed back into our ontology, it was argued, by acknowledging their role as *causal intermediaries* between sensory inputs and behavioral outputs, intermediaries with complex and mutually embracing causal profiles of their own, unique profiles that constituted the essential nature of each distinct kind of internal mental state. Those characteristic causal profiles might be realized in diverse physical

[8] What this most fundamental learning process consists in, at the neurobiological level, is explored in some detail in Churchland, P.M., *Plato's Camera: How the Physical Brain Captures a Landscape of Abstract Universals* (forthcoming: Cambridge, Mass., The MIT Press. 2011).

substrates, cautioned the Functionalists, but those possibly diverse substrates were not what was important. What was essential to the shared essence of these internal mental states was not the metaphysically simple qualities they displayed, nor the underlying medium in which they were realized – it was their shared causal/functional profile. It hardly mattered whether those profiles were realized in neural brain-stuff, in Cartesian mind-stuff, or in electronic computer-stuff, although Functionalists typically claimed that the empirical evidence was converging massively in favor of the first alternative, massively *against* the second alternative, and that the third alternative was growing as a future technological possibility.

What is noteworthy here is that once again we find genuinely gifted philosophers taking a close look at the domain of the mental, as that domain is currently comprehended within our existing conceptual framework, and then announcing, on the basis of that arm-chair examination, that the *essential* features of the elements of that domain are . . . (place your favorite ontological prejudices here). No substantive experiments are cited to sustain the analysis. No empirical theories are proposed or evaluated. And yet the 'analysis' here proposed was advanced with considerable confidence and authority, *despite* the fact that it stands in stark opposition to the 'analysis' proposed by Descartes, who possessed the same conceptual framework for the relevant internal domain that we do, but whose take on what were and were not its essential features was diametrically opposed to that of the Functionalists.

That philosophers can disagree is not news. But this is ridiculous. A possible explanation for this situation is that both forms of analysis are actually correct, but they comprehend *distinct dimensions* of our inner lives – the subjective/conscious/qualitative dimension in the one case, and the objective/causal/functional dimension in the other. This is in fact the line taken by the contemporary tradition that runs from Nagel through Chalmers discussed earlier, and it has at least a *prima facie* appeal. Not just because it resolves an awkward dilemma, but because the ontological division it proposes is antecedently plausible, at least to some. This line also constitutes, note well, yet a *third* 'analysis' of the structure of essential features within the domain of the mental, one no less born of the armchair than were the first two. It just bifurcates the elements within that domain, in ways that neither Descartes and Putnam were apparently willing to embrace. If one's confidence in the armchair route to understanding the mental were already growing thin, how likely is it to recover in the face of yet another such analysis, one that simply pastes together two antecedent failures?

Still, and its armchair origins aside, this third analysis might be correct. For the sake of argument, let us suppose that it is. What would follow about the real nature of our mental states? Absolutely nothing. For we still have to address the question of whether or not our current conception of the domain of mental states is an *accurate* or *faithful* portrayal of the actual elements and the real nature of that domain. Even if we have finally gotten it straight what our current conceptual convictions and commitments *are*, it remains a separate question whether those convictions and commitments are *correct*. After all, we humans have repeatedly been forced, by developments in the natural sciences, to *reconceive* a variety of things that were and remain central to our

dealings with the world. We used to think that the Earth was essentially motionless: indeed, it was thought to be the essential background bedrock or reference frame against which all genuine motions had to be reckoned in the first place. But it isn't. We used to think that Light was essentially that-which-made-things-visible. But the vast majority of kinds of light -- i.e., all wavelengths outside the tiny 'optical window' -- do no such thing, at least for humans. And even within that tiny window, making environmental information available to terrestrial creatures is an extremely peripheral feature of light, hardly its essence.

We used to think that the essence of Life was some kind of Soul or Vital Spirit. But it isn't. We used to think, without question, that Mass and Length were simple, one-place properties. But they both turned out to be more penetratingly and accurately reconstructed as *two*-place predicates, denoting a variable relation between a thing and a variety of reference-frames. And so on. Evidently, being a constituting element of one's current conceptual framework is hardly a guarantee of genuine essentiality, or even of bare truth, come to that.

At this point one might anticipate that I am going to argue that our current conceptual framework for mental states is defective in some major way. It might be, and I have so argued in the past, at least where the propositional attitudes are concerned.[9] But that is not my purpose in this essay. The focus of the authors cited earlier is on the ontological status of mental states with a distinctive *qualitative* character, and that will be my focus also. My aim is to eviscerate their arguments that the qualitative characters of these states are forever beyond an explanatory reduction in terms of the physical dimensions of brain activity.

III. Subjective Knowledge versus Objective Knowledge. I begin with Nagel's original argument, which leaned so hard on the distinction between subjective knowledge and objective knowledge. The basic idea was that the *objects* of these two kinds of knowledge, respectively, are completely disjoint and mutually exclusive, as are the two kinds of knowledge themselves. Accordingly, since the knowledge supplied by the physical sciences is always (and essentially?) objective, science can never give us subjective knowledge, nor, therefore, knowledge of its typical objects, namely, subjective qualitative characters.

Arguing for fundamental ontological distinctions on the basis of the idiosyncratic, historically relative, and changeable profiles of our supposed *knowledge* of those ontological categories is a dubious undertaking on its face, especially when, as in the present case, our knowledge of those categories is relatively paltry. But this is a general complaint, and I wish to register a highly specific objection to Nagel's prima facie compelling argument. *The typical objects of the two forms of knowledge at issue are not at all disjoint and mutually exclusive, even by the standards of current common sense.*

To begin, a great many objective and plainly physical facts and features are made directly available to oneself by introspection. The current configuration of one's body,

[9] (1981) "Eliminative Materialism and the Propositional Attitudes," *J. Phil.* **78**, no. 2.

for example, is something of which one is directly and continuously aware, such as being in a sprinter's starting crouch, or being seated, or having one's arms folded in front of one. The (wholly physical) state of contraction and tension of every muscle in the body is made continuously available to the brain by the body's *proprioceptive* system. (And a good thing, too, if the brain is to exercise continuous control over an ever-changing bodily configuration.) One knows one's own bodily configuration in a way that no one else can or ever will. Those others have such subjective access to *their own* bodily configurations; but not to yours.

Similarly, even with a constant bodily configuration, one knows directly if one is being gently rotated (as in a barber's chair) or gently rocked to the left and right (as on a large ferry boat), even if one's eyes are closed and one's tactile senses are disabled. The *vestibular* apparatus of the inner ear (the innervated semicircular canals) provides the brain with a super-sensitive monitor of any rotational changes in the head's position. One is aware of such changes in one's own case in a way that no one else is. Others may see you rotate in various ways, but they will never access your rotation as you access it. And yet, and as with static bodily configurations, such rotations are as objective and as physical as can be.

To continue, one knows that one's stomach is full, or that one's bladder is full, in ways that no one else can know it. Yet these are objective facts about physical things. As well, one knows when one's muscles are seriously overtired, as after protracted stressful work (they are then awash in lactic acid, a chemical byproduct of biological energy use) in a way that no one else knows their weary condition. And one knows, with arresting introspection, when that condition occasionally produces a 'cramp,' a spontaneous maximal contraction that freezes a given muscle into excruciating immobility. One knows that one's sinus cavities are swollen, as with the common cold, in a way that is difficult to articulate, but unambiguous even so. And no one else will know their swollen state as you do.

These examples can be multiplied indefinitely. They are all instances of what psychologists and neuroscientists call *interoception*. Evidently, one knows a great deal about what is going on inside oneself, and knows it in ways forever denied to creatures not identical with oneself, even though those introspectively accessible states and activities are ostentatiously physical. Introspective apprehension, accordingly, is clearly *not* confined to states of a non-physical nature, let alone to states that are 'qualitative simples'. It includes physical facts of substantial complexity, although that complexity is often only partially or only dimly apprehended, depending on the particular physical state involved. Indeed, the degree to which one spontaneously appreciates the complexities that may be involved in these internal states can *vary* as a function of how much one has *learned* about the relevant kinds of physical states.

For example, an young infant's proprioceptive apprehension of its own bodily configuration is presumably of much lower 'resolution' than that of an older child's, save perhaps where the mouth, lips, and tongue are concerned, elements of the infant's body that are already hard at work. Similarly, the proprioceptive apprehension, by a pianist,

guitarist, or harpist, of his complex hand-and-finger positions while playing his instrument is markedly superior to the same apprehension by one untrained in musical performance. A skilled typist shows the same advantage over a non-typist. And in the case of a professional ear-nose-and-throat doctor, her interoceptive appreciation of the details of her sinus infection (e.g., which of the several cavities is affected, and how) will be rather greater than in the case of a naïve youngster. In sum, there simply is no dividing-line that excludes physical facts from the domain of introspectively accessible facts. That domain includes a multitude of physical facts and physical things. And *how far* these undoubtedly physical facts intrude into that 'favored' epistemological domain *varies* both with time and with increasing knowledge.

We can illustrate, experimentally, this sort of epistemological intrusion without your even having to put aside the page you are now reading. When one reads black text on a white background, one's visual system is chronically fixed on some one or other horizontal line of *text*, rather than on the empty (white) horizontal spaces that separate those successive lines of dark graphical elements. Your eyes move left and right and up and down, to be sure, but while reading text, your visual field contains a roughly constant grid of white and dark horizontal lines. The white lines, being vertically fixed in your visual field, and being brighter than the lines of text, produce a form of fatigue or adaptation in the visual neurons that code what you are seeing, a fatigue that is confined to the neurons that chronically code for those white lines. The result is that, when you suddenly shift your vision to a surface of *uniform* brightness (such as the empty margin at the bottom of this page), the fatigued neurons all fail to respond normally to the relevant parts of the now-uniform surface. They represent those parts as being darker than they really are. The result, subjectively speaking, is an *after image* of light and dark horizontal lines, an image that is brightness-inverted relative to the original page of text. To see this vividly for yourself, simply fixate rigidly on some word in the middle of this line for ten seconds or so (count slowly), and then relocate your gaze on the empty page-margin below. The after-image will be obvious, although it will begin to fade within a second or two as the relevant neurons begin to recover from their induced fatigue.

What you are then noticing, perhaps for the first time, is the fatigued or energy-depleted state of a specific subset of the neurons in your visual system, an entirely physical condition. You know your condition subjectively, that is, in a way that makes it evident to you, but to no one else. After all, it is your after-image. And you would probably never have known of your neuronally fatigued condition, but for the physical description I gave you, and the physical instructions that went with it. But in principle, this case is no different from the familiar case of knowing, introspectively, that your *muscles* are fatigued.

In fact, such visually evident fatigue-patterns are a constant intrusion into our visual experience, but the brain tends, quite rightly, to ignore them or look past them as regrettable noise. We are mostly unaware of them. But once they have been pointed out to you, you start to notice them almost everywhere, especially in cases where one's external visual experience involves sharp brightness-contrasts.

Subjective knowledge, then, is *not* confined to some ontologically special class of nonphysical sensations. It regularly concerns the condition of one's stomach, one's viscera, one's muscles, one's visual nervous system, one's sinuses, one's overall skeletal configuration, and one's bodily motions – physical conditions all.

Still, it will be said, subjective knowledge itself remains distinct from the various forms of objective knowledge, even if their typical objects frequently overlap. And those epistemological objects that are knowable *only* subjectively, if there are any, might yet form an ontologically special class of things. That is, if there is a domain of phenomena that *cannot* be known by any objective means, then perhaps the fortunes of nonphysical qualia might be worth betting on after all.

Sensations themselves and their qualitative characters (as opposed to the physical conditions that they frequently signal) are the preferred candidates for this supposed role. These things, it is often claimed, are known *only* subjectively, never objectively. But this claim is false on its face. I have systematic and ongoing knowledge of the sensational states of the people near and dear to me – of their pains, their hunger, their anxieties, the warmth they enjoy or the cold they endure, the tastes they encounter (and like or dislike), even (as we saw) the detailed quality of the visual after-images they may have. To be sure, I do not know these things in the way that they do, but I certainly have knowledge of these things. That same knowledge governs much of my daily behavior. I can even *explain*, in neurophysiological terms, some of the more interesting facts about at least some of their subjective lives. Short of a blanket skepticism about our knowledge of Other Minds, then, we seem once again denied any uniform essence that would mark off the domain of sensations as ontologically distinct in some way. One knows about physical conditions both objectively and subjectively. And one knows about phenomenal conditions both objectively and subjectively. So far, no dividing essence has emerged.

But we have not yet addressed the most salient and the most widely cited element of Nagel's overall argument, the element that motivated his paper's title. Once again, it involves a thought experiment, but an admittedly compelling one. He asks you to imagine that you have somehow come to know *everything* about the physical nature and the physical activities of a bat's brain. However, and despite your exhaustive command of the physical details of bat cognition, you still wouldn't know *what it is like* to *be* a bat-style cognizer. You still wouldn't have the *subjective* knowledge of the bat's cognitive and sensory life that the bat himself has. From this, he concludes that there must be something missing – something real and something important – from the purely physical story that you have learned. Knowing the complete physical theory of bat-style cognition wouldn't *make you* a bat-style cognizer.

Indeed it wouldn't. What is required to make you a bat-style cognizer – to make you enjoy the special dimensions of a bat's subjective cognitive activity here at issue – is that the complete physical theory of bat-style cognition *be true of you*. Whether or not you happen to *know* that theory is utterly irrelevant to whether or not you actually have bat-style cognitive states. The natural-born bat doesn't have any inkling of that theory (even though it is true of him), and yet he clearly doesn't need it to enjoy the subjective

dimensions of his own existence. And you may know the physical theory in exhaustive detail, as in Nagel's thought-experiment, but that wouldn't give you what the bat has. Simply knowing the theory doesn't make the theory *true* of you, not in this case, and not in any other case either.

To cite parallel examples, having complete knowledge of the physical nature of superconductivity doesn't make you a superconductor. (That would require that the theory of superconductivity be true of you.) Having complete knowledge of the physical nature of pregnancy doesn't make you pregnant. (That would require that the theory of pregnancy currently be true of you.) Having complete knowledge of the physical nature of diabetes doesn't make you a diabetic. (That would require that the theory of diabetes currently be true of you.) The fallacy involved in these parallel cases is immediately obvious. Why wasn't the case of bat-cognition similarly obvious? Because the bat-case concerned your gaining, or rather, *failing* to gain, a certain form of *knowledge*, in a circumstance where your scientific/physical/objective knowledge of bat-style cognition was supposedly complete. The failure here, accordingly, looked like a failure in the reach of that scientific/physical/objective knowledge, at least where subjective phenomena are concerned. But it isn't a failure of that kind at all. The proper test of that scientific/physical/objective theory of bat-style cognition is whether, when that theory happens to be genuinely *true of* some given creature, then the creature actually *has* the subjective experiences of a bat. And nothing in Nagel's paper suggests, even for a second, that a complete theory of bat neurophysiology would fail *this* test.

In sum, Nagel is implicitly demanding or expecting that mere possession of a certain body of *theoretical* knowledge should *constitute* (as opposed to describe or explain) a quite *distinct form* of knowledge: bat-style subjective cognition. But there is not the remotest reason to expect any such thing, and no ontological lessons to be drawn from the utter failure of the neurophysiological account to actually *provide one with* bat-style cognition, subjective *or* objective. As is illustrated in the parallel cases just listed, that expectation is wholly unreasonable in the first place, and its failure to be fulfilled is entirely without significance for the adequacy of the particular theory involved. Most especially, the 'failure,' in every case, is without any *ontological* significance for anything.

These same considerations undermine Jackson's closely similar argument concerning the effect (or the lack of it) of possessing complete physical knowledge (of the brain activities of color-normal people) on the visual experiences of color-blind Mary. Mary, you will recall, was supposed to be neuroscientifically omniscient, but despite this distinction, she *still* didn't know what it was like to have the subjective visual experience of red. As it was often argued, "She knows all of the physical facts, but there is still something she does *not* know; so there must be some *non*physical facts."

But here again, Jackson is expecting, quite wrongly, that one form of knowledge should *constitute* a quite different form of knowledge. He is expecting that explicit/discursive/scientific knowledge should somehow *constitute* subjective knowledge of visual experiences. But that expectation is unreasonable on its face. As well expect

that your exhaustive discursive knowledge of the micro-organization of a professional golfer's motor cortex would constitute actual practical 'knowledge,' on your part, of how to hit a golf ball 200 yards down the middle of the fairway, even if you have never swung a golf club before in your life.

This last analogy gives a specific voice to a classic objection to Jackson's original argument, namely, that it is formally invalid by reason of equivocating on the term "knows." Nemirow[10] and Lewis[11] pointed out that the first occurrence of the term "knows," in the brief argument quoted in the preceding paragraph, denotes explicit or discursive knowledge, whereas the second occurrence of the term concerns a suite of cognitive *abilities* (to recognize red visually, to imagine red, to remember red, etc.). Given this equivocation, the case for 'nonphysical facts' evaporates.

One may indeed wonder exactly how to understand or analyze the distinct nature of 'knowing' in the subjective case, and in the intervening years much space in the philosophy journals has been spent pursuing that question. But even in advance of a settled analysis, it is plain that we are here looking at two distinct kinds of knowledge. Having discursive scientific knowledge of anything requires having a *language*. Knowing what it is like to see red requires nothing of the sort. As well, some three decades later, it now seems that the most promising place to *find* a discursive account of what-it-is-to-know-the-colors-subjectively lies in the emerging *science* of how biological brains actually represent the domain of colors, and how the visual system activates specific representations therein. Such neuronal accounts already exist, and they do not derive their plausibility from the armchair. More on such accounts anon.

IV. Back to the Sensations Themselves. If our sensations and their properties are the true claimants to some special ontological distinction, then perhaps we should focus on *them* in order to reveal that distinction, rather than on the not-so-distinctive profile of how we happen to know them. This is the line that Chalmers takes, and we need now to examine his rather different approach to our question.

Chalmers also focuses our attention on the *intrinsic qualitative characters* of our sundry sensations, in contrast to the causal/relational/functional profiles that those same sensations may also display. That latter dimension of sensational reality, he is happy to concede, is extremely important for our ongoing explanations of human behavior, and it may well find a successful and exhaustive reductive explanation in terms of physical neuroscience. Indeed, he positively expects this to happen. But the qualitative features of our sensations are a different matter, according to him. Those features are ontological *simples*, he avers, and for that reason, they offer nothing in the way of internal structure that a physicalist theory of the brain might hope to *re*construct, as a successful intertheoretic reduction would require. The qualitative character of my visual sensation-of-red, for example, simply confronts me, or so it seems. It does so in a manner quite

[10] Nemirow, L. (1980), "Reveiew of Thomas Nagel's *Mortal Questions*," in *Philosophical Review* Vol. 89, no. 3.
[11] Lewis, D. (1983), "Postscript to 'Mad Pain and Martian Pain'," *Philosophical Papers, vol. 1* (New York: Oxford U.P.).

distinct from any alternative sensation-of-color, but not because it invites any hope of some signature 'decomposition' into anything else at all, let alone into something physical. Such *qualia*, as they have come to be called, should therefore be counted as something outside the physical order, as something beyond the causal/functional profile in which our sensations are admittedly embedded. Chalmers' dualistic conclusion here is thus an instance of the long-familiar position called *epiphenomenalism*, a term he explicitly embraces as descriptive of his position.

We may open our examination of this argument by noting that the bulk of one's sensational life is characterized, not by simplicity, but by an extraordinary and ever-changing *complexity*. Listening to a conversation, looking around a flower garden, tasting a braised-lamb stew, smelling the aromas in a wood-working shop – our sensations in such cases display intricacies that are amazing. And not always obvious. A young child may not appreciate that the distinctive taste of her first ice-cream cone *resolves* intself into sensations of sweetness, creaminess, and strawberry. And it may take her awhile to learn that such decompositions are both common, and useful to keep track of. For the complexities we encounter are indeed composed, quite often, of simpler elements or constituting dimensions. In time, we do learn many of those simpler dimensions. A dinner-table conversation contains my brother's unique voice as an identifiable element; the complex flower-garden displays the striking orange of a typical poppy blossom; the lamb stew displays the distinctive taste of thyme, sprinkled into the mix at the outset; and the smell of yellow cedar stands out from the other smells in the wood shop, at least to a seasoned carpenter. Each of these particular qualitative features of one's inner phenomenological life is certainly a *simpler* dimension of a more complex whole.

But is each of these examples, or any of them, itself an ultimate, undecomposable simple? Perhaps. But how does one tell? I may indeed be *unable* to specify any sub-dimensions whose peculiar concatenation constitutes the sound of my brother's voice, or the poppy's visual orange, or the taste of thyme, or the smell of yellow cedar. But neither could the still-learning child specify, at least at the outset, the taste of sweetness, the taste of creaminess, and the taste of strawberry-ness as constituting sub-dimensions of her taste of the ice-cream cone, even though those elements were undoubtedly there, and even though she subsequently came to appreciate them. How, then, do I know when I have genuinely 'hit bottom' in a given case, as opposed to merely having reached the current limits of my capacity to articulate *how* I manage to discriminate the qualitative feature at issue?

This question has a certain bite to it in the present context, because hitting at least a *local*, *current*, or *apparent* 'bottom,' note well, is absolutely inevitable on *both* a dualist *and* a materialist account of how we discriminate the qualitative characters of our sensations. The only alternative to an apparent or presumptive 'bottom,' *some*where or other, is an infinite sequence of qualities discriminated via a recognizable concatenation of simpler qualities, each discriminated via a recognizable concatenation of still-simpler sub-qualities, where each of those is discriminated in turn via a recognizable concatenation of still-simpler sub-sub-qualities, and so on without end. Qualitative

characters that are at least *apparent* simples are thus utterly inevitable on *both* approaches to understanding the mind, dualist and materialist.[12] Their undoubted existence, accordingly, implies nothing one way or the other about their underlying ontological character, despite a widespread presumption that it speaks, somehow or in some degree, in favor of dualism. It doesn't. Every cognitive creature, even in an *exhaustively physical* universe, must display a current limit on how far it can decompose the qualities it can discriminate. Let us not be too impressed in our own case, then, by the mere existence of apparent 'qualitative simples,' however robust their apparent simplicity. Their existence is entailed by *both* of the philosophical approaches here at issue. Such 'simples' simply have to be there, if only to mark the limits of our current understanding. Their existence is not only *consistent* with both of the ontological positions at stake here; it is positively *entailed* by both of them. Accordingly, to infer the ontological simplicity of a given qualitative character from its *apparent* simplicity is to commit the fallacy of Arguing from Ignorance, as in, "I am *unaware* of any constituting elements in this qualitative feature, therefore, there *aren't any* constituting elements."

This *a priori* point looms larger when we reflect on the fact that the domain of *external, objective* things and properties displays exactly the same contrast between complex, decomposable features and (apparently) simple features as is found in the subjective realm. The *objective* red of an apple and the *objective* temperature of warm water, for example, are also *apparent* simples, ontologically; they are without definitional analysis, semantically; and they are 'immediately' accessible, epistemologically. But no one since the Eighteenth Century supposes that such objective perceptual properties are thereby revealed as genuine ontological simples. The physical sciences of objective color, temperature, sound, and so forth have provided us with decisive analyses of the underlying ontological *complexities* that constitute the objective perceptual qualities here at issue. Those qualities are entirely real. But they are also entirely physical and more than a little complex. It just took the physical sciences awhile to learn about their constituting elements. Our *internal* phenomenological qualities may be awaiting a precisely similar fate.

Indeed, the waiting period seems already to be over. But I will return to the matter of the emerging Neuroscience of qualitative states in a few pages. Let us here focus on the earlier and quite independent complaint that nothing of any ontological significance *follows from* either the epistemological opacity of our current sensational discriminations, or from the semantic/analytical simplicity of those qualities as judged by the lights of our current conceptual framework. As we saw, the claim that the subjective qualitative characters at issue are ontological simples is evidently not the outcome of a sound demonstrative argument based on either or both of these two premises. Rather, that ontological claim looks more like an *explanatory philosophical hypothesis* whose

[12] This important fact is evident even in the case of Chalmers' 'zombies.' On his own hypothesis, the zombies behave, speak, and argue exactly as we do, and therefore encounter the same decompositional limitations that we do when addressing their own inner states. They, too, *despite being purely physical*, confront what they, too, describe as a family of 'qualitative simples,' and they are no less puzzled by them than we are. Indeed, if they embrace Chalmers' line of argument (and it will be as plausible to them as it is to us), they will end up believing that they, too, have nonphysical qualia, when, *ex hypothesis*, they *don't*.

hope is to provide a uniquely compelling explanation of those two premises. After all, if our conscious qualia really *are* ontological simples, wouldn't you expect that our discrimination of them, one from another, would be inarticulable? And wouldn't you expect that our concepts of them would be without internal structure?

Perhaps so, but we must remind ourselves that we can already point to independent explanations of both premises, explanations that do not engage in weakly-motivated ontological profligacy. Moreover, if the postulation of ontologically simple supra-physical qualia is to purchase its plausibility by means of its comparative explanatory virtues, as the above interpretation suggests, then that postulation must be prepared to have its own explanatory virtues explored and evaluated in some detail. To that end, let us look into its actual performance.

V. The Explanatory Performance of Epiphenomenal Qualia.

Exactly *what are* the phenomena that the postulation of epiphenomenal qualia is supposed to explain? It is hard to see what they might be, for the simple reason that the postulated qualitative simples at issue are held to be *epi*phenomena, phenomena that are caused *by* physical phenomena in the brain, but *which have no causal properties of their own* – not within the physical realm, and not among each other either. They are held to be causally inert: a dynamically impotent sideshow, continuously reflecting the brain's activity, to be sure, but with absolutely no causal effects of their own.

How, then, can they possibly provide systematic *explanations* of anything at all? On the epiphenomenalist's own hypothesis, qualia are precluded from explaining anything about our bodily behavior: that must be done by appealing to the facts about our physical environment and its interactions with our brains. They are precluded from explaining anything about the behavior of our brains themselves: that job is exhausted by the physical neurosciences. And finally, they are precluded from explaining anything about *each other*: that job is exhausted by the idiosyncratic physical activities of each person's brain. Epiphenomenal qualia have no causal effects on one another, nor, indeed, on anything whatever. On the face of it, then, they are explanatorily impotent.

"Well, no," it will be objected, "for they do explain the *existence* of consciousness. Collectively, the complex flux of your epiphenomenal qualia *constitutes* your ongoing consciousness. Without that supra-physical flux, there would be no genuine consciousness. There might be the purely 'functional' form of consciousness displayed by Chalmers' zombies during their 'waking' hours, but there would be no *qualitative* consciousness."

This is the core claim of epiphenomenalism. But there remains a stubborn problem. Indeed, there are at least two of them. The qualitative features at issue cannot constitute someone's consciousness unless they are somehow *apprehended* by that someone, unless their local instantiations are *detected*, *noticed*, *registered*, or *recognized* by that someone. But on the epiphenomenalist's own story, to state the first problem,

those qualitative features are wholly without impact or causal effect of any kind on anything. In what, then, does their *apprehension* consist?

And to state the second problem, there is no 'someone' there to do the apprehending or conceptualizing in any case. The epiphenomenalist explicitly eschews any form of substance dualism, and, *ex hypothesi*, the qualitative features at issue can have no causal effects on the physical brain. Who and/or what, then, is 'home' to host, enjoy, or somehow *respond* to this qualitative 'show'? Evidently, no one and/or nothing. To be sure, the physical brain, or some part of it, is the true *subject* of each proposed qualitative feature, on the epiphenomenalist's own account. They are supposed to be supra-physical features of the *brain*. But on that same account, the brain itself is supposed to be totally and eternally blind to the occurrence (and to the *non*-occurrence as well: recall Chalmers' zombies) of any and all such supra-physical features. The price of epiphenomenalism, apparently, is the absence of any*thing* to be aware *of* the supra-physical features that the position itself proposes. Accordingly, those qualitative features themselves disappear from the causal matrix of the world in general, forever undetectable by anything, into an inaccessible metaphysical vacuum, where, beyond merely existing, they do precisely nothing, even to each other.

As an explanation of consciousness, this is a train wreck. Aside from failing to provide *any* positive explanations concerning the qualitative contents of consciousness and their causal role(s) in our cognitive economy and our physical behavior, and aside from leaving it an absolute mystery what these 'ontological simples' are and why they should exist at all, epiphenomenalism is flatly inconsistent with the core conviction of our common-sense conception of mental phenomena, namely, the conviction that our conscious mental states are *causally involved* in the unfolding drama of our conscious mental lives, and *causally responsible* for the unfolding physical behaviors to which it continuously gives rise. The point being made here is that the epiphenomenalist's claim to be faithful to our antecedent conception of our mental states is a five-star fraud to begin with. The allegedly fundamental division the epiphenomenalist draws between our conception of the causal/relational/functional aspects of our inner states, on the one hand, and our conception of the qualitative/introspectible aspects of those states on the other, is not a mutually-exclusive division that our common-sense conceptual framework respects at all. On the contrary, commonsense ascribes *both* kinds of aspects/properties to one and the same internal states, and it portrays their qualitative characters as an integrated *part* of the avowedly causal activities in which those states participate.

To illustrate this point, the state of *pain* is perhaps the first of many hundreds of examples that jump to mind. If a pain is strong enough to register in one's consciousness in the first place, then the familiar and unwelcome *qualitative* character that it displays will prompt one's attention to its possible causes, provoke aversion to its presence, kindle practical reasonings aimed at relieving it, distract one from one's antecedent activities, occasion regret at whatever you did to run afoul of it, and ultimately drive behaviors that one hopes will make it go away. The qualitative character of your pain is not a disconnected bystander to this modest explosion of causal consequences: it is typically what ignites them all in the first place. That is to say, as our current Folk Psychology

conceives of things, the qualitative character of pains is a fully integrated *part* of the dynamical profile that pains typically display, not a causally impotent bystander to a causal process that, strictly speaking, does not include it.

The case of pain is typical. The qualitative characters of *all* of our sensational and emotional states are causally potent elements in the dynamical profiles of each of those states. The dynamical profiles vary, of course, across the wide range of such sensational and emotional states, but those diverse causal profiles are just a further reflection of the diverse qualitative characters that give rise to them.

The situation here, *within* the narrow dynamical domain of human and animal cognition, is not different from the situation within the much larger dynamical domain of the physical world at large. As we noted, that larger domain *also* displays a great many *qualitative* features such as the objective *pitch* of a sound, the objective *warmth* of the air in an oven, the objective *redness* of a ripe strawberry, and so on at considerable length. (Throughout history, these, too, have often been thought to be 'ontological simples.') And these qualitative features are *also* causally integrated elements in the dynamical profiles that the objective physical world displays. The pitch of a sound – a middle A, or 440 Hz, for example – is causally related to many things: to the wavelength λ of that sound, for one, via the equation $\lambda=\nu/\omega$, where ω is the pitch and ν is the velocity of sound. (The wavelength of that sound must therefore be 340 m/s divided by 440 Hz = .773 meters. Change the pitch – the qualitative feature at issue – and you will thereby cause the wavelength to change.)

The warmth of the air in an oven – 300^{o}F, say – is also causally related to many things: to the fact that a cup of water will eventually come to the boil if placed in that oven, for example. The redness of a ripe strawberry – which has an overall electromagnetic reflectance peak at around .63 μm – will have characteristic causal effects on a spectrometer, and on the angle at which reflected light will be refracted through a prism, and (of course) on the human eye itself. These *external* qualitative sensory characters, familiar to us all, are certainly not causally impotent, supra-physical epiphenomena. On the contrary, they are an integrated part of the world's causal structure, they and thousands of other robustly qualitative objective features as well. And we can see *how* and *why* they are thus integrated when we finally appreciate how they are constituted within the underlying ontological complexities of the physical world.

Why should we think that their inner analogs – the qualitative features of our own conscious states – are any different? Why should the states of the physical brain and nervous system, which even Chalmers agrees are characterized by the causal/functional profiles here at issue – *not* have qualitative features that are just as causally integrated within the relevant dynamical profiles as are their manifold external brethren? Why should being located *inside* the skin introduce such an enormous ontological contrast with qualitative states that are located *outside* the skin? What *motivates* this lack of parity in one's construal of these two classes of qualitative features, especially when it flies in the teeth of the evident convictions of common sense, and of the daily explanatory practices that they make possible for all of us?

This presumptive parity between the semantic, ontological, epistemological, and causal status of the qualitative features of both our inner states and the world's many outer states finds a further parallel when we look at the business of *explaining* their various phenomenological characters and causal profiles in terms of the underlying physical reality that Physics, Chemistry, and Biology have been slowly revealing to us. We all know that the pitch of a sound is the oscillatory *frequency* of a compression wave in the atmosphere. We all know that the temperature of the air in an oven is the mean kinetic *energy* of the molecules that make up the air. We all know that the redness of a ripe strawberry is a peculiar *reflectance-efficiency profile* that leans strongly toward the long wavelengths within the optical range of the electromagnetic spectrum. These 'outer' qualitative characters, and thousands more besides, have all found highly revealing reductive explanations from the relevant sciences, explanations that positively *account* for their causal/functional integration with the rest of the world.

Moreover, those same explanations also account for the *structure* of the mutual *similarity-and-difference* relations among the diverse *qualitative* features within a given qualitative domain. Thus, different *pitches* and different *temperatures* are each arrayed on a one-dimensional similarity continuum, as befits features that vary in only one dimension. (Namely, oscillatory frequency, and mean molecular kinetic energy, respectively.) And different *colors* are arrayed within a three-dimensional similarity space, as befits a feature that varies in three significant dimensions. (Namely, the spectral *location* of its global reflectance peak (its hue), the degree of *concentration* of that reflectance peak (its saturation), and the overall *area* under its energy-reflectance profile (its brightness).)[13] Here we see the underlying physical theories providing systematic explanations of central *qualitative* facts concerning the qualitative features themselves, and not just of their causal/functional profiles.

Finally, and perhaps most importantly, note that the explanations that science now provides for the external, objective qualitative features discussed above reveal that *they are not ontological simples at all*, despite a fairly convincing first impression. The oscillatory frequency of a compression-wave train in the atmosphere is a modestly *complex* phenomenon. So is the mean of the kinetic energies of the millions of ballistic molecules that make up any gas. And so is the 3-dimensional configuration of the relevant three aspects of the strawberry's electromagnetic reflectance profile. Our native sensory organs are *causally sensitive* to these complex properties, to be sure, which is why we can detect them so reliably, but neither our sensory organs nor our brains have any initial cognitive inkling of the ontological complexities that constitute them. That difficult matter is for the relevant sciences to address and reveal, not for our unaided mechanisms of bare discrimination. And so, in our uninstructed ignorance, we are naturally but wrongly tempted to construe these several properties – the pitches, the

[13] For an accessible account of the underlying nature of objective colors, see Churchland, P.M. (2007), "On the Reality (and Diversity) of Objective Colors: How Color-Qualia Space is a Map of Reflectance-Profile Space," in *Philosophy of Science*, Vol. 74, no. 2. Reprinted as Ch. 10 of Churchland, P.M. (2007), *Neurophilosophy at Work* (New York: Cambridge University Press). Also reprinted in Matthen, M., and Cohen, J., eds. (2010), *Essays in Honor of Larry Hardin* (Cambridge: The MIT Press).

temperatures, and the colors – as qualitative and ontological simples, even though they are nothing of the sort.

These external cases provide a clear lesson for addressing our focal case of *internal* qualitative characters. The qualia of our inner states are *also* spontaneously discriminable, one from another, by our native interoceptive mechanisms, whatever those mechanisms might happen to be. Not surprisingly, those native, internal discriminatory mechanisms are *also* cognitively blind to whatever ontological complexities might happen to underlie those internal qualitative characters, and cognitively blind to how those complexities might play a causal role in the discriminations at issue. Just as we found in the outer case. But here also, this does not mean, not for a second, that such underlying ontological complexities are *not there*. Indeed, given that the brain is more complex, by far, than a compression-wave train, or an oven full of gas, or a light-reflecting surface, we should positively *expect* that its internal states will possess extraordinary ontological and causal complexities, complexities *that are initially opaque to our native discriminatory and cognitive comprehension*. Those internal states may be spontaneously discriminable by us, one from another, but finding out exactly *what* it is that is being discriminated, and *how*, is a job for the *sciences* of the brain, in strict parallel to the external cases discussed in the preceding paragraph.

That job, as was pointed out earlier, is already well under way. How sounds are processed and represented in the cochlea of the inner ear, so as to send a range of qualitatively distinct (and highly complex) neuronal activation-patterns to the auditory cortex, is a matter that is already understood. The cochlea is wonderfully configured to do a fine-grained energy-profile analysis across the frequency spectrum of any incoming sound. So also – though here the story is still provisional – is the manner in which those peripheral inputs are synaptically transformed and subsequently coded within a multi-dimensional similarity-and-difference 'space' of activation-patterns within the auditory cortex itself. (Pitch, of course, turns out to be only one of many dimension of variation among sensations of sound.)

The same is true for the brain's internal representations of color, both at the retina and in the brain's downstream cortical area V4. The Hurvich-Jameson model mentioned earlier, of how chromatic information is both processed and represented in the brain, gives us a detailed account of the neuronal niceties that underlie subjective human color experience, an account that gives a highly illuminating explanation for the internal *phenomenological* structure of human color-qualia space itself, that is, of the similarity-and-difference relations that severally unite all of the colors.[14] It further provides predictions of and systematic explanations of the tens of *thousands* of distinct *color after-images* that are produced when one fixates for time on any one of a hundred (different) colored circles, and then relocates one's gaze on any one of a hundred (different) uniformly colored backgrounds. The resulting circular after-image will have a distinct, predictable, and entirely explicable color-quality different from either of the two contributing stimuli. The model even predicts the existence of, and tells us how to

[14] See again Churchland, P.M. (2005), "Chimerical Colors: Some Phenomenological Predictions from Cognitive Neuroscience," *Op cit.*

produce, color sensations of qualitatively *novel* sorts, such as sensations of a 'red,' or a 'blue,' or a 'green,' each of which is simultaneously as black as the blackest-possible black. I know this description sounds impossible, or semantically ill-formed, but the predictions turn out to be true and the mechanisms involved are straightforward.

Evidently, and as rightly expected, the domain of internal qualitative features is not at all explanatorily impenetrable by the resources of the physical sciences. Just as in the case of the external qualitative features, we already possess some striking explanatory accounts of the nature and contents of our internal qualitative lives, and it would be foolish not to expect more. None of this strictly *entails* that epiphenomenalism is mistaken about the ontological status of our inner qualia, but that position is currently being overwhelmed by an alternative tide of explanatory success, and the position's initial strength derived from nothing more than a highly prejudicial 'analysis,' and a whopping 'argument from ignorance' in any case, both of which missteps are unmasked by the considerations of the preceding pages.

Accordingly, the truth would seem to be that absolutely *none* of the 'apparently simple' qualitative characters that grace our inner lives are genuine *ontological* simples at all. They are, all of them, complex neural and physiological states, states whose qualitative characters are embodied in that precious physical complexity. The dynamical activities of the brain are positively *driven* by those very physical complexities, and so the philosophical claim that these alleged 'simples' are also causally impotent bystanders to the brain's dynamical adventures is flatly inconsistent with the recent insights of Neuroscience, as well as with the antecedent convictions of Folk Psychology. It may be that the overall cognitive profile that characterizes *conscious* brain activity remains to be understood. Indeed it does.[15] But the account of our qualitative conscious states offered by epiphenomenalism holds out no analytical or explanatory virtues to tempt us towards that position, and the competing neurobiological account of those very same states already holds out a broad range of ontological, explanatory, and predictive virtues that pull us in precisely the opposite direction. Add up their respective contributions to our current understanding and there is simply no contest. Epiphenomenalism will soon be a museum piece.

The more common forms of Property Dualism – which do not attempt to disconnect our inner qualitative characters from the dynamics of our cognitive activities – are not quite so badly off as is epiphenomenalism, for they do not fly in the face of the constituting convictions of Folk Psychology and the explanatory practices they sustain. But, as has been known for more than fifty years, these less extreme forms of Dualism do fly in the face of basic Physics itself, a rather more damning matter, since any position that includes non-physical elements in the causal dynamics of the brain must violate both the law that energy is neither created nor destroyed, and the law that the total momentum in any closed system is always conserved. In short, you simply can't get a change in any aspect of the physical brain (for that would causally require both energy changes and

[15] Although, for an opening stab at what such a cognitive profile might look like, and how it might be embodied in the recurrent structure of the brain's global 'wiring diagram,' see Churchland, P.M. (1995), *The Engine of Reason, The Seat of the Soul* (Cambridge: The MIT Press), pp. 211-226.

momentum changes) save by a compensatory change in some other *physical* aspect of the brain, which will thereby lay claim to being the cause at issue. There is simply no room in a physical system for ghosts of any kind to intervene in some fashion to change its dynamical behavior. Any physical system is 'dynamically closed' under the laws of Physics. (Indeed, it was this very difficulty, over a century ago, that initially motivated the desperate invention of Epiphenomenalism in the first place.)

Still, one might choose to simply reject, or somehow to circumscribe, the currently accepted laws of Physics, and contrive to make a case for an 'interactive' Dualism based on its comparative explanatory and predictive successes, relative to the same successes displayed by the physicalistic Neurosciences. This, I propose, is the only possible route by which an honest Dualism of any kind can hope to succeed. Any other route, as we have seen above, will involve nothing but subterfuge and self-deception. But if this honest route is to be taken, it must begin by acknowledging that, to date, ". . . Dualism is less a theory of mind than it is an empty space waiting for a genuine theory of mind to be put in it."[16] If the 'explanatory successes' of Dualism are to be fairly weighed against those of current Cognitive Neuroscience and of basic Physics, they must first be brought into existence. So far, there is nothing there to permit such a comparative evaluation to even begin. But while we are waiting, we can fairly contemplate the steadily accumulating and highly enlightening explanatory successes produced by our theoretical and experimental probings of the physical brain, even on the topic of qualitative states. After all, we will need to *know* about those successes, and in some detail, should the contest just imagined ever materialize.

[16] This quotation is drawn from a textbook published over a quarter-century ago: Churchland, P.M. (1984), *Matter and Consciousness* (Cambridge: The MIT Press), p. 19. Little or nothing has changed since then.