The Nature of Consciousness: Philosophical □ Debates/ Edited by Ned Block, Owen J. □ Flanagan, Güven Güzeldere; Cambridge: MIT □ Press, 1997. (1-69 p.)

INTRODUCTION

The Many Faces of Consciousness: A Field Guide

Güven Güzeldere

There is perhaps no other phenomenon besides consciousness that is so familiar to each of us and yet has been so elusive to any systematic study, philosophical or scientific. In thinking about consciousness, the puzzlement one often finds oneself in is rather like St. Augustine's riddle in his contemplations about the nature of time: When no one asked him, he knew what it was; being asked, however, he no longer did. (Augustine of Hippo 1961: Book 11.)

What is at the heart of this puzzlement? Is there a genuine difficulty that underlies it? What are the specific issues that comprise *the* problem of consciousness? (Is there really a "*the* problem of consciousness"?) And are we facing a phenomenon the understanding of which lies forever beyond our intellectual capacities? These are the questions that I will pursue below.

The overarching goal of this introduction is to provide a *field guide* (with a particular perspective) for anyone interested in the history and present status of philosophical issues in the study of consciousness. Part One is a preliminary overview of the current philosophical positions in the literature, as well as a discussion of the unique difficulties inherent in the concept and nature of consciousness. Part Two is an account of the study of consciousness in the history of modern psychology. Finally, Part Three is an exposition of the mosaic of philosophical puzzles of consciousness, as well as an exploration of their interrelations.[1]

PART ONE
CONCEPTUAL FOUNDATIONS

The feeling of helplessness was terrifying. I tried to let the staff know I was conscious but I couldn't move even a finger or eyelid. It was like being held in a vice and gradually I realized that I was in a situation from which there was no way out. I began to feel that breathing was impossible, and I just resigned myself to dying.
—Patient: Male, aged fifty-four, bronchoscopy, 1978.

This testimonial was one of the many unsettling personal accounts of "becoming conscious" under general anesthesia, gathered in response to the following advertisement, which appeared in four national newspapers in Great Britain in 1984:

SURGERY: Have you ever been conscious during a surgical operation when you were supposed to be anaesthesized? A medical research team would like an account of your experiences. Write in confidence.

The goal of this advertisement was to gather firsthand accounts of gaining consciousness under general anesthesia, in order to investigate the truth of a number of patients' discomforting post-surgery reports and to provide legal guidance for the accumulating court cases.[2]

Whatever philosophical problems may be associated with the term *consciousness*, it might be thought that it would be a straightforward matter to specify an operational definition of being conscious for anesthesiologists to work with. Can consciousness simply not be detected on the basis of the patient's being alert and responsive? What is in question, after all, is neither the notoriously elusive problem of phenomenal experience nor the concept of the evanescent Humean self.

A brief look at the anesthesiology literature, brimming with terms like *real awareness* and *incipient consciousness*, quickly proves otherwise. (Cf. Rosen and Lunn 1987.) If anything, the consensus is that "with the spectral edge of the EEG [electroencephalogram] or median frequency, or any other processed EEG signal, there does not seem to be a clear cut-off, without overlap, between consciousness and unconsciousness" (Vickers 1987, p. 182). The phenomenon of consciousness does not have clear-cut boundaries, and its complex structure does not admit any easy formulations. (See, for instance, the *Roche Handbook of Differential Diagnosis* on "Coma" [1979] and on "Transient Loss of Consciousness" [1989].) Even if it is in principle possible to invent

a "consciousness monitor," a device that would detect the physical signs of the presence of consciousness in a patient, no such technology is anywhere in sight, because it is not even known what exactly is to be measured.

The root of the problem lies deeper than the inadequacy of the technology or the lack of sufficient data, however. What seems to be critically lacking is a solid theoretical framework to ground and facilitate the experimental research. For example, there is no established consensus, even in the medical field, as to what should count as the criteria of consciousness, so as to demarcate the domain of the conscious from that of the unconscious or the nonconscious. The problem with building a consciousness monitor is not confined to a lack of sufficiently fine-grained measuring instruments; it ultimately has to do with not knowing where to begin measuring and where to end up with the measured quantities.[3]

Worse, it is not clear whether everyone means the same thing by the term *consciousness*, even within the bounds of a single discipline. There is considerable variation in people's pretheoretic intuitions, for instance, regarding the kinds of creatures to which consciousness can be attributed.[4]

And in the absence of well-grounded theories, the lack of robust pretheoretical intuitions becomes even more importunate. Consider again the case of anesthesia. A person who is totally unresponsive to stimuli can, in one very important sense, be said to have lost consciousness. Nonetheless, can she still be said to be conscious in another sense—in the sense of passively experiencing the sensations caused by the stimuli, for instance? Similarly, are we to grant consciousness to a patient in a vegetative state, even when she lacks a well-functioning brain stem? Or what would justify granting consciousness to the patient if she did have a functioning brain stem that maintained the autonomic functions of her body?[5] How many senses of consciousness are there anyway, and how are we to taxonomize them?

I The Puzzle of Consciousness

These questions do not have any easy, obvious answers. Nor is there at present anything that could be regarded as a received view on problems of consciousness in the scientific and philosophical community. Furthermore, it is common to find serious doubts expressed in the literature about whether there can ever be a complete understanding of the phenomenon of consciousness. The gloomy opening lines of Thomas Nagel's famous essay, "What Is It Like to Be a Bat?" have become formative for many in thinking about consciousness: "Consciousness is what makes the mind-body problem really intractable.... Without consciousness the mind-body problem would be much less interesting. With consciousness, it seems hopeless" (Nagel 1974, pp. 165–166).

The puzzle of consciousness can be regarded in various ways, all the way from a supernatural mystery that will forever elude naturalist explanations, to a natural but extremely complicated phenomenon about which we know very little. And some of the time, the blue line that lies in between becomes very thin. There are also those who express skepticism about the existence of consciousness as a real phenomenon or about the coherence of its conceptual grounding, as well as others with a much more positive outlook, busily constructing their own accounts of consciousness to solve the puzzle. A brief look at some of the representatives of these different positions is in order.

The Mystery of Consciousness and the Explanatory Gap

In the opening pages of *Consciousness Explained*, Daniel Dennett (1991) remarks:

> Human consciousness is just about the last surviving mystery.... There have been other great mysteries: the mystery of the origin of the universe, the mystery of life and reproduction, the mystery of the design to be found in nature, the mysteries of time, space, and gravity.... We do not yet have the final answers to any of the questions of cosmology and particle physics, molecular genetics, and evolutionary theory, but we do know how to think about them. The mysteries haven't vanished, but they have been tamed.... With consciousness, however, we are still in a terrible muddle. Consciousness stands alone today as a topic that often leaves even the most sophisticated thinkers tongue-tied and confused. (pp. 21–22)

Dennett should not be taken as promoting the sense of mystery, however. After all, his book is entitled *Consciousness Explained*. Of course, it is hard to say that everyone (or even many) agrees with Dennett's conviction. In fact, the general sentiment among those who work on consciousness (including philosophers, psychologists, and neuroscientists) seems to be on the "puzzled" side.

Moreover, in the wake of a recent rise in interest in the study of consciousness, almost each appearance of consciousness as a subject matter in the popular press has been tagged with some element of mystery. For instance, Francis Crick and Christof Koch called consciousness the "most mysterious aspect of the mind-body problem" in their article that appeared in a special issue of *Scientific American* titled *Mind and Brain* (September 1992). *Discover* magazine enlisted consciousness as one of the "ten great unanswered questions of science" (November 1992), and *Omni* published a special issue on consciousness but titled it "Science and the Soul" (October 1993). (Perhaps the rather unusual title Francis Crick chose for his book that appeared shortly afterward was a response to *Omni*'s inquiry: *The Scientific Hypothesis: The Scientific Search for the Soul*.) Finally, *Time* magazine, in an issue that featured consciousness research, put the words "that evanescent thing called consciousness" on its cover (July 17, 1995).

Now, no one refers to other biological or psychological phenomena in such terms. There is never a special magazine issue that pronounces the problem of cell mutation with the question of the soul in the same breath, nor does anyone refer to language as "that evanescent thing." So it seems obvious that consciousness is perceived as special, possibly unique, and not readily amenable to ordinary scientific or philosophical explanation.

Some take this sense of mystery even further, and this attitude is not at all restricted to the popular press. It is in fact possible to find the same sentiments expressed in philosophical and scientific circles, by those whom Owen Flanagan (1991) calls the "New Mysterians." For instance, Colin McGinn (1989) finds it humanly impossible ever to understand "how technicolor phenomenology can arise from grey soggy matter," and approvingly quotes the English biologist Huxley who famously stated: "How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue is just as unaccountable as the appearance of Djin when Aladdin rubbed his lamp" (p. 349).[6]

The expression of this sort of puzzlement is hardly new. Similar perplexity has been expressed by a number of people over the years, especially since the mid-nineteenth century, with the advancement of neurology and neuropsychology and the consequently well-grounded conviction that facts about consciousness *must* have some explanatory basis in the facts about the brain. For instance, in 1874, physicist John Tyndall made the following remark: "We can trace the development of a nervous system, and correlate with it the parallel phenomena of sensation and thought. We see with undoubting certainty that they go hand in hand. But we try to soar in a vacuum the moment we seek to comprehend the connection between them. An Archimedean fulcrum is here required which the human mind cannot command; and the effort to solve the problem ... is like that of a man trying to lift himself by his own waistband" (p. 195).[7] In contemporary literature, Karl Popper, in a similar vein, finds "the emergence of full consciousness ... which seems to be linked to the human brain ... one of the greatest

miracles" (Popper and Eccles 1993, p. 129). And most recently, McGinn (1989) delivers what he considers to be the final verdict on the mind-body problem: "We have been trying for a long time to solve the mind-body problem. It has stubbornly resisted our best efforts. The mystery persists. I think the time has come to admit candidly that we cannot solve the mystery" (p. 349).

There is, however, more than one way to read assertions about the mystery of consciousness. Accordingly, it is important not to lump together everyone who expresses puzzlement about consciousness into the same category. In particular, it is important to pay attention to the following two questions: Is the mystery essentially a result of a commitment to a materialist framework? Is the mystery essentially inherent in our (lack of) cognitive capacities?

As such, these questions constitute an ontological and an epistemic axis, respectively, that cross-cut each other. Not every combination in the matrix receives equal philosophical attention. A negative answer to both questions essentially leaves one out of the circle of those who find something mysterious in consciousness, and not many defend a view that find consciousness mysterious in both aspects. Rather, the focus is on views based on an exclusively positive answer to one or the other question.

Those who think that consciousness will remain a mystery in a materialist ontology suggest that the proper place to pursue investigation is instead an immaterial realm—such as the realm of the *res cogitans* for Descartes, or *World 2* of mental entities for Karl Popper. This move brings with it a problem perhaps larger than that it was presumed to solve: how to account for the link between consciousness in the immaterial realm and brains (and bodies) in the material realm. Descartes's notorious solution was to postulate the pineal gland as the gatekeeper of interaction between the two essentially different kinds of substances. Alternatively, Leibniz chose to rely on divine intervention to secure a "pre-established harmony" between the events of the two realms.

More than three centuries after Descartes, John Eccles (1991, pp. 190–191) makes a repeat attempt, although in neurologically sophisticated dress, by postulating *psychons* (mental units) as counterparts of dendrites in brains, to connect Popper's mental *World 2* to physical *World 1*. Another contemporary expression of the view that defends an antimaterialist framework for consciousness is given by Robert Adams (1987), who finds theism theoretically advantageous to materialism in explaining the relation of consciousness to bodily physical states.[8]

On the other hand are those who do not cut the bill of the mystery of consciousness to the presumed immaterial ontology of consciousness but rather to our lack of cognitive capacities that would enable us to understand the nature of the "psychophysical link" between brains and minds. Tyndall seems to be in this group, and so is McGinn, who states that although "we know that brains are the *de facto* causal basis of consciousness," we have no idea about how "the water of the physical brain is turned into the wine of consciousness" (McGinn 1989, p. 349). This is an epistemic rather than ontological problem. For McGinn, "there is, in reality, nothing mysterious about how the brain generates consciousness," but we human beings are forever "cognitively closed" to understanding the nature of this process, much the same way the understanding of quantum mechanics lies beyond the cognitive capacities of monkeys.[9]

The general difficulty involving consciousness forms a basis for what Joseph Levine (1983, 1993) called the "problem of the explanatory gap." Almost everyone agrees that there is indeed some explanatory gap in this area; what is controversial is, as I will note later, whether there is just an epistemic or also an ontological lesson that needs to be drawn from it.

In any case, these positions constitute only a fraction of the whole dialectical space. There are also the skeptics and the naturalists, and to them I now turn.

Skepticism About Consciousness

The skeptics among philosophers fundamentally doubt the coherence of the very concept of consciousness, and the merits of consciousness itself as a phenomenon fit for scientific or philosophical investigation. Patricia Churchland, in one of her early papers, compares the concept of consciousness (under a certain reading that she explicates) to such now-defunct concepts as ether, phlogiston, and demonic possession—concepts that "under the suasion of a variety of empirical-cum-theoretical forces ... lose their integrity and fall apart" (Churchland 1983, p. 80).

In a similar vein, Kathleen Wilkes claims not only that "science can dispense with the concept of consciousness and lose thereby none of its comprehensiveness and explanatory power," but "so too could ordinary language." She then suggests that "perhaps 'conscious' is best seen as a sort of dummy-term like 'thing', useful [only] for the flexibility that is assured by its lack of specific content" (Wilkes 1984, pp. 241–242). Along the same lines Georges Rey goes a step further and suggests that there are "reasons for doubting that oneself is conscious and ... thinking that nothing is conscious." Consciousness, Rey suggests, "may be no more real than the simple soul exorcised by Hume" (Rey 1988, p. 6).

Notice that this sort of skepticism about consciousness is a very different attitude from any form of "mysterianism." In particular, those who think that consciousness is mysterious are committed to the existence of some significant phenomenon, however elusive it may be in relation to scientific investigation or philosophical analysis. Consciousness skepticism, on the other hand, embraces an eliminativist stance: the concept of consciousness is defunct, and the phenomenon itself may actually be inexistent, at least so far as it is construed in the literature that the skeptics are attacking.

The Consciousness Naturalists

Finally, there are those who believe that consciousness is a real and perfectly natural phenomenon and that there will remain no mysterious unexplained residue about consciousness in a completely naturalist, but surely more advanced and mature, theoretical framework. Among the naturalists, however, there is a wide spectrum of positions representing different levels of confidence in the success of a naturalist program. For instance, one can straightforwardly distinguish between full-blown naturalists and naturalists-at-heart.[10]

Into the first group fall a number of philosophers who have explicitly defended a naturalist framework to explain consciousness, without theoretical reservations. However, some of the full naturalists, most significant among them Paul Churchland (1988) and Daniel Dennett (1991), have been charged with trying to do away with consciousness for the sake of explaining it. Some others have been more careful not to fall under this decree. For instance Owen Flanagan (1992), who proposes what he calls the *natural method*—a triangulated approach for studying consciousness that combines phenomenology, psychology, and neuroscience—states: "Consciousness exists, and it would be a mistake to eliminate talk of it because it names such a multiplicity of things. The right attitude is to deliver the concept from its ghostly past and provide it with a credible naturalistic analysis.... It will be our proudest achievement if we can demystify consciousness" (Flanagan 1995, p. 20).

Among the naturalists who give accounts of consciousness in terms of causal and functional roles (broadly construed) are Armstrong (1980b, 1993), Lewis (1966, 1972, 1980, 1995), Shoemaker (1975, 1991, 1994), Lycan (1987, 1997b), Van Gulick (1988, 1989, 1993), and Rosenthal (1986, 1997).

Most recently, Fred Dretske (1995) and Michael Tye (1995) came up with naturalist accounts that explain consciousness in entirely

representational terms. Finally, one can add John Searle to this group, who dubs his view "biological naturalism":

> The "mystery" of consciousness today is in roughly the same shape that the mystery of life was before the development of molecular biology or the mystery of electromagnetism was before Clerk-Maxwell equations. It seems mysterious because we do not know how the system of neurophysiology/consciousness works, and an adequate knowledge of how it works would remove the mystery.... [T]here has been no question of "naturalizing consciousness"; it is already completely natural. (Searle 1992, p. 102, 93, respectively)

Despite Searle's fully naturalist convictions, his views have, as I will point out later, significant disagreements with those of the above on many other (relevant) points. Nonetheless, his position also differs from a position that I call naturalism-at-heart.

Naturalists-at-heart are those who openly feel the pull of naturalism, while not quite being able to find a satisfactory place for consciousness in a naturalist framework. For example, Levine (1983, 1993), who thinks that the problem of explanatory gap poses ultimately no ontological problems for materialism, nonetheless expresses the following troubled sentiments: "The absent and inverted qualia hypotheses are thought experiments which give concrete expression to what I will call, following the Churchlands, the 'proqualia' intuition. This is the intuition that there is something special about conscious mental life that makes it inexplicable within the theoretical framework of functionalism, and, more generally, materialism" (Levine 1988, p. 272). Short of disbelieving materialism, Levine finds it difficult to place consciousness in a naturalist framework.

Perhaps the most eloquent proponent of this position is Thomas Nagel, possibly more so in his earlier writings than his later work. In one of his earlier works, Nagel examines and rejects "the reasons for believing that physicalism cannot possibly be true" and concludes: "My attitude toward physicalism [is that it] repels me although I am persuaded of its truth" (Nagel 1965, p. 110). Later he moves to a position he calls "dual aspect theory"—a position that lies between asserting falsity of physicalism but remaining short of postulating nonphysical substances for accounting for the ontology of the mind—while admitting that "to talk about a dual aspect theory is largely hand waving" (Nagel 1986, pp. 29–30). So perhaps the early Nagel was a naturalist-at-heart, and now it is more accurate to characterize him as a half-hearted-naturalist. In either case, Nagel's position seems, at least fundamentally, somewhere in the naturalist camp, despite the skeptical and pessimistic undertones that make him sometimes look closer to the mysterians or skeptics.

To recapitulate: It is noteworthy that the spectrum of disagreements ranges over not only particular accounts of consciousness but, more fundamentally, whether any satisfactory naturalistic explanation of consciousness can *in principle* be given. Part of this disagreement owes, no doubt, to the difficulty in the *nature* of the phenomenon of consciousness. But there is also a part that stems from a conceptual disarray surrounding the *notion* of consciousness. It is thus instructive to examine these two dimensions that contribute to the puzzle of consciousness separately.

II Approaching Consciousness: A Multitude of Difficulties

> What we are when we are awake, as contrasted with what we are when we sink into a profound and perfectly dreamless sleep or receive an overpowering blow upon the head—*that it is to be conscious*. What we are less and less, as we sink gradually down into dreamless sleep, or as we swoon slowly away: and what we are more and more, as the noise of the crowd outside tardily arouses us from our after-dinner nap, or as we come out of the midnight darkness of the typhoid fever crisis—*that* it is to become conscious.

This is how George Trumbull Ladd (1909), noted American psychologist (whose definition of psychology as the "description and explanation

of states of consciousness *as such*" was adopted and promoted by William James), characterized the phenomenon of consciousness (p. 30). This characterization seems straightforward, commonsensical, and familiar to everyone. So familiar that perhaps, as George Stout (1899), another psychologist of the same era, declares in the opening pages of his *Manual of Psychology*, no precise definition is necessary, or even possible: "What *is* consciousness? Properly speaking, definition is impossible. Everybody knows what consciousness is because everybody is conscious" (p. 7). Similarly, William James never attempts to give a definition of consciousness anywhere in his two-volume work, *Principles of Psychology* (1950a, 1950b). This is not because James had no interest in, or nothing to say on, consciousness; on the contrary, many of the chapters in his two volumes are *about* consciousness—its underpinnings in the nervous system, its function in evolution, its streamlike phenomenology, and so on. Rather, according to James, consciousness was a phenomenon *too familiar* to be given a definition. James (1950a) was convinced that everyone took themselves to be possessors of conscious states that were accessible by introspection, and he regarded this belief as "the most fundamental of all the postulates of Psychology." In his refusal to discuss this postulate any further, James adds that he would "discard all curious inquiries about its certainty as too metaphysical" for the scope of his book (p. 185).

This somewhat peculiar "all-too-familiar a phenomenon" attitude toward consciousness has indeed been quite common among many other prominent investigators of consciousness. Sigmund Freud (1964), for instance, supports Stout's and James's convictions in his introductory lectures on psychoanalysis: "What is meant by consciousness we need not discuss; it is beyond all doubt" (p. 70). Closer to our times, neuroscientist Francis Crick and Christof Koch (1990) endorse the same line in the opening paragraphs of their article, "Towards a Neurobiological Theory of Consciousness," even though they are attempting to lay out the foundations of a new theory of consciousness: they need not provide a precise definition of consciousness since "everyone has a rough idea of what is meant by consciousness" (p. 263).

On the other hand, it is not uncommon to come across statements about consciousness that convey a conviction opposite to those mentioned above: that not only is there no clear and generally accepted definition, but we are not even in possession of a stable *pretheoretical* conception of consciousness. And this view, too, has been around for quite a while, as expressed by Edward Titchener (1915) who cites two British psychologists, Alexander Bain and James Ward, of the late nineteenth century: "'Consciousness,' says Professor Ward, 'is the vaguest, most protean, and most treacherous of psychological terms'; and Bain, writing in 1880; distinguished no less than thirteen meanings of the word; he could find more today" (pp. 323–324).

Unfortunately, there is no hope of receiving help from antonyms, either. Here is the entry for "unconscious" in a psychology dictionary: "It is said that there are no less than 39 distinct meanings of 'unconscious'; it is certain that no author limits himself consistently to one. And nearly all meanings are closely linked to debatable theories. Any user of the term therefore risks suggesting agreement with theories he may deplore" (English and English 1958). Finally, Julian Jaynes (1976) rhetorically asks: "This consciousness that is myself of selves, that is everything, and yet nothing at all—what is it?" (p. 1).

Moreover, not only is there no consensus on what the term *consciousness* denotes, but neither is it immediately clear if there actually is a single, well-defined "*the* problem of consciousness" within disciplinary (let alone across disciplinary) boundaries. Perhaps the trouble lies not so much in the ill definition of the question, but in the fact that what passes under the term *consciousness* as an all too familiar, single, unified notion may be a tangled amalgam of several different concepts, each inflicted with its own separate problems.

What exactly, for example, is the problem of consciousness in philosophy, in psychology, and in the neurosciences? Are philosophers concerned with the same problem, or set of problems, as psychologists and neuroscientists who work on consciousness? Whereas Thomas Natsoulas (1992), a psychologist, questions: "Is consciousness what psychologists actually examine?" (p. 363), Kathleen Wilkes (1988), a philosopher, gives the following advice: "Just as psychologists do not study 'mind' *per se*, so they need not bother with consciousness [because] in all the contexts in which it tends to be deployed, the term 'conscious' and its cognates are, for *scientific* purposes, both unhelpful and unnecessary" (pp. 38–9).

An even more pessimistic view is enunciated by Stuart Sutherland (1995) under the entry "consciousness" in his *International Dictionary of Psychology*:

> CONSCIOUSNESS: The having of perceptions, thoughts, and feelings; awareness. The term is impossible to define except in terms that are unintelligible without a grasp of what consciousness means. . . . Consciousness is a fascinating but elusive phenomenon: it is impossible to specify what it is, what it does, or why it evolved. Nothing worth reading has been written about it. (p. 95)

Why are there such glaring polarities? Why is consciousness characterized both as a phenomenon too familiar to require further explanation, *as well as* a source of obscurity that remains typically recalcitrant to systematic investigation by those who work largely within the same paradigm? There is something uniquely peculiar here. Is it the phenomenon of consciousness that is more puzzling, one sometimes wonders, or the magnitude of the puzzlement itself and the theoretical dissonance surrounding consciousness? Could it perhaps be as R. J. Joynt (1981) predicted: "Consciousness is like the Trinity; if it is explained so that you understand it, it hasn't been explained correctly" (p. 108)?

George Miller (1962), faced with these difficulties, tentatively entertains the interesting proposition that "maybe we should ban the word for a decade or two until we can develop more precise terms for the several uses which 'consciousness' now obscures." Nonetheless, he ultimately decides against it: "Despite all its faults, however, the term would be sorely missed; it refers to something immediately obvious and familiar to anyone capable of understanding a ban against it" (p. 25).[11]

In the end, I find myself in agreement with Miller's positive conclusion. It is historically accurate to note that consciousness as a phenomenon in need of not only explanation, but also definition, has persistently kept resurfacing. It also seems reasonable to think that further attempts to provide carefully constructed conceptual tools could only help the situation, providing a common platform of interaction among all who choose consciousness as their object of study. This is not to say that it is not crucial to proceed cautiously to steer clear of conceptual dead ends, as well as to make sure that one does not fall into the trap of reinventing the wheel of consciousness over and over. After all, if we hope that anything toward a better understanding of consciousness will come out of the joint efforts of different disciplines, it is of utmost importance to minimize crosstalk and make sure that common terms actually point to the same referents. As a result, it seems even more imperative to look for and try to delineate the specific *conceptions* and *aspects* of consciousness under which the different problems arise. The next section is a brief attempt to address conceptual issues along these lines, before turning to the epistemological and ontological difficulties that arise from the nature of consciousness.

Difficulties with the Concept of Consciousness

Because *consciousness* is a word whose semantics has shifted over time, a brief lexical and etymological exposition may be of some service as a preliminary step. Let us start with present definitions. The *Oxford English Dictionary* (*OED*: unabridged second edition), for example, gives eight definitions of "consciousness," and twelve definitions of "conscious." For the purposes of this chapter, the eight *OED* entries for "consciousness" can be divided roughly into two groups. On the one hand, there is a largely social aspect of the term *consciousness*: joint or mutual knowledge shared by a community of people. This was indeed the earliest sense of consciousness, derived from the Latin term *conscius*.[12] It is this sense of consciousness that is used in talking about "class consciousness" in Marxist thought, or that appears in titles like Gerda Lerner's recent book, *The Creation of Feminist Consciousness* (1993).

On the other hand, *consciousness* has a largely psychological (mental) sense which relates to individuals, rather than groups, with no particularly ethical or political overtones.[13] This sense of consciousness, too, can be subdivided into two meanings—it means either "the state or faculty of being conscious, as a condition and concomitant of all thought, feeling, and volition" or "the state of being conscious, regarded as the normal condition of healthy waking life" (*OED*). The former sense is in accord with Descartes's usage of "consciousness," but it is more closely associated with, if it does not originate from, Locke's *An Essay Concerning Human Understanding*. It is also this sense of consciousness that bears an intentional component, inherent in the intentional states it subserves, leading to the *transitive* usage: "consciousness of." The latter is the *intransitive* usage of "consciousness," something more basic than, or perhaps a necessary constituent of the former, unless one wants to insist that all consciousness is consciousness of. The distinction between the transitive and the intransitive senses of the psychological concept of consciousness is alluded to numerous times in contemporary philosophy and psychology, and I will accept it here as such.[14]

Among the contemporary analytic philosophers, David Rosenthal (1997) has written substantially about the distinction between the transitive and intransitive senses of consciousness. He calls these two senses "creature consciousness" and "state consciousness," respectively. In taking consciousness as a property of organisms one can talk about a person's being conscious in the sense of being awake and alert, as opposed to being in a transient state of no consciousness, in a deep coma, or in non-REM sleep. In another vein, one can ask of a bat, or a spider, or a stickleback, or perhaps a robot if it is conscious. All of this has to do with *creature consciousness*. In addition, it makes sense to talk about whether a particular mental state is conscious. This is not quite the same as someone's being conscious. The creature sense of consciousness denotes an overall state one is in; the other classifies one's (mental) states as of one type or another. Further, a state that is not conscious can be among those that in principle cannot become conscious (e.g., certain computational states postulated by cognitive psychology), or those that can be made conscious only by such specific methods as Freudian psychoanalysis. In any case, this sense of consciousness which functions as a type-identifier for mental states is what I have in mind by *state consciousness*.[15] (Another important distinction is that of characterizing creature or state consciousness in causal versus phenomenal terms. This distinction will be introduced in the next section, and used recurrently throughout the chapter.)

In the rest of this discussion, I will largely put aside the social conception of consciousness and proceed with the psychological conception. This move approximately halves the size of the literature to be examined; yet there is still plenty that needs teasing apart and sorting out.[16]

Difficulties with the Nature of Consciousness

It is often remarked that conscious experiences are "immediately familiar" to any subject of such experiences. It is also argued that they are "so immediately close" that it is at times difficult, if not impossible, to separate their *appearance* from

their *reality*. This ambiguity between the appearance versus the reality of consciousness is also regarded as unexampled; presumably nothing else in the world suffers from it. As such, it tends to uniquely blur the line between the epistemology and ontology of consciousness: if all there is to (the reality of) conscious states is their appearing in a certain way to subjects, and if they have no existential status independent of their so appearing, the ontology of consciousness seems to collapse into its epistemology.[17]

Ironically, on the other hand, the problem appears to be just the opposite from a different angle: the appearance of consciousness seems so different from its "physical reality" that a comprehensive theory that bridges this gap is regarded as a near impossibility. In Sellarsian terms, the scientific and manifest images of consciousness are considered to embody a theoretical gap perhaps greater than in any other subject matter. (Cf. Sellars 1991.)

The view that consciousness (or, in general, the mind) and its physical basis (or, in general, the body) seem essentially so different from one another that they must have distinct existences is based on a deep-rooted idea in the history of philosophy. This idea and its variants were constitutive of arguments for the metaphysical independence of mind and body throughout early modern philosophy of the seventeenth and eighteenth centuries, perhaps most notably exemplified in the work of Descartes. The essential and complete nature of mind, generally speaking, seems to consist solely in thinking, and, as such, it must be unextended, simple (with no parts), and essentially different from the body, and therefore immaterial. This was Descartes's argument in a nutshell, ultimately drawing a strong ontological conclusion (regarding the distinctness of mind and body) from a starting point constituted by epistemic considerations (regarding the distinctness of their appearances). As Ben Mijuskovic (1974) observes, in this type of argumentation, "the sword that severs the Gordian knot is the principle that what is conceptually distinct is ontologically separable and therefore independent" (p. 123).

Mijuskovic, in locating this form of reasoning in its historical context, also notes the presence of the converse of its inference: "If one begins with the notion, explicit or implicit, that thoughts or minds are simple, unextended, indivisible, then it seems to be an inevitable step before thinkers connect the principle of an unextended, immaterial soul with the impossibility of any knowledge of an extended, material, external world" and consequently, of the nature of the relation between them (p. 121). That is, this time an epistemological conclusion (regarding an epistemic gap between mind and body) is reached from a starting point constituted by ontological considerations (regarding the distinctness of their natures).

The difficulties inherent in the nature of consciousness constitute many of the philosophical problems that will be discussed in depth in the rest of this chapter. In particular, the nature and validity of inferences between matters epistemological and matters ontological, especially those that go from the former to the latter, will continually appear as a leitmotif. Of course, some of these difficulties may be overcome quite rapidly as the study of consciousness advances; others may prove more obstinate. But it may also turn out that certain problems that seemed unsolvable had appeared that way because of the specific ways in which they were formulated, or the implicit assumptions they rested on. This possibility, too, will emerge as a relevant concern later in the discussion.

The best way to gain insight into this sort of a difficulty may very well be through locating the analysis in a broad historical perspective. The long history of consciousness research no doubt contains hints that can be parlayed not only to draw methodological lessons for further study, but also to reveal the constituent fibers of the past paradigms which couched persistent core problems under different guises over and over. It is in fact the most striking feature of the consciousness

literature, as I try to exemplify numerous times in this exposition, that the very same problems, analyses, and suggested solutions repeatedly appear, and the very same theoretical moves repeatedly get introduced at different times in the history of philosophy and psychology with little (if any) acknowledgement of past attempts and failures. In this regard, a historical approach that exposes the misleading implicit assumptions common to the past failures should prove useful in illustrating, at minimum, which steps *not* to take in approaching consciousness at present.

III Looking Ahead: The Two Faces of Consciousness

Before proceeding further, I will offer a brief first pass at a diagnosis that will be made at the end of this chapter: a principal reason underlying the confusion and seeming mystery surrounding the concept and phenomenon of consciousness lies in the presence of two influential, equally attractive, pretheoretic characterizations. These two characterizations not only shape the methods with which consciousness is studied, but more fundamentally, shape the way the problems to be studied are defined and delineated. They can be summarized in the following mottos: "Consciousness is as consciousness *does*" versus "Consciousness is as consciousness *seems*." The former is the *causal characterization*: it takes the causal role consciousness plays in the general economy of our mental lives as basic. The latter, in contrast, is the *phenomenal characterization*: it takes as fundamental the way our mental lives seem (or "feel," for lack of a better term) to us—that is, the phenomenal qualities that we typically associate with our perceptions, pains, tickles, and other mental states.

Most of the time, these two characterizations are taken to be mutually exclusive for explanatory purposes, to the extent that accounts of consciousness built around one characterization are typically accused of failing to capture the other. I believe that this undesirable consequence, which often seems to deadlock debates on consciousness, stems from a fundamental and ultimately misleading intuition that I will call the *segregationist intuition*: if the characterization of consciousness is causal, then it has to be *essentially* nonphenomenal, and if it is phenomenal, then it is *essentially* noncausal. (I call this formulation an "intuition" rather than a "thesis" due to its widely diffused, often implicit and unarticulated, but highly influential nature.)

In contrast with the segregationist intuition is what I call the *integrationist intuition*: what consciousness does, *qua* consciousness, cannot be characterized in the absence of how consciousness seems, but more importantly, that how consciousness seems cannot be conceptualized in the absence of what consciousness does. This counter intuition underwrites the project of trying to dissolve the stalemate between accounts of consciousness respectively based on the causal and the phenomenal characterizations, and marry them into a single unified account.

This introduction is not an attempt towards accomplishing such a project. Nevertheless, while presenting a conceptual mapping of the territory and locating in it contemporary problems and debates that center around consciousness, I hope to provide support for the integrationist intuition that motivates it. In doing so, I also aim to substantiate an *antiskeptical* position with respect to consciousness: there is a a deep-rooted and continuous theoretical thread connecting a set of recurrent problems in the history of philosophy and psychology typically associated with consciousness, indicating the presence of a persistent, significant, and challenging object of study.

PART TWO
A BRIEF HISTORY OF CONSCIOUSNESS: PHILOSOPHY AND PSYCHOLOGY

There is possibly no other subject matter in the history of philosophy and science with as fascinating a historical record as consciousness. Even

within the past one hundred years, consciousness has more than once been crowned as the most significant aspect of human mentality, to be followed by periods of scapegoat treatment for the failures of philosophy and science (in particular, psychology) to give a satisfactory account of the mind. In either case, consciousness was hardly ever ignored. Explicitly or implicitly, it was an ever-present concern for everyone thinking about the human mind.

Following is a brief journey through the historical path that consciousness research has traversed in approximately the last hundred years, in particular via the schools of introspectionism, behaviorism, and cognitivism in psychology, with early modern philosophy taken as a starting point.

IV Consciousness in Early Modern Philosophy

In accord with the fact that the origins of the word *consciousness* go back to early modern philosophy, it is generally agreed that Descartes gave the mind-body problem its modern formulation. Descartes's own account respectively characterized mind and body as thinking versus extended substances, and postulated that the nature of their relation was that of interaction between the *res cogitans* and the *res extensa*. But how did his notion of consciousness compare with his notion of mind? And to what extent does his notion of consciousness capture the notion that presently figures in contemporary debates?[18]

Descartes claimed that consciousness was an essential component of everything that was mental, and by "consciousness" he meant something akin to one's awareness of one's own mental states: "As to the fact that there can be nothing in the mind, in so far as it is a thinking thing, of which it is not aware, this seems to me to be self-evident. For there is nothing that we can understand to be in the mind, regarded in this way, that is not a thought or dependent on a thought" (Descartes 1993b, p. 171, fourth set of replies to Arnauld). By "thought" Descartes must have had in mind something very similar to one of the contemporary usages of "consciousness," or "awareness," especially given his definition in the *Principles of Philosophy*: "By the term 'thought' I understand everything which we are aware of as happening within us, in so far as we have awareness of it" (Descartes 1992, p. 174).

Locke, coming after Descartes's rationalism from a distinctively empiricist tradition, was nonetheless largely in agreement with his predecessor with respect to the nature of the relation between what was mental and what was conscious: they were conceptually tied. In Locke's words, "thinking consists in being conscious that one thinks," and "the idea of thinking in the absence of consciousness is as unintelligible as the idea of a body which is extended without having parts" (Locke 1959, bk. 2, chap. 1, p. 138).

There is also another sense in which Descartes and Locke seem to be in agreement: the idea of construing consciousness in roughly something like higher-order awareness. For Descartes, proper sensations in adults exist only insofar as they are accompanied by a second-order reflective awareness: "When an adult feels something, and simultaneously perceives that he has not felt it before, I call this second perception *reflection*, and attribute it to the intellect alone, in spite of its being so linked to sensation that the two occur together and appear to be indistinguishable from each other" (Descartes 1991, p. 357: letter to Arnauld, 29 July 1648, AT V, 221). Along this theoretical line, Descartes concludes, for instance, that "pain exists only in the understanding" (Descartes 1991, p. 148: letter to Mersenne, 11 June 1640, AT III, 85).

In a somewhat similar vein, Locke famously stated that "consciousness is the perception of what passes in a man's own mind" (Locke 1959, bk. 2, chap. 1, §19, p. 138). However, it may be unfair to read too much into the "higher-order awareness" construal and make Descartes's and Locke's views seem more similar than they actually are. For instance, it is not altogether obvious that Locke's second-order "perception" is as cognitively loaded as Descartes's "reflective perception," although they seem to serve the same purpose in being responsible for consciousness of "first-order" mental goings on.[19] There are also contemporaries of Descartes and Locke, who located the epistemic locus of mind in qualitative conscious states rather than thoughts or reflective perceptions. Most notably, Malebranche holds this view.[20]

This brief characterization of the early modern philosophical thought on consciousness no doubt fails to do justice to the subtleties involved. But for the sake of finding a starting point common to both philosophy and "scientific psychology" in the study of consciousness, and tracing the issues in double-track to the contemporary debates, I will leap ahead to the late nineteenth century and, skipping over the problem of unity of consciousness and Kant's treatment of "unity of apperception," continue with the work of William James.

V The Last Hundred Years: William James's Puzzle

William James may be the philosopher and psychologist who thought and wrote more about consciousness than anyone else in history. Interestingly enough, the record of his stance(s) toward consciousness is also the most curious one. James allots a great deal of space to discussing the neural underpinnings, the evolutionary function, and the phenomenal nature of consciousness in his monumental work, *Principles of Psychology*. According to James of this book, consciousness is the starting place of all psychology, the most crucial aspect of human mentality. In a chapter on the methodology of psychology, he states: "*Introspective Observation is what we have to rely on first and foremost and always.* The word introspection need hardly be defined—it means, of course, the looking into our own minds and reporting what we there discover. *Every one agrees that we there discover states of consciousness*" (James 1950a, p. 185; originally published in 1890). But only fourteen years later, James would bitterly denounce consciousness in an article titled, "Does Consciousness Exist?" with the following verdict:

> For twenty years past I have mistrusted "consciousness" as an entity; for seven or eight years past I have suggested its non-existence to my students, and tried to give them its pragmatic equivalent in realities of experience. It seems to me that the hour is ripe for it to be openly and universally discarded....
>
> [Consciousness] is the name of a non-entity, and has no right place among first principles. Those who still cling to it are clinging to a mere echo, the faint rumor left behind by the disappearing "soul," upon the air of philosophy. (James 1971, p. 4; originally published in 1904)

The reasons for this remarkable change of mind may partly lie deep in James's personal history, but they also have to do with the unique place of consciousness as a subject matter in philosophy and psychology.[21] A somewhat similar, almost neurotic shift of attitude, though in a much larger scale, spanning the whole discipline of psychology and, to some extent, philosophy, occurred in a relatively short period of transition, early in this century. This transition involved the collapse of the then very established school of introspectionism and the subsequent rise of behaviorism.[22]

VI Introspectionism

Introspectionism can be regarded as the first offspring of the effort of pulling psychology apart from philosophy and establishing it as an independent, "scientific" discipline on its own. Ironically, behaviorism would later denounce introspectionism as having tangled with metaphysics and present itself as the true, alternative scientific school of psychology. In actuality, in their struggle for identity, both schools borrowed a great deal from the scientific methodology of their times, and neither one's approach was intrinsically more "scientific" than the other.

Introspectionism's fundamental assumption was that psychology was the study of the "phenomenology" of the human mind; it attempted to give a full description of the mental landscape *as it appeared to the subject*. The data points consisted of discriminations in subjects' sensations of colors, sounds, smells, and the like. In doing so, introspectionism largely modeled its methodology after the modern chemistry of the day, which was enjoying a high reputation due to its successes in having put together the atomic table. The fundamental belief underlying most of introspectionist research was that a full understanding of the mind was possible only after completing an exhaustive inventory of its "atomic units," most elemental sensory impressions one can discriminate. Introspectionism, in other words, was in the business of constructing an atomic table of the human mind.[23] (Cf. Külpe 1901; Titchener 1915.)

Giving a full inventory of anything is no easy feat, and attempting this for the totality of the human "sensory space" was a daunting task, even in the hands of scrupulous researchers and meticulously trained subjects.[24] However, the fall of introspectionism did not result from depleted patience or the lack of a sufficient number of experiments. The failure had deeper reasons, both external and internal.

The external reason for failure was the overall changing intellectual climate in Europe and the United States, especially the rising influence of positivism in all sectors of science, as well as the humanities. The general positivist attitude constituted a significant motivation for psychologists, who had been trying hard to sever their professional ties with philosophers and to move away from anything "mental" in an attempt to relocate psychology among natural sciences. Consciousness was the subject matter of no natural science, so it could not be the subject matter of psychology either.[25]

The positivist atmosphere further provided a context in which it was easier for behaviorists to make introspectionism appear as a scientifically baseless enterprise, further burdened with the metaphysically dubious cargo of consciousness—so much so that John Watson, in the opening pages of his book that served as the behaviorist manifesto, belittled introspectionists' concern with consciousness by likening it to witchcraft: "Behaviorism claims that consciousness is neither a definite nor a usable concept. The behaviorist, who has been trained always as an experimentalist, holds, further, that belief in the existence of consciousness goes back to the ancient days of superstition and magic" (Watson 1970, p. 2).

In reality, introspectionism was as much an attempt to bring psychology up to par with natural sciences—to make it a "science of the mental," with ideas and methods inspired largely by chemistry. Although the founding adherents of behaviorism wholeheartedly denied any intellectual debt to their predecessors and did a good job of making themselves appear completely detached from their past, they were very similar to the introspectionists in aspiration and professional policy with regard to methodology.[26]

In any case, it would not be fair to place on solely external causes all responsibility for introspectionism's formidable downfall, which brought with it the downfall and disgrace of both consciousness (as a subject matter for research) and introspection (as a method for studying the mind) for several decades. There were serious reasons internal to the paradigm as well. Most important was the apparently irreconcilable conflict between results coming out of different laboratories. The most significant polarity was constituted by two main streams of research pursued in two different continents: the Würzburg school, represented by Külpe and his students in Leipzig, versus the Cornell school, lead by Titchener and his associates in Ithaca, New York. For instance, Titchener's laboratory reported that they discovered a total of "more than 44,435" discriminably different sensations, largely consisting of visual and auditory elements. In contrast, Külpe's published results pointed to a total of fewer than 12,000 (Boring 1942, p. 10). Who was telling the truth?

Conflicting results are no surprise in any experimental discipline. What led introspectionism to a dead end was an additional methodological shortcoming: the lack of a generally agreed-upon method of falsifying any of the results. The nature of introspective reports constituting the core of the data in the introspectionist paradigm was colored by the subjects' previous training. Titchenerian introspectionists were very careful not to work with "naive subjects," enforcing strict procedures to avoid "stimulus errors," but this policy worked against them in the end. Although a rigorous and careful training program and meticulous repetition of the experiments provided an acceptable degree of statistical consistency within individual laboratories, the results across different laboratories were sometimes highly contradictory. Unfortunately, when individual "seemings" (of colors, sounds, tactile sensations) were what counted as the sole data, each "phenomenal" report had to be taken at face value. And within such a framework, the degree and nature of previous training, which apparently was not standardized, made all the difference.

Edwin Boring recounts an anecdote in which Titchener and Edwin Holt debated, in front of an audience of other psychologists, whether green was an "atomic color" or a combination of blue and yellow. Each side insisted on his own judgment, and there was no means to settle the issue. One of the most serious conflicts, somewhat similar to this but larger in scope, involved a staunch disagreement between the followers of Titchener and those of Külpe on the existence of "imageless thought." Titchener was convinced that all conscious thought involved some form of imagery, at least some sensory elements. However, subjects from Külpe's laboratory came up with reports of having experienced thoughts with no associated imagery whatsoever. The debate came to a stalemate of, "You cannot experience X," of Titchenerians versus "Yes, we can!" of Külpeians, and remained the theoretical knot that it was until introspectionism, as a whole, eventually disappeared against the rising tide of behaviorism.[27]

In sum, the fact that introspectionism ultimately located the locus of authority with regard to the data in the word of the subject, while training procedures for subjects were not standardized across laboratories to immunize against "stimulus-error," brought the death sentence to the movement. When the subjects' reports showed statistical inconsistencies, the whole introspectionist community found itself up against a theoretical wall. This impasse, which surfaced as a result of several years of careful laboratory work, brought with it a sad ending to a research paradigm of hundreds of experiments and thousands of subjects.[28]

VII Behaviorism

In contrast to introspectionism, behaviorism arrived with an extremely straightforward methodology (that would ultimately cut, rather than try to untangle, the knot of consciousness), and it appeared as a fresh alternative in the troubling times of introspectionism. There was one and only one element in its research agenda: publicly observable behavior. In the natural sciences, behaviorists argued, all phenomena under scrutiny were open to third-party observation. Behavior was a perfect candidate as a subject matter of this sort. Moreover, behaviorism was able to avoid introspectionism's fatal problem of irreconcilably conflicting subjective reports by, in Boring's words, shifting "the locus of scientific responsibility from an observing subject to the experimenter who becomes the observer *of* the subject" (Boring 1953, p. 184).

Watson championed this shift of locus and the change in the subject matter of the new psychology from "facts of the internal" to "facts of the external" in a rather rallying manner in the following advice to his colleagues:

> Psychology as the behaviorist views it is a purely objective experimental branch of natural science. Its theoretical goal is the prediction and control of behavior. Introspection forms no essential part of its method nor is the scientific value of its data dependent upon the readiness with which they lend themselves to interpretation in terms of consciousness. (Watson 1913, p. 158)

> You, as a psychologist, if you are to remain scientific, must describe the behavior of man in no other terms than those you would use in describing the behavior of the ox you slaughter. (Watson 1970, p. ix)

The term *consciousness* had never figured in the vocabulary of any natural science, and it had to leave the vocabulary of the scientific psychology as well. Watson was confident that behaviorism marked the beginning of an era that was also the point of no return for consciousness:

> The time seems to have come when psychology must discard all reference to consciousness; when it need no longer delude itself into thinking that it is making mental states the object of observation.... This suggested elimination of states of consciousness as proper objects of investigation in themselves will remove the barrier from psychology which exists between it and the other sciences. (Watson 1913, p. 163, 177)

Behaviorism remained a very influential paradigm for psychology for over half a century and managed to have the words *consciousness* and *introspection* disappear from the face of the Anglo-American world.[29] There were obvious reasons for the enthusiastic acceptance of behaviorism by psychologists, motivated by its promising, "trouble-free" methodology. However, behaviorism became influential as a doctrine not only of methodology but also one of ontology. The behaviorist line turned into a fundamental belief not only that whatever psychology—*the discipline*—could study could be studied by observing behavior, but also that all there was to psychology—*the phenomenon*—was observable behavior.

This was what made (ontological) behaviorism both very strong and very weak: strong as a doctrine in its metaphysical claims and weak in grounding the strong claims it was making on what it took the world's constituents to be. The ultimate expression of the extreme view behaviorism came to hold about the ontology of consciousness is reflected in the formula Karl Lashley used to characterize "strict Behaviorism": "Consciousness is the particular laryngeal gesture we have come to use to stand for the rest" (Lashley 1923, p. 240). However, the metaphysical foundations of behaviorism, what it so passionately tried to detach itself from, turned out to be its own Achilles' heel, the cracked brick in the edifice. Even during the heyday of behaviorism, when all talk about consciousness was strictly taboo, consciousness was always present as a hidden variable in the minds and research agendas of psychologists. Boring was cognizant of this fact as early as the 1930s, when he declared: "Behaviorism owes its *ism* to consciousness. And what would it be without its *ism*? Well, it would be physiology" (Boring 1963, p. 275). Much later, Julian Jaynes would retrospectively note that "off the printed page, behaviorism was only a refusal to talk about consciousness" (Jaynes 1976, p. 15).

This make-believe attitude about the absence of *anything*, let alone consciousness, occurring somewhere between the input impinging on the subject and the subject's subsequent behavior was also precisely what provided cognitive psychology the fulcrum it needed to topple behaviorism. In Neisser's words, "the basic reason for studying cognitive processes has become clear as the reason for studying anything else: because they are there. Our knowledge of the world *must* be somehow developed from the stimulus input" (Neisser 1967, p. 5).[30]

VIII Cognitivism (and Beyond)

Ulric Neisser's *Cognitive Psychology* became a mark of a new era in psychology and proclaimed the name of the new game in its title. Neisser, in the introduction to his book, mentions the change in the intellectual atmosphere among psychologists in a wry tone: "A generation ago, a book like this one would have needed at least a chapter of self-defense against the behaviorist position. Today, happily, the climate of opinion has changed, and little or no defense is necessary. Indeed, stimulus-response theorists themselves are inventing hypothetical mechanisms with vigor and enthusiasm and only faint twinges of conscience" (Neisser 1967, p. 5).

For consciousness research, the era of cognitive psychology was marked with a few timid overtures. With the advent of cognitive psychology, whose fundamental ideas were largely inspired by computational models, consciousness found a new niche, though in terms completely foreign to its past: it became a kind of component or aspect of information-processing models. Although only a small percentage of the models developed at the time secured a role for consciousness, cognitivism brought about the first signs of the dissolution of a taboo. Nonetheless, even these cautious beginnings were not easy; consciousness would have to wait until the current ongoing ascent of neuropsychology research to come back under the spotlight. In cognitivism, cognition needed defense over behavior no more, but consciousness over cognition still did.

In this context, George Mandler's manifesto "Consciousness: Respectable, Useful, and Probably Necessary," even though it was not the first article that came out of the cognitivist literature on consciousness, and despite being written in a somewhat gingerly manner, stands out as a cornerstone.[31] Mandler opens his article with the following historical remarks: "I welcome this opportunity to act as *amicus curiae* on behalf of one of the central concepts of cognitive theory—consciousness. Another statement, however imperfect, may be useful to undo the harm that consciousness suffered during fifty years (approximately 1910 to 1960) in the oubliettes of behaviorism. It is additionally needed because so many of us have a history of collaboration with the keepers of the jail and to speak freely of the need for a concept of consciousness still ties the tongues of not a few cognitive psychologists" (Mandler 1975, p. 229). Of course, Mandler was not alone in pointing to the importance of consciousness in cognitive psychology. Tim Shallice, for instance, had observed a few years earlier that "theoretical developments in cognitive psychology and the increasing use of introspective reports require a rationale, and that this should involve consideration of consciousness" (Shallice 1972, p. 383).

Interestingly, it was the success, not the failure, of information-processing models in explaining learning, memory, problem solving, and the like—actually almost everything except consciousness—that brought some attention to consciousness itself. The fact that consciousness seemed to be the last remaining unexplained phenomenon in an otherwise successful new research paradigm helped highlight old questions about consciousness buried during the behaviorist era. Furthermore, similar developments were taking place in philosophy. Functionalist accounts, largely inspired by computational ideas, were being met with noticeable success in explaining propositional attitudes, whereas consciousness (in the sense of the subjective character of experience, or qualia) was largely being regarded as the only aspect of mind escaping the net of functionalist explanation. (See, for instance, Ned Block's influential article, "Troubles with Functionalism" 1978, as well as Block and Fodor 1980.)

Shallice was one of the first to point out the special place consciousness occupied in the problem space of cognitive psychology: "The problem of consciousness occupies an analogous position for cognitive psychology as the problem of language behavior does for behaviorism, namely, an unsolved anomaly within the domain of the approach" (Shallice 1972, p. 383). Attempts to find some role for consciousness in a cognitive economy turned up results that at times exceeded expectations. Mandler, for instance, pointed to the possibility that consciousness might be the

missing central element in a cognitivist framework, able to tie together several separate lines of cognitive research: "I hope to show that consciousness is ... probably necessary because it serves to tie together many disparate but obviously related mental concepts, including attention, perceptual elaboration, and limited capacity notions" (Mandler 1975, p. 229).[32]

Of course, there were others on whose work Mandler was basing his claim. Most notably, Norman (1968) and Atkinson and Shiffrin (1968) had used consciousness as a property demarcating processes of different kinds (conscious versus unconscious processes) in their respective unistore and multistore models of memory. Treisman (1969) and Posner and Boies (1971), among others, talked about consciousness as a limited capacity processing mechanism. Shallice's idea was to equate consciousness as selector input in his cognitive model of the dominant action system. Johnson-Laird characterized the "contents of consciousness" as the "current values of parameters governing the high-level computations of the operating system" (1983a, p. 465; 1983b). All in all, the common presupposition driving the cognitivist research on consciousness was that "the basic phenomenological concept—consciousness—can be mapped onto an information-processing concept" (Shallice 1972, p. 383).[33]

Most of these models came complete with their flowcharts, with each functionally defined element confined to its own black box and arrows indicating the direction of information flow among them. Consciousness, then, became a box among boxes—a module connected to various other modules of processing in which input was registered, intermediate results were transmitted, and output was delivered. This approach to consciousness has, according to Neisser, a special strategical advantage: "It represents a theoretical coup: not only are the facts of attention apparently explained, but psychology's most elusive target is finally nailed down to a box in a flow chart" (Neisser 1976, p. 103).

A prominent account of consciousness in recent cognitive psychology, Bernard Baars's (1988) *A Cognitive Theory of Consciousness*, is similarly given in information-theoretic terms with substantial use of functional diagrams. This trend of diagramming in cognitive psychology, inspired largely by flowcharts of computational models in computer science, also got imported into philosophy of mind by empirically minded philosophers. A primary example is Daniel Dennett's model in his "Toward a Cognitive Theory of Consciousness" (1986). Similar functional flowchart models are also being used in some of the present day neuropsychological accounts. (Cf. Schacter 1988 and Shallice 1988, especially chap. 16.)

IX The Study of the Unconscious

One important line of thought in the study of consciousness that has not yet been addressed in this chapter is the foundation of the crucial distinction between the conscious and unconscious aspects of mentality. According to Johnson-Laird, "The division between conscious and unconscious processes is the best available clue to the structure of the mind" (Johnson-Laird 1983a, p. 466). Freud would probably agree. Nonetheless, conceptions of the unconscious have changed from their Freudian origins to their cognitivist incarnations. Following is a brief historical account of the unconscious.

The Freudian Unconscious

Until the time of Freud, there was no proper theoretical framework in which to reject the Cartesian idea of equating the mind with whatever lay within the scope of one's consciousness. In other words, consciousness was generally taken to be "the point of division between mind and not mind" (Baldwin 1901, p. 216)—the mark of the mental.

The received conception of the transparency of the mind to one's consciousness, found in Descartes and Locke, was not without exceptions, however.[34] Most notably, Leibniz, in his visionary reply to Locke in *New Essays*, can be said to have anticipated some very important developments to come in psychology two centuries ahead of their time, especially those with regard to the nature and role of the unconscious: "There are a thousand indications which lead us to think that there are at every moment numberless *perceptions* in us, but without apperception and without reflection.... In a word, *insensible* [unconscious] *perceptions* are of as great use in psychology as insensible corpuscles are in physics, and it is equally as unreasonable to reject the one as the other under the pretext that they are beyond the reach of our senses" (Leibniz 1951, pp. 374–378).

Nonetheless, taking consciousness as marking the boundaries of mind by and large remained an influential maxim until the time of Freud. For instance, the entry for "consciousness" in the 1901 edition of the *Encyclopedia of Philosophy and Psychology* reads as follows: "[Consciousness] is the distinctive character of whatever may be called mental life" (Baldwin 1901, p. 216). Within this context, the introspectionist conviction of the time—that psychology is the "science of the mental"—provided an especially strong basis for rejecting the unconscious as part of the mental, and hence as a subject matter for psychology. Titchener, for example, was resistant to the idea of the unconscious, to the extent of declaring it a theoretically dangerous construct for psychology: "The subconscious may be defined as *an extension of the conscious beyond the limits of observation*.... [T]he subconscious is not a part of the subject-matter of psychology.... In the first place, *the construction of a subconscious is unnecessary*.... Secondly, *the introduction of a subconscious is dangerous*" (Titchener 1915, pp. 326–327, emphases in the original text).[35]

None of this should be taken as claiming that the concept of the unconscious as a part or aspect of the mental was completely unheard of or unacknowledged, however. In other words, Freud was not really the inventor (or discoverer) of the concept of the unconscious in any way. On the contrary, the general intellectual atmosphere of the times preceding Freud's appearance allowed talk about mental activity of various sorts that occurred without the subject's awareness, at least in any direct way. For instance, the well-known metaphor of the mind as an iceberg, consisting of consciousness as the tip above the surface and of a subsurface unconscious component, constituted by hidden currents but nonetheless effective on one's conscious mental life, was generally recognized and used.

In particular, toward the end of the nineteenth century, the idea of the unconscious mind had become operative among many scientists, philosophers, and literary scholars, in a lineage traceable back from Rousseau to Goethe, to Fichte, and to Nietzsche (Whyte 1960). Freud apparently acknowledged this, as reported by Ernest Jones, one of the most prominent Freud scholars, in the following statement he made at his seventeenth birthday celebrations: "The poets and philosophers before me discovered the unconscious. What I discovered was the scientific method by which the unconscious can be studied" (quoted in MacIntyre 1958, p. 6). There were also attempts to study the unconscious empirically. For instance, Henri Ellenberger credits Gustav Fechner, a pioneer of psychophysics research, as the first person who tried to reveal the nature of the unconscious by experimental methods, though his work did not prove fruitful (Ellenberger 1970, chap. 5).

However, none of these ideas about mental processes going on in one's mind without being conscious were well formulated: there was no coherent account to explain the structure, functional role, or operation of the unconscious, or the modality of its relation to consciousness in the general scheme of an individual's mental life. There was consensus regarding neither the nature of the unconscious, nor its place in regard to

consciousness, in the intellectual community. To this situation Freud brought a steadily evolving theoretical framework in which, for the first time, construction of hypotheses to answer each of these questions became possible. This is the sense in which Freud can be said to be the pioneer of the unconscious.[36]

In Freudian theory, the *unconscious proper* consists of repressed processes, exerting stress on the conscious component of the subject's mind and shaping his or her daily life in substantial ways. This is in contrast to the *preconscious*, which includes those processes that only contingently happen to lie outside awareness. What is preconscious can easily become conscious without special techniques or effort; what is unconscious has to be "brought to the surface" through the psychoanalytic technique with the help of an analyst.

The Freudian unconscious, although related, is not the same sense of unconsciousness employed in the current cognitive psychology research regarding unconscious processes—the "cognitive unconscious." Unconscious processes of both kinds are opaque to introspection, but there is a difference between them. The Freudian unconscious exists because of past events, explainable by repression mechanisms and the like, and is not *in principle* inaccessible. The cognitive unconscious, on the other hand, exists due to the way our perceptual-cognitive system is constituted and lies *in principle* outside our access. The mechanisms that subserve depth perception, for instance, are taken to be hard-wired: they are not there because of repression, and they can never become conscious through any method, psychoanalytic or otherwise.

The recognition and study of the cognitive unconscious goes even further back than Freud, at least to von Helmholtz's work on perceptual constancy, and spans a substantial period, all the way up to the thesis of "unconscious perceptual inference" by Rock (1983). Despite these differences, however, Freud's approach to the unconscious was very modern and in anticipation of the "cognitive revolution."

An encompassing account of the Freudian unconscious, including its structure and dynamics, is given in Erdelyi (1985). Erdelyi also makes a strong case that Freudian psychology was indeed very close, in essence, to the cognitive psychology of our day—especially in terms of its approach to understanding mental phenomena, and research methodology. He even goes on to reconstruct Freudian schemas of the structure of consciousness, quite plausibly, in modern flowchart style. Neisser (1976) also refers to Freud's diagrams depicting the structure of the tripartite division of consciousness, preconsciousness, and unconsciousness as "flowcharts" (See Freud 1950, p. 394).[37] It is true not only that Freud anticipated some of the developments in cognitive psychology but also that the Freudian unconscious, even if under different names, has played a significant role as an influential construct in cognitive psychology.

The Cognitive Unconscious

There has also been a whole research industry in contemporary cognitive psychology involved in investigating the nature of the unconscious: mental processes that underlie cognition but are themselves not conscious.[38] Over the past few decades, there has been an enormous wealth of data accumulated, operative in current psychological theory, in this area—from rules of Chomskian universal grammar, to computational mechanisms underlying vision and the $2\frac{1}{2}D$ *sketch* inspired by the work of the late David Marr (1982), and to Newell and Simon's work on cognitive constraints in planning, problem solving, and game playing (1972). Consequently, the classification of mental processes as conscious versus nonconscious is useful and not unusual (though controversial) in psychological practice, especially in research on psycholinguistics, attention, and perception.[39] Furthermore, as evidenced from contemporary psychology literature, research on type identifying mental states as conscious versus nonconscious, and research on the nature of consciousness of the subjects who have

such states is being pursued on independent conceptual grounds. In fact, the dichotomy of conscious versus nonconscious processes is not the only such ground on which current research in cognitive psychology rests. There are several other such distinctions, all overlapping in various ways in their function to distinguish mental processes that are directly available to the subject ("introspectable," reportable, etc.) and those that are opaque and unavailable, as reflected in a recent note by cognitive psychologists Holyoak and Spellman:

> Theorists of diverse persuasions have been led to propose cognitive dichotomies, which have been given a rather bewildering array of labels: unconscious vs. conscious, procedural vs. declarative, automatic vs. controlled, reflexive vs. reflective, and many others.
>
> These distinctions do not always divide cognition along the same lines ... [but] there are tantalizing similarities among the proposed dichotomies. In particular, the first member of each pair is generally viewed as involving unconscious mental processes, a topic that has seen a recent resurgence of interest among experimental psychologists. (Holyoak and Spellman 1993, p. 265)

In sum, whether or not Johnson-Laird is right in his claim about the distinction between the conscious and the unconscious being the most important theoretical tool to study the mind, one can easily say that the investigation of the unconscious in cognitive psychology has proved to be at least as fruitful as the investigation of the conscious.[40]

X Status Report: From Information Processing to Qualia

What is the current status of cognitive psychology? Information-processing models in psychology are still popular, but they do not constitute the sole dominant paradigm any more. This is also reflected in models of consciousness. But there are other reasons, too, for the shift in the research paradigm with respect to consciousness. One of them is the recognition that the functional diagrammatic depictions of consciousness seem to leave out something important: the subjective, experiential aspect of consciousness. Perhaps there is something about consciousness that makes its identification with specific modules of isolated functions fundamentally inadequate.

Interestingly, it was Neisser who registered such a concern about the information-processing models of consciousness during the heyday of cognitivism:

> The treatment of consciousness as a processing stage is unsatisfactory in a still more fundamental way. It does justice neither to the usages of the word "consciousness" in ordinary discourse nor to the subtleties of experience. A better conception of consciousness, which has been suggested many times in the history of psychology, would recognize it as an aspect of activity rather than as an independently definable mechanism.... Consciousness is an aspect of mental activity, not a switching center on the intrapsychic railway. (Neisser 1976, pp. 104–105)

Many people, including philosophers, proceeded with Neisser's intuitions in the past few decades. Something essential to (at least our commonsense conception of) consciousness; it was largely believed, was necessarily left out in characterizing consciousness only by specifying its functional role in the cognitive economy of human mentation and behavior. This something—the phenomenal face of consciousness—brings us back full circle to the problem of the two faces of consciousness.

In the last part of this chapter, I will examine the dialectic of the opposition between the segregationist and the integrationist intuitions, in the context of the causal and phenomenal characterizations of consciousness.

PART THREE
PROBLEMS OF CONSCIOUSNESS: A PERSPECTIVE ON CONTEMPORARY ISSUES AND CURRENT DEBATES

"Consciousness is a word worn smooth by a million tongues," George Miller once said. "Depending upon the figure of speech chosen it is a

state of being, a substance, a process, a place, an epiphenomenon, an emergent aspect of matter, or the only true reality" (Miller 1962, p. 25).

The conceptual and historical analyses I have presented are in agreement with Miller: it is probably best to regard and treat consciousness as a cluster concept. There are simply too many connotations that go under the term, and it seems futile to try to specify a single concept that would cover all aspects of consciousness or a single "*the* problem of consciousness." Nonetheless, I have tried to illustrate, there is a coherent theoretical thread constituted by certain problems and not others, that one can trace from texts in early modern (if not ancient Greek) philosophy to the emergence of scientific psychology in the nineteenth century, to the present.

The most troublesome feature of this thread is what has been most difficult to explain, and it is the topic I arrived at by the end of the historical analysis: the qualitative, or phenomenal aspects of consciousness, or qualia. Of course, the notion of consciousness theoretically outstrips the notion of qualia, and there are many fascinating aspects to consciousness that do not necessarily have a qualitative component (e.g., its representational aspect, its attentive and control components, and mechanisms of the unconscious). But it is also questionable whether qualitative and nonqualitative aspects of consciousness can really be understood or explained independent of one another. These are the questions I will focus on and pursue below.

XI Consciousness and Intentionality: Two Dimensions of Mind

Jerry Fodor once remarked: "There are, I think, three great metaphysical puzzles about the mind. How could anything material have conscious states? How could anything material have semantical properties? How could anything material be rational?" (Fodor 1991a, reply to Devitt, p. 285). Having enumerated these three questions, Fodor chooses to stay away from the first, despite his career-long devotion to the latter two (which he takes to be closely related). This attitude is not at all uncommon. It is generally accepted as a received view that the two fundamental aspects of mind, consciousness and intentionality, can be studied in the absence of one another—at least that intentionality can be so studied with no reference to consciousness. Here, "consciousness" typically refers to the qualitative aspects of consciousness, and "intentionality" is taken sufficiently broadly to embrace questions about semantics, as well as rationality. Fodor's justification is the following:

> It used to be universally taken for granted that the problem about consciousness and the problem about intentionality are intrinsically linked: that thought is ipso facto conscious, and that consciousness is ipso facto consciousness of some or other intentional object.... Freud changed all that. He made it seem plausible that explaining behavior might require the postulation of intentional but unconscious states. Over the last century, and most especially in Chomskian linguistics and in cognitive psychology, Freud's idea appears to have been amply vindicated.... Dividing and conquering—concentrating on intentionality and ignoring consciousness—has proved a remarkably successful research strategy so far. (Fodor 1991b, p. 12)

Not everyone agrees, however. In particular, John Searle recently argued for what he called the "connection principle": the thesis that consciousness and intentionality are immanently linked, and, contra Fodor's thesis, any research strategy that tries to explain the latter without recourse to the former is doomed to failure. Searle states the connection principle as follows: "Only a being that could have conscious intentional states could have intentional states at all, and every unconscious intentional state is at least potentially conscious.... [T]here is a conceptual connection between consciousness and intentionality that has the consequence that a complete theory of intentionality requires an account of consciousness" (Searle 1992, p. 132).[41] More recently, Strawson (1994) takes the similar position that consciousness is the only distinctive characteristic of the mind.

The beginnings of this line of thought can be traced back to Brentano's discussion of the relation between mental and physical phenomena in *Psychology from an Empirical Standpoint*. Brentano is acknowledged as being first to postulate intentionality as the mark of the mental in modern terms. For instance, he says: "Every mental phenomenon is characterized by ... the intentional inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction toward an object (which is not to be understood here as meaning a thing).... This intentional inexistence is characteristic exclusively of mental phenomena. No physical phenomenon exhibits anything like it" (Brentano 1874, pp. 88–89). And regarding consciousness, Brentano states, for instance, that "no mental phenomenon is possible without a correlative consciousness" (p. 121). Although the connotations of the terms *consciousness* and *intentionality* have somewhat shifted from Brentano's time to the present, I think it is fair to note that there is a great deal of theorizing in his work that lay the foundations of an account of the mental that attempts to incorporate these two dimensions of the mind in a principled way.

Another attempt to characterize (phenomenal) consciousness and intentionality as the two hallmarks of the mind, though for the purposes of a critique, is given by Richard Rorty (1979, p. 24) in terms of the diagram in figure I.1. There are in fact various ways to fill in such a diagram, and which elements of the mental are to occupy which cells is a matter of controversy. For instance, not everyone thinks that beliefs and desires are without a phenomenal component (Searle 1992), or that pains have no representational or intentional aspects (Dretske 1995, Tye 1995). And the adherents of panpsychism would probably maintain that the cell that holds what Rorty labels "the merely physical" is bound to remain nil.[42] Nonetheless, this way of depicting the dimensions of the mental is useful in terms of illustrating what has been the most problematic aspect of the study of consciousness. In this diagram, it is what Rorty calls "raw feels."[43]

Most of the current debates involving consciousness revolve around the (possible) inhabitants of this particular cell and their nature. Are there really such things as nonrepresentational but phenomenal properties? If there are, what is their ontological nature, what kinds of special epistemological problems do they present, and how can their semantics be given? Can they ever be captured in naturalistic explanatory scheme,

	With phenomenal properties	Without phenomenal properties
Intentional, representational	Occurrent thoughts, mental images	Beliefs, desires, intentions
Nonintentional, nonrepresentational	Raw feels (e.g., pains and what babies have when they see colored objects)	"The merely physical"

Figure I.1
Two dimensions of mind

or are they inherently bound to remain mysterious? These are the questions that constitute the consciousness debates today.

XII Perspectivity and Epistemic Asymmetry

Naturally, there can be several different entry points to the kind of exposition I aim to present here. The investigation of consciousness is as fascinating as it is difficult, and it presents unique epistemological and ontological difficulties. Although my overarching goal is to provide an overview of the contemporary problems of consciousness rather than try to present my own solutions to them, I will start by presenting a brief profile of what I take to be the primarily responsible component in the consciousness puzzle: the epistemic element of perspectivity. Perspectivity, or the fact that consciousness is a phenomenon that admits a distinction between "perspectives," or "points of view" in its explication, lies deep at the roots of the common understanding of consciousness, as well as the attitude of puzzlement. Furthermore, epistemically based theses about consciousness seem much less controversial than ontologically based theses. I start by sketching a commonsense conception of consciousness and try to reveal just how perspectivity figures in it. Then I will proceed to examine its possible ontological ramifications.

Why does consciousness keep appearing as an unsolved puzzle for philosophy, psychology, and neuroscience? There do not seem to be similar puzzles associated with the study of, say, memory or learning, or biological development and growth. What is so special about consciousness?

George Miller thinks that perhaps the unique difficulty involved in the understanding of consciousness stems from the fact that consciousness is both the phenomenon we try to investigate and the very tool we need to use to pursue this investigation. "Turning a tool on itself," he says, "may be as futile as trying to soar off the ground by a tug at one's bootstraps." He continues: "Perhaps we become confused because whenever we are thinking about consciousness, we are surrounded by it, and can only imagine what consciousness is *not*. The fish, someone has said, will be the last to discover water" (Miller 1962, p. 25).

Miller's observation is intriguing. One cannot, in principle, study the minute details of a microscope's outer surface, for instance, by using the very same microscope. This would be impossible simply because of the way the microscope, as a tool, is designed and used. Neither can one directly take the picture of a camera by using the camera itself. But why should these considerations apply to the study of consciousness using consciousness itself? One can certainly lay the body of a microscope under another microscope for examination or take pictures of one camera with another. It may be that the sort of recursive impossibility involved in the self-study of tools applies to the phenomenon of self-consciousness—for example, one's study of one's own consciousness by introspection. But Miller is concerned with the study of consciousness in general here, not only self-consciousness, and it is not clear why the analogy should hold.[44]

Nonetheless, Miller's point is related to what I see as the source of what makes consciousness puzzling. The difficulty lies in the curious duality inherent in the (epistemic) *study* of the phenomenon. This duality does not need to be inherent in the (ontological) *nature* of the phenomenon of consciousness itself or its properties. In fact, as I have mentioned, the ontology of consciousness is an issue open to current debates. But as far as the epistemology of the matter goes, there appears to be a genuine asymmetry between the *mode of access* to facts of one's own consciousness and the mode of access to facts about others' conscious states. This asymmetry is what grounds the important distinction between systematic approaches to consciousness from the *first-person perspective* versus the *third-person perspective*.

On the one hand, nothing is more intimately known by conscious human beings than the way the world (including themselves) appears to them.

We are all subjects of a variety of perceptual experiences, thoughts and ideas, pains and tickles, joys and sorrows. Under normal circumstances, there is nothing more familiar with the way the face of one's spouse looks, the way a favorite drink tastes, the way the chronic heartburn starts to make itself felt. We all have, it seems, firsthand, immediate, direct knowledge of the rich phenomenology of colors, sounds, tastes, aromas, and tactile sensations that embellish our experiences—the *qualia*. All these are constituents of a specific mode of being for every individual; they determine, in Thomas Nagel's famous phrase, *what it is like to be* that individual (Nagel 1974).

Moreover, we all seem to have a "privileged" way of knowing about our own thoughts, feelings, and sensations. Epistemological problems about knowledge notwithstanding, and even the question of the incorrigibility of the mental put aside, it seems that there is at least a special mode in which one's own experiences are present to one, in an immediate, direct way, not available for anyone else.[45]

Further, the common wisdom goes, we cannot genuinely entertain the possibility that we may be lacking consciousness; the very fact that we are questioning our own consciousness renders the possibility of our not being the entertainer of some occurrent thoughts logically contradictory. If in nothing else, Descartes was perhaps right in this regard: The mere fact of being the bearer of (these) thoughts is, in the Cartesian sense, unmistakable evidence, *for oneself*, that one is conscious. This is the characteristic of the first-person perspective; from the inside, consciousness seems all-pervasive, self-evident, and undeniable.

On the other hand, contemporary science tells us that the world is made up of nothing over and above "physical" elements, whatever their nature (waves, particles, etc.) may be. Where does this leave us with respect to the place of consciousness in an entirely physical world? "How can technicolor phenomenology arise from the soggy grey matter of brains?" as Colin McGinn asks (McGinn 1989, p. 349). Can one accommodate a respectable account of consciousness that does justice to the richness of our conscious experiences of sights and sounds within a framework based on a monistic materialist ontology? Consciousness just does not seem to be the kind of phenomenon that is amenable to the sort of scientific explanation that works so well with all other biological phenomena, such as digestion or reproduction. The facts that would settle the question of whether some organism—an animal or a fellow human being—is digesting do not seem to be available in the same way when it comes to the question of consciousness in others, especially in the case of organisms phylogenetically distant from ourselves. There seems to be no ordinary way to peek into the inner lives of others—to feel their pains, go through their sensations, or directly observe their consciousness.[46] That is, there seems to be an epistemic impossibility for anyone to have direct access to the qualia of others—literally share their first-person perspective, in short, to partake in the mode of what it is like to be them. These are the limitations of the third-person perspective: *from the outside*, firsthand exploration of the consciousness of others just seems to be out of the reach of ordinary scientific methods, others' experiences being neither directly observable nor noninferentially verifiable. And therein this asymmetry between the first- and the third-person perspectives lies the epistemic duality in the study consciousness.

But what exactly follows from this asymmetry? What do the limitations of the third-person approach entail, for example? Are there any insurmountable problems for a systematic study of consciousness—its nature, underlying mechanism, evolutionary function, ontological status? After all, the third-person perspective is what is and has successfully been operant in the scientific practice of the past several centuries, and no one doubts that it can provide valuable advances in the understanding of consciousness. But the issue is whether such an approach is always doomed to leave something essential to consciousness out of its explanatory scope. In short, is there an

unbridgeable "explanatory gap" inherent in the third-person approaches to consciousness, and if so, can it be remedied by the deployment of a crossbred conceptual scheme that embodies a first-person approach in the investigation of consciousness?

XIII First-Person versus Third-Person Approaches to Consciousness

The epistemic asymmetry inherent in the study of consciousness can be found as manifested under different names, roughly as variations of one another and as occupying critical roles in theoretical junctions. The notion of the first-person versus third-person perspectives is one such contrastive pair. Yet another similar distinction is that of the subjective versus the objective, or the "phenomenal" versus "physical."

One way of describing a particular experience, say, of tasting a particular vintage of a certain kind of wine, would be to try to state how the wine *tastes to me*—that is, what it is like for me to have that particular gustatory experience. This is indeed the ordinary, even if not so easy, way. It involves the usage of qualitative terms such as "fruity," "with a hint of tobacco," or "full-bodied," and the hope of conveying some sense of what the tasting of that wine would be like had it been experienced by the listener.

The other way would be to proceed by way of giving a description of the specific ways in which my tastebuds are excited, my olfactory nerves are activated, my blood chemistry has changed, and so forth. This would not be the most ordinary way of describing one's gustatory experiences, but perhaps one can overhear two devoted neurologists talking this way to one another at a conference reception. In any case, in the right context, it is clear that this alternative method would also be informative in conveying *something* about the nature of one's experience.

There is clearly an important difference between the two methods.[47] The first one attempts to describe an experience by stating its qualitative aspects as they seem to the experiencer. The fruity character is directly experienced *only* by the person whose gustatory and olfactory nerves are excited by the wine. As such, the experiencer has a privileged status; she gets to have the experience, whereas the listener only gets to hear the description. As far as the second method is concerned, however, the experiencer and the listener are epistemically on a par. The description of the perturbations in the experiencer's nervous system is open to public observation and verification, and ordinarily, no qualitative terms about the experience (how it feels) need to be involved. In other words, whereas the instantiation of the phenomenal properties of an experience is directly accessible only to the experiencer, the instantiation of its intrinsic neurophysiological properties can be equally observed by many.[48] In the latter case, what is at issue are the publicly observable aspects of the experience—not how it feels but what it does.

The important question is to determine whether these two methods have distinct scopes of explanation and whether they are necessarily committed to distinct ontologies. I have outlined two general approaches to consciousness, each of which respectively takes one of the two above methods as primary. The first of these approaches takes consciousness "as consciousness seems" and, in accord with the phenomenal characterization, regards its qualitative aspects as the primary components of any explanatory scheme. The second one takes consciousness "as consciousness does," and, in accord with the causal characterization, tries to account for consciousness in terms of what it does and the role it plays in one's cognitive economy. Put in different terms, one can call these the "first-person-perspectival" versus the "third-person-perspectival" approaches to consciousness.[49] The dichotomy of the "two faces of consciousness" manifests itself in yet other distinctions and under other names, to which I now turn.

XIV The Two Faces of Consciousness Revisited

The phenomenal and the causal characterizations are merely expressions of what seems most important, or primary, in the understanding of the nature of consciousness. They are not, in themselves, in opposition with one another. It is only under the dictum of the *segregationist intuition* that they are considered essentially antipodal and mutually exclusive. The issue of how to locate the phenomenal and the causal characterizations with regard to each other is central to the dialectic of certain ongoing debates surrounding the "phenomenal" versus "access" senses and the "easy" versus "hard" problems of consciousness.

By "phenomenal characterization of consciousness" I mean a characterization given fundamentally in first-person terms, describing episodes of inner life in terms of how they feel or seem to the subject who experiences them. William James was interested in both the temporal and the spatial structure of consciousness, and his chapter "Stream of Thought" (James 1950a, pp. 224–290) provides an excellent example of such a characterization.[50] Apart from James and the continental phenomenologist philosophers, introspectionist psychologists were paradigmatically interested in the phenomenal aspects of experience, and they relentlessly pursued the project of "mapping the boundaries of the inner space of consciousness." Along those lines, Titchener defined consciousness as the occurrent parts of one's mind, accessible by introspection, at any given moment: "My 'consciousness' is the sum of mental processes which make up my experience *now*; it is the mind of any given 'present' time. We might, perhaps, consider it as a cross-section of mind" (Titchener 1902, p. 13).

But there is more to the phenomenal characterization in the way "problem of phenomenal consciousness" is understood today. In particular, the problem has now transformed into the exploration of explanatory laws that would account for how particular phenomenal aspects of consciousness could arise from their physiological underpinnings.[51] Deep down, this problem is a manifestation of the gap that separates our direct understanding of consciousness in first-person terms, versus the objective, physicalist accounts of consciousness given in third-person terms. The roots of this problem are indeed unique; no other phenomenon presents us with two distinct epistemic perspectives from which it can be investigated. Given this duality, how does the "causal characterization" of consciousness fare against its phenomenal counterpart?

A causal characterization of consciousness can be given in many dimensions. One can try to account for consciousness in terms of behavioral manifestations, or of its role and place in the general mental economy. The former approach was behaviorists' failed solution to account for (or, rather, do away with) consciousness. The former became a canonical characterization in behaviorism's successor, cognitive psychology and in functionalist schools of philosophy.

Behaviorism, in its explicit form, is no longer around. But it is worth mentioning again how the most obvious difficulty in relying entirely on external criteria gave way to information-processing accounts of consciousness in particular, and of mental phenomena in general. Behaviorism left no room for the possibility of the presence of consciousness in the absence of external behavior. Put differently, the absence of evidence from the third-person perspective implied the theoretical rejection of all experience that is generally characterized in first-person terms. Given that the most familiar aspects of consciousness have to do with its phenomenology (think of James' stream), this result stood out as the most difficult one to accept. Even many behaviorists balked at biting the bullet and claiming that a person who is sitting perfectly still with no vocal cord activity whatsoever (behaviorists' characterization of "thinking") would ipso facto be unconscious.[52] This claim, it seemed, was readily refutable in one's own everyday phenomenology. There was, after all, an epistemic component to the

phenomenon of consciousness that cried for a characterization in first-person terms. However, behaviorism, in its attempt to associate consciousness and behavior conceptually, and thereby fully externalize consciousness, left no room for talking about consciousness as it is experienced by the subject. Everyone smiled at the joke about one behaviorist's asking another, after having made love, "It was great for you; how was it *for me?*" but (fortunately) not many took the scenario as a serious possibility.[53]

A logical next step in trying to account for consciousness in causal terms was to reverse the behaviorist direction and, to some extent, re-internalize the causal criteria of consciousness. This provided a groundwork for functionalist philosophy of mind, and cognitive psychology. Under such a more relaxed framework, consciousness was allowed to be individualized by the role it played, as an integral component of the larger network of mental states and processes. With the promising application of computational ideas and information-processing models in psychology, it was canonically characterized as a process accomplishing a specific task, a module with a specific function in a cognitive diagram, or an abstract property of the overall system. Here is a paradigmatic characterization in the cognitivist framework:

> Consciousness is a process in which information about multiple individual modalities of sensation and perception is combined into a unified multidimensional representation of the state of the system and its environment, and integrated with information about memories and the needs of the organism, generating emotional reactions and programs of behavior to adjust the organism to its environment.... The *content* of consciousness is the momentary constellation of these different types of information. (Thatcher and John 1977, p. 294)

What is important to note here is that this characterization is given largely in a third-person perspective. Consciousness is identified with *what it does* but not necessarily *how it feels* to the experiencing subject.

Once again, it is essential to ask at this stage whether these two perspectives are mutually exclusive regarding their explanatory roles—that is, whether an account of causal consciousness provides us with no understanding of how consciousness seems to the first-person subject, and whether an account of phenomenal consciousness has no elements that figure in the understanding of what consciousness does. As I will suggest below, my answer is "not necessarily." A further question along this line would be what ontological consequences the perspectival asymmetry in the epistemology of consciousness entails. There, my answer will be "not very much." But first let me bring into the picture another distinction, proposed by Ned Block, which aligns well with the distinction between the causal and phenomenal conceptions of consciousness.

Access Versus Phenomenal Consciousness

Block (1995) distinguishes between "access consciousness" and "phenomenal consciousness" as follows:

> *Access (A) consciousness*: A state is access-conscious if, in virtue of one's having the state, a representation of its content is (1) inferentially promiscuous, that is, poised for use as a premise in reasoning, (2) poised for rational control of action, and (3) poised for rational control of speech.... These three conditions are together sufficient, but not all necessary.
>
> *Phenomenal (P) consciousness*: P-consciousness is experience. P-conscious properties are experiential ones. P-conscious states are experiential, that is, a state is P-conscious *if* it has experiential properties. The *totality* of the experiential properties of a state are "what it is like" to have it. (pp. 230–231)

Defining A-consciousness is a straightforward matter. In the case of human beings, A-consciousness is a cognitively interwoven aspect of mental life, underlain by three crucial capacities centered around rationality: rational cogitation, speech, and action. (For the general case, not all three conditions are necessary for A-consciousness because, Block maintains, animals without speech can have mental states of the A-conscious type.) Construed as such, A-consciousness fits well in the domain of propositional attitudes in philosophy of mind, and it is just the perfect sort of subject matter for cognitive psychology.[54]

P-consciousness is more problematic. Block starts out his analysis of P-consciousness by stating the difficulty particular to it: "Let me acknowledge at the outset that I cannot define P-consciousness in any remotely noncircular way.... The best one can do for P-consciousness is ... *point* to the phenomenon" (Block 1995, p. 230). The way Block himself goes about characterizing P-consciousness is either "via rough synonyms" or by examples. P-consciousness, as expected, is what I have been referring to as the phenomenal aspect of consciousness. Among the P-conscious properties that endow a mental state with P-consciousness in virtue of its having them are, for instance, the way it feels to "see, hear, smell, taste, and have pains" and more generally, "the experiential properties of sensations, feelings, and perceptions". Furthermore, Block maintains, P-conscious properties are "distinct from any cognitive, intentional, or functional property" (p. 230).

Block thinks that it is not an embarrassment that he cannot provide a noncircular definition of P-consciousness. But why is it difficult to provide a straightforward definition of P-consciousness, and why should this not be considered a cause of disconcertment? According to Block, that there is no way to give a reductive definition of P-consciousness is not embarrassing given the "history of reductive definitions in philosophy," presumably full of failures.

It is still not clear, however, whether the inability to define P-consciousness reductively is sufficient reason to think that the only other alternative must be an ostensive definition. Even if it is, it may be useful to ask why the definition of P-consciousness is, unlike other definitions, thus obliged to ostention. In the case of A-consciousness, there seems to be no such problem. Block's definition of A-consciousness is not, strictly speaking, reductive, and it serves its purpose well with no need for ostension. Could there be something inherent in the pretheoretical construal of P-consciousness such that it does not allow a nonreductive but also nonostensive definition? More importantly, could it be that the particular way Block's distinction carves out phenomenal consciousness, separating it completely from its causal and functional aspects in accord with the "segregationist intuition," renders its investigation by means of scientific methods theoretically impossible? Put differently, could we be painting ourselves into a corner by a conceptual commitment to Block's distinction such that we end up with a number of straightforward problems about A-consciousness and a conjured-up "hard problem" of P-consciousness that in principle admits no solution?[55] This last question leads directly into a related debate that has its roots in the "explanatory gap" problem, recently dubbed by David Chalmers the "easy and hard problems of consciousness."

The "Easy Problems of Consciousness" and the "Hard Problem"

Chalmers (1995) characterizes the "easy problems" as those concerning the explanation of various cognitive functions: discriminatory abilities, reportability of mental states, the focus of attention, the control of behavior. Of course these are not trivial problems at all, and labeling them "easy problems" should not be taken as downplaying their complicated nature. Rather, Chalmers's point is that "there is no real issue about whether these phenomena can be explained scientifically." They can be. "All of them are straightforwardly vulnerable to explanation in terms of computational or neural mechanisms" (p. 201). What makes "the hard problem" of consciousness a different *kind* of problem is, Chalmers maintains, its resistance to all the methods that explain, or have the potential to explain, the rest of the problems. Put differently,

there is a different kind of problem about consciousness that may evade the successes of all standard scientific advances. Such a problem would be a hard problem indeed. What is it?

According to Chalmers, "The really hard problem is the problem of experience." More specifically, it is the "subjective" aspect of every experience that resists explanation. The notion of "subjective aspect" is given, as Block does, in Nagelian terms: There is something it is like to be a conscious organism and have experiences. In other words, "what it is like to be" constitutes the subjective character of the experiences of the organism in question. This much is also in line with Block's characterization of phenomenal consciousness.

I will give a sorted-out schema of the theoretically interwoven notions of "phenomenal aspect," "subjective character," and "what it is like to be" in section XVI below. For now, it is useful to observe that the line that separates Chalmers's "easy" and "hard" problems is the counterpart of the line that separates "access" and "phenomenal" consciousness in Block, which also mirrors the distinction between the causal versus phenomenal characterizations of consciousness outlined earlier. Given these distinctions Chalmers states the "hard problem" as the problem of bridging the explanatory gap between accounts of the causal-functional (physical) kind and the occurrence of specific phenomenal aspects. He asks:

> Even when we have explained the performance of all the cognitive and behavioral functions in the vicinity of experience—perceptual discrimination, categorization, internal access, verbal report—there may still remain a further unanswered question: *Why is the performance of these functions accompanied by experience?* ... This further question is the key question in the problem of consciousness. Why doesn't all this information-processing go on "in the dark", free of any inner feel? (Chalmers 1995, p. 203).

There are two related questions Chalmers raises, and the roots of both go back at least a hundred years in the history of psychology and philosophy. The first has to do with how to bridge the explanatory gap between physical mechanism and phenomenal appearance, or brain and mind, as discussed in the "Mystery of Consciousness" section in section I. The second asks whether all the activity on the "physical" side could go on as usual in the total absence of any counterpart phenomenology. The former question is based on Levine's (1983, 1993) original formulation of the problem of the "explanatory gap," which has antecedents in considerations raised by Saul Kripke (1980) and Thomas Nagel (1974, 1986) (though reaching different conclusions). The latter question is a version of the so-called absent qualia problem.[56] Similar considerations also underlie what William Seager (1991) calls the *ultimate problem of consciousness*. He asks: "Why is it so hard to think about consciousness, to formulate reasonable models of the relation of particular modes of consciousness to their physical bases?" The answer Seager offers, in agreement with Nagel, has to do with the uniqueness of consciousness as a phenomenon: "There is no model by which we can satisfactorily understand the relation between conscious experience and subvening physical state since this relation is absolutely unique in nature" (pp. 223–234).

Before discussing the status of the hard problem further, let me first sketch a larger framework in which a number of relevant questions can be located and pursued. This framework will also be useful in revealing just how much explanation along one question-path can be useful in explaining issues in neighboring problems.

XV The Four W Questions and the Further-How Question

It is true that the enterprise of approaching consciousness within a scientific discipline has traditionally been very problematic, largely due to the inadequacy of the scientific third-person perspective *all by itself* as a penetrating tool for the study of the phenomenal character of consciousness.

The most influential assumption about consciousness, as evidenced by the diversity of the literature, is that what makes it a tough nut to crack is in some crucial way related to those properties of consciousness that have to do with its phenomenal aspect. Problems such as the irreducibility of consciousness, its imminent subjectivity, the status of its relation to its physical underpinnings, and so on all relate to the phenomenal side of consciousness.

In order to get into some of the inner structure of this difficulty, consider the following five questions, which I will call the four *W questions* and the *further-How question* of consciousness:

1. *What* are the media and mechanisms of consciousness? Can consciousness occur in any type of material substance, or does it have to have a specific kind of underpinning (e.g., a carbon-based molecular structure)? And what are the underlying mechanisms that facilitate consciousness?

2. *Where* is, if anywhere, the locus of consciousness? Can consciousness be localized in a specific organ, the brain (or a module in the brain), or is it endemic to the whole of the nervous system? Where is the seat of consciousness?

3. *Who* can be said to be a conscious being? Using consciousness as a type-identifying predicate, one can ask: Is a chimp, a spider, a protozoan, or a robot conscious or *non*conscious? (In a slightly different sense of consciousness, one can also ask of a person in a coma, or in sleep, or in a petit mal seizure whether she is conscious or *un*conscious.)

4. *Why* is there consciousness at all, and what is the role it plays in the general scheme of mental life and behavior of an organism? To put it in evolutionary terms, *which* function does consciousness serve such that it was selected as a trait in the phylogeny of certain classes of living things?

5. *How* does consciousness arise in, or emerge from, its underlying substance, structure, and mechanism, in the way it does?

Notice that the How question seems to be a further question, the answer to which may not be completely revealed even if all the previous four W questions are already adequately answered. The answer to the How question may involve the postulation of, in Chalmers's terms, an *extra ingredient*, which makes the question difficult in a unique way. Even when all the underpinnings of consciousness, including its medium, locus, and mechanism, are revealed, and conscious and nonconscious things are, at least according to some operational definition, properly categorized and explained, a further question may remain: Just *how* is it that one experiences the particular sort of phenomenal quality that one does, rather than a different quality, or even none at all? Or, more generally, how does any physical mechanism give rise to any kind of phenomenal experience? Because of this extra ingredient seemingly inherent in the How question, I call it the *further-How* question.

As must be clear from the formulation of the further-How question, the difficulty surrounding the extra ingredient, the gap that remains not bridged, owes its difficulty to the phenomenal aspect of consciousness. The further-How question is generally considered to be categorically more difficult compared to the other four—hence the dubbing: "the hard problem of consciousness" versus the rest, the "easy problems."

It is important to notice that the term *extra ingredient* can carry greatly different theoretical weights. For instance, the missing extra ingredient may be merely explanatory, due to an undeveloped concept, or some other theoretical tool. That would raise only an epistemological problem. But it may also mean a missing ingredient in the part and parcel of the world, in its ontology. It has been suggested that the missing ingredient is indeed ontological, and consciousness should be added to the list of fundamental physical elements of the universe. For instance, Chalmers (1995) claims that "a theory of consciousness requires the addition of *something* fundamental to our ontology" and suggests we

take experience as fundamental "along-side mass, charge, and space-time" (p. 210; see also Chalmers 1996). Nagel and Searle have respectively made the same point in terms of the subjective properties of consciousness, in, for instance, the following passages:

> The subjectivity of consciousness is an irreducible feature of reality—without which we couldn't do physics or anything else—and it must occupy as fundamental a place in any credible world view as matter, energy, space, time, and numbers. (Nagel 1986, pp. 7–8)

> Conscious mental states and processes have a special feature not possessed by other natural phenomena, namely subjectivity. It is this feature of consciousness that makes its study so recalcitrant to the conventional methods of biological and psychological research, and most puzzling to philosophical analysis.... The world ... contains subjectivity as a rock-bottom element.... In the sense in which I am here using the tem, "subjective" refers to an ontological category, not to an epistemic mode. (Searle 1992, pp. 93, 95)[57]

The same idea has also been favored among those who try to find fundamental theoretical connections between consciousness and quantum physics, as well as those who popularize on this theme. For instance, an interview with Nick Herbert, the author of *Elemental Mind: Human Consciousness and the New Physics*, outlines his position as arguing that "consciousness itself must be considered a 'fundamental force' of the universe, 'elemental', on a par with such irreducible phenomena as gravity, light, mass, and electrical charge" (quoted in "The Consciousness Wars," *Omni*, October 1993, p. 56; see also Herbert 1993). In a somewhat similar spirit, theoretical physicist Henry Stapp claims that "an analysis of the measurement problem of quantum theory points to the need to introduce consciousness, per se, to physics," stressing as well that a complete account of consciousness can be given not in an "ontologically and dynamically monistic conceptualization of the world provided by classical-mechanics" but only "within a dualistic quantum-mechanical conceptualization of nature" (Stapp 1996, p. i).

The study of consciousness can take any one of the above five questions as its entry point to investigation. Indeed, various people have made attempts to approach the phenomenon of consciousness by respectively addressing each of these issues. But the further-How question has typically generated less success than others. As a matter of fact, it led to grim diagnoses about the "explanatory gap," thought by some, such as Colin McGinn (1989), to lie possibly forever beyond the grasp of human understanding.

The seeming uniqueness of the further-How question, given the lack of apparent promising directions to pursue it in any existing methodology, led McGinn (1989) to take its conclusions perhaps too seriously. The very same considerations, on the other hand, can lead one to think that there is perhaps something fishy about the whole setup. The way the problem is presented relies obviously on a set of presumptions about the metaphysics of phenomenal consciousness, as well as the nature of scientific explanation. Could it be that the reason we seem to have no clue about how to explain the further-How question is that there is really nothing there to explain?[58] This brings up a metalevel issue: whether a complete explanation of the four W questions will in fact leave some further aspect of consciousness unexplained, such that the further-How question will remain untouched, unscathed, and in need of explanation as ever? To take up a favorite example of the Churchlands from the history of scientific explanation, what can assure us that the further-How question will not evaporate in time just as did questions about élan vital and phlogiston?

Note that this sort of skepticism against the further-How question need not entail a deflationary attitude toward consciousness in general. One can remain convinced that consciousness presents fascinating and real problems for philosophy and science and that this is already justified in the history of its study, while not believing that there is a further-How question in the way it is formulated, isolable from the four W questions such that no degree of understanding there will shed any light on it.

I do not know if there is a decisive way to settle the metaissue at this stage of our understanding of consciousness, and thereby decide the fate of the further-How question. I do not know if it is useful, or even yet possible, to settle it at present. It seems that the opposing attitudes toward consciousness stem largely from pretheoretical, though (or perhaps, hence) deep-rooted and very strongly held, intuitions. Of course, it is crucial to try to systematically examine and uncover the often implicit presumptions that these intuitions embody, but doing that also requires understanding what is currently known and accepted about consciousness at the present theoretical level—that is, understanding what is known about the four W questions. Each of these W questions is interesting in its own way, and each has generated some fruitful thinking independently in different fields. Thus, I now turn to a brief exposition of their current status, primarily the What and the Where questions.

The What Question

With regard to functionally characterized varieties of access consciousness, there is hardly any suspicion that consciousness is medium independent. But regarding phenomenal consciousness, this question is open to speculation. The functionalist intuitions suggest that if the existence of all mental phenomena, including P-consciousness, is a matter of the functional organization of the elements in the nervous system, then the possibility that consciousness is a trait that is not restricted to carbon-based animal brains of this planet should be allowed. Denying this possibility would be "neural chauvinism."

Perhaps Searle comes closest to claiming that consciousness, and actually the mind in general, can occur only in human and animal brains, or their causal (but not necessarily functional) equivalents, because of the "special powers of the brain," which cannot be matched by, for example, digital computers.[59] Notice that the question here is different from those in the various absent qualia arguments. The possibility being questioned is not one of non-emergence (i.e., absence) of consciousness in functional equivalents of human brains or in human brains themselves. Rather, somewhat symmetrically, it is the possibility of the emergence of consciousness in *non*-brains.

What about mechanism? Regarding the underlying mechanism of a very important component of consciousness, the binding of the various sensory features into a coherent whole in experience, the most promising recent results come from the work of Christof Koch and Francis Crick. In "Towards a Neurobiological Theory of Consciousness" (1990), they hypothesize that what underlies the phenomenon of binding is the pattern of synchronous oscillations in the brain within the 40 to 70 Hz range during visual experience.[60] (See also Llinás and Ribary 1994, in support of the 40 Hz hypothesis in the context of dream experiences. Metzinger 1995a explores how the binding problem relates to the integration of phenomenal content.)

Now, let's examine this hypothesis in light of the distinctions introduced so far. Does it, for instance, explain the access or the phenomenal senses of consciousness (or both)? Since Crick and Koch do not have such a distinction, it is hard to know what they think. According to Block, the hypothesis is designed to explain P-consciousness; failing that, it can explain, if anything, only A-consciousness. A true explanation of P-consciousness, Block maintains, has to explain further questions about why, for instance, it is the 40 to 70 Hz range and not some other. The discovery of an empirical correlation does not suffice to bridge the explanatory gap between the phenomenon as it appears to the subject and what its underlying mechanism does.

This formulation is just another expression of the "hard problem" and, as such, falls in the purview of the further-How question. Thus, while the Crick-Koch hypothesis (so far as it is correct) can be considered to explain successfully the What question of consciousness for some, it remains essentially incomplete for the defenders of a Blockean conception of phenomenal consciousness.

The Where Question

Is there a seat of consciousness? This question in its various incarnations has been discussed from the time of the ancient Greeks. What was once the question of the organ of reason in humans (e.g., the brain versus the heart) has now transformed into the question of the whole brain or a module in it, and if the latter, which?

As early as the late nineteenth century, James had discussed the question of the seat of consciousness and declared that the cortex, and not the rest of the brain, is what is responsible for consciousness:

> For practical purposes, nevertheless, and limiting the meaning of the word consciousness to the personal self of the individual, we can pretty confidently answer the question prefixed to this paragraph by saying that *the cortex is the sole organ of consciousness in man.* If there be any consciousness pertaining to the lower centres, it is a consciousness of which the self knows nothing. (James 1950, pp. 66–67)

James's view was based on the experimental results of his day, which showed a significant correlation between cerebral processes and subjective reports of conscious experience. Note that James does not attempt to give an explanation of how the brain can possibly subserve conscious experience any further than outlining the relevant mechanism. In other words, James does not seem to be after anything beyond the ordinary W questions. Clearly this sort of explanation does not satisfy those who are after the further-How question.

But regardless of whether the further-How question is a well-formed formulation of inquiry, there is a lot of work to be done in explaining the mechanism of how and where consciousness emerges in a given organism. James was perhaps one of the first "consciousness modularists" by proposing that it was only a certain component of the brain that subserved consciousness.[61] Although it has always been in the scientific agenda, the belief in modularity in brain function has gained particular popularity over the last decade, especially due to the results coming from neuropsychology. Recent discoveries involving certain types of brain damage, such that the subjects become deprived of only very specific, encapsulated perceptual or cognitive abilities (e.g., prosopagnosia—the deficit of recognizing faces while almost all other visual capabilities remain intact), have provided support for theses of modular architecture.[62]

Extending this idea, one can transform the question of the modularity of mental function in general into the question of whether phenomenal consciousness in particular may be subserved by a module of some sort. Tim Shallice (1988) puts forth such a view, and a modularity hypothesis seems to lie behind Daniel Schacter's DICE model, where consciousness is depicted as a separate, functionally individuated box in the wiring diagram sketch (roughly speaking) of a nervous system (Schacter 1988). Block is also sympathetic to these models and calls the view "that treats consciousness as something that could be accomplished by a distinct system in the brain" *Cartesian modularism*, in contrast to Dennett and Kinsbourne's *Cartesian materialism*.

Cartesian materialism is the name Daniel Dennett and Marcel Kinsbourne (1992) give to the general belief that there is literally a place in the brain "where it all comes together"—something like a spatial or at least a temporal finish line that determines the outcome of various brain processes as a coherent, unitary, single experience. Dennett (1991) calls this the "Cartesian Theater" model of consciousness.[63] The idea of such a logical line in the brain makes it possible to ask questions about the temporality of certain events that take place inside the brain against the milepost of the phenomenology of experience. Denying that such a line exists makes it logically impossible to impose a fine-grained order on brain processes as having occurred prior to or following a particular experience. Dennett and Kinsbourne present a forceful argument against Cartesian materialism; for them it is the whole brain, if anything, that is in some sense the seat of consciousness. Today, the Where question, just like the What question, remains a hotly debated issue.[64]

The Who and the Why/Which Questions

The question of who can be classified as a conscious being is largely subordinate to the question of what the underlying medium and mechanism of consciousness are, at least in a materialist framework. Roughly speaking, those beings whose physical constitution (medium) allows the instantiation of those properties that indicate the working mechanism of consciousness can be safely allowed into the "charmed circle" of consciousness (barring difficulties inherent in the What question itself). It is also common practice in medicine to have a more or less circumscribed set of behavioral and psychological criteria to determine the occurrent presence or absence of consciousness in patients (e.g., see the *Roche Handbook of Differential Diagnosis* on "Transient Loss of Consciousness," 1989).

Of course, the issue is not *so* straightforward, especially when it comes to phenomenal consciousness. Is there anything it is like to be a bat catching prey with its sonar system, or a dogfish detecting electromagnetic fields in the ocean, or a robot clumsily walking about in an artificial intelligence laboratory? The answers and, more important, the advice on how to obtain these answers greatly vary. This question also leads to the discussion on "zombies"—whether there could be, in nomological or just logical possibility, human replicas who nonetheless lack phenomenal consciousness. I come back to this issue in the discussion of epiphenomenalism.

Regarding the *Why/Which* question, the literature is somewhat barren. Perhaps this is partly as a result of the fact that it is nearly impossible to find any evolutionary role for *phenomenal* consciousness to play under the decree of the segregationist intuition, whereas the evolutionary contribution of consciousness, when it is taken as causally efficacious in accord with the causal characterization, is just too obvious. In other words, so long as consciousness is characterized as essentially noncausal and nonfunctional, rendered an *epiphenomenon* that makes no difference in the world, it drops out of the pool of factors that have survival value, and thus becomes explanatorily irrelevant to evolutionary theory. This is the conclusion Frank Jackson (1982) defends (using the term "qualia" for phenomenal consciousness): "[Qualia] are an excrescence. They *do* nothing, they *explain* nothing, they serve merely to soothe the intuitions of dualists, and it is left a total mystery how they fit into the world view of science.... Epiphenomenal qualia are totally irrelevant to survival" (p. 135)[65] On the other hand, if consciousness is taken as a *genus* for different modalities of perceptual awareness under a causal-representational characterization, pace Dretske (forthcoming), there remains no philosophically puzzling question about its evolutionary role. It would clearly be somewhat difficult for any creature to survive without sight, hearing, touch, smell, and so forth. (See also Dretske 1996 for a discussion of what kind of differences qualia make vis-à-vis judgments and beliefs.) Similarly, Armstrong (1980) attributes to introspective consciousness the biological function of making us aware of current mental states and activities of our own mind, such that it becomes "much easier to achieve *integration* of the states and activities, to get them working together in the complex and sophisticated ways necessary

to achieve complex and sophisticated ends" (p. 65). See also Van Gulick (1988, 1989) for attempts to locate a functional role for phenomenal consciousness, and Dennett (1991), Flanagan (1992), Dretske (1995), and Flanagan and Polger (1995) for further evolutionary considerations.[66] Finally, Jaynes (1976) and Crook (1980) take entirely different approaches to the idea of the evolution of consciousness (characterized in terms closer to what I called the social sense of consciousness). Searle (1992) tries to strike a balance between defending a version of the essentialist intuition while assigning an evolutionary role to consciousness. Finally, a number of neuropsychological accounts identify consciousness with a specific information-processing module, in terms of a specific function it serves in the whole system. Although the concern is almost never evolutionary in such accounts, they can be mentioned here for their effort to find a specific function for consciousness (see, for example, Schacter 1988 and Shallice 1988). But on the whole, there is much about the Why/Which question that remains to be written than what is already there.

Having considered the various characterizations of consciousness and the various questions one can ask about them, I now turn to the examination of questions about phenomenal consciousness in the landscape of current philosophical debates.

XVI A Road Map for Phenomenal Consciousness and the Unbearable Lightness of *Whatitisliketobe*

As we have seen, the concept of consciousness is a hybrid that lends itself to several different characterizations. Part of my goal in this chapter was to tease them apart and treat them separately. Having done so, however, one sees that the problem of consciousness is like a Chinese box puzzle; for every distinction made, one discovers that further embedded distinctions are required.

In any case, given that the philosophical problems all revolve around the phenomenal characterization of consciousness, it is reasonable to focus discussion there. Doing so actually reveals that what is commonly referred to as "phenomenal consciousness" is also itself a hybrid. Consequently, it becomes imperative to bring the analytical microscope over there and to dissect the different elements in the tangle of phenomenal consciousness. Here I present a conceptual road map for locating various different philosophical problems, each associated with phenomenal consciousness in one way or another.

The term *phenomenal consciousness* is often used interchangeably with a variety of others, such as *qualitative character, qualia, phenomenal properties, subjective awareness, experience,* and *what it is like to be a certain organism.* (See, for instance, Block 1994, pp. 210–211.) This is a bunch. And to make matters worse, each of these concepts is known for its notorious elusiveness. Traditionally, the properties that go under the various names of "raw feels," "qualia," "qualitative character of experience," "phenomenal aspect of consciousness," and so on have all proved to be recalcitrant to systematic explanations. Dennett points out that attempts to give a straightforward account of phenomenal properties have typically been frustrating; "no sooner does [the concept of qualia] retreat in the face of one argument that 'it' reappears, apparently innocent of all charges, in a new guise" (Dennett 1988, p. 42).[67]

This elusiveness actually goes to the heart of the particular and long-standing problem of phenomenal consciousness, which is often labeled a mystery. Elusiveness by itself is not what makes the problem persistent, however. Otherwise, eliminativism could appear as a more appealing option. Rather, it is our unique epistemic relation to consciousness: phenomenal consciousness is perhaps the most difficult aspect of the mind to give up. An eliminativist stance toward the phenomenal aspect of mental life seems the most counter-intuitive of all eliminativist attitudes. That is why the question of phenomenal consciousness does not just disappear out of the philosophical and, in other guises, psychological and scientific landscapes. I thus find it important to lay out properly each conceptual component that contributes to the puzzle.

It is worth noting, however, that among all notions that are associated with phenomenal consciousness, one has particularly captured philosophical intuitions more than any other—so much so that it has become the central notion underlying almost any discussion about consciousness during the two decades since its publication. Unfortunately, it is also the most difficult to pin down or muster theoretical agreement upon. I have in mind Nagel's (1974) notion of "what it is like to be" a certain creature, or subject of experience.[68]

Nagel's notion of "what it is like to be" has been so influential that it seems to have an omnipresence in several distinct (even if related) problems with regard to consciousness. In particular, it gets pronounced in an intertwined way with the problem clusters that can be grouped under the headings of qualia, subjectivity, and the knowledge argument.

Nagel himself presents the issue of what it is like to be a certain creature as a theoretical basis for establishing the claim about that creature's having a certain ontologically irreducible point of view, which furnishes certain facts about the creature with subjectivity. For others, however, the notion of "what it is like to be" is taken to lay the ground for arguing for the reality of qualia, and for others, for the persuasiveness of the knowledge argument which claims that physicalism, as an ontological doctrine, is false. But the nature of the relations among each one of these problems is hardly ever spelled out in any detail. In fact, it seems that the notion of "what it is like to be" has become the *wild card* of consciousness problems. I will henceforth refer to it simply as the notion of *whatitisliketobe*.

Given this tangle, let me present the following schema as a conceptual road map to distinguish problems typically associated with phenomenal consciousness:

1. Qualia: Experiences have phenomenal and thus noncausal, nonrepresentational, nonfunctional, and perhaps nonphysical properties.

2. Subjectivity: Certain facts about experiences are subjective, that is, they cannot be completely understood except from a single kind of point of view.

3. Knowledge Argument: Certain facts about experiences are nonphysical.

To this, one can add the "base element" in the formula:

* Whatitisliketobe: There is something it is like to have experiences for a certain organism (or, simply, something it is like to be that organism).

I call *whatitisliketobe* a wild card, because it gets alluded to in discussions concerning any of the three problems mentioned above. To have certain qualia, it is generally presumed, is *whatitisliketobe* an organism undergoing a certain experience; certain facts about an experience are subjective because there is *somethingitisliketobe* having that experience; and finally, *whatitisliketobe* having a certain experience constitutes nonphysical facts about that experience. I think, however, that it can be questioned whether this common denominator is not in fact theoretically vacuous. Perhaps *whatitisliketobe* has turned into nothing but a wild card—a convenient way of talking about any one of the three problems of phenomenal consciousness, without, due to its intuitive charm, having to specify anything further. Then there would be no reason to look for a shared ingredient in need of explanation, above and beyond the explanation of these three problems.[69]

Nagel's original intention in introducing the notion of *whatitisliketobe* was, I think, to use it as an "intuition pump" for instating subjectivity

rather than as a tool to talk about qualia. Further, for Nagel, the scope of applicability of the notion of subjectivity greatly transcends the problem of qualia, or consciousness in general; it also underlies problems about free will, personal identity, and the self, as well as the ontological doctrine of physicalism.[70]

The knowledge argument, formulated in contemporary literature by Frank Jackson, is also much closer in nature to issues surrounding subjectivity than to the problem of qualia. In fact, it can be seen as a logical conclusion of the difficulties Nagel raises about accommodating subjectivity in a physicalist ontology. In a nutsbell, the knowledge argument is based on the claim that certain facts about experiences evade all physicalist accounts, and no matter how much one learns about the physical (causal, functional, representational, and so on) aspects of an experience, some facts about how the experience feels (to oneself, but more important, to others) will remain in the dark until one actually *has* that experience.

Jackson attempts to establish this claim by means of a thought experiment that involves an imaginary vision scientist, Mary, who learns "everything physical there is to know" about color experiences without ever having color experiences herself. Jackson's contention is that upon seeing a colored object for the first time in her life, Mary will learn something new, belying physicalism. The pivotal issue here is whether the having of an experience constitutes a special class of irreducible "first-person facts" or whether what is lacking in Mary has to do with her experiential "mode of access" to facts that she is already acquainted with (in the form of propositional knowledge); on this point of contention the knowledge argument has generated a fair amount of literature.[71] The interrelations between these problems need to be pursued further, but I will stop and opt for focusing on the most central player of the phenomenal consciousness debate: qualia. (I will henceforth use "phenomenal consciousness" and "qualia" interchangeably.)

XVII The Qualia Battles

The problem of qualia is one that surfaced under different guises in the philosophy literature during different periods. It is probably fair to state that qualia was the single most recalcitrant notion that resisted the rising wave of materialists in their program of giving an account of the mental by means of identity theory. For example, J. J. C. Smart mentions in his now-classic "Sensations and Brain Processes" (1959) that among the eight objections he considers, he feels the least confident in his answer to the one about phenomenal properties. (*This is Objection 3, attributed to* Max Black.) Both U. T. Place (1956) and B. A. Farrell (1950), philosophers of the same era, note that the identification of the so-called raw feels with the straightforwardly physical properties of the nervous system has been the most elusive component of the overall program of identity theory in "Is Consciousness a Brain Process?" and "Experience," respectively. Herbert Feigl also wrestles with the same problem in his lengthy manuscript, *The "Mental" and the "Physical"* (1967).[72]

A second wave in philosophy of mind came about, this time that of functionalism, in the 1970s. The problem of qualia was again on stage; the phenomenal feels were considered the "Achilles' heel of functionalism" (Shoemaker 1981a)—the only aspect of mentality that escaped the net of functional explanations.[73] It is during this period that the problems of *absent qualia* and a reincarnation of Locke's puzzle of *inverted spectrum* reached celebrity status. Critics of functionalism argued that a functionalist framework can provide an account of all components of mental life but cannot capture its qualia, lacking the theoretical tools to settle decisively questions about whether any two functionally equivalent systems differ (e.g., can be inverted) in their phenomenal aspects, or even whether a given system has any qualia at all. Thomas Nagel gives a concise characterization of the problem of

absent qualia as follows: "The subjective character of experience . . . is not captured by any of the familiar, recently devised reductive analyses of the mental, for all of them are logically compatible with its absence. [E.g.,] It is not analyzable in terms of any explanatory system of functional states, or intentional states, since these could be ascribed to robots or automata that behaved like people though *they experienced nothing*" (Nagel 1974, pp. 166–167; my emphasis).[74]

In a footnote to this passage, Nagel also entertains the possibility of the *impossibility* of absent qualia, but rejects it: "Perhaps there *could not* actually be such robots. Perhaps anything complex enough to behave like a person *would have* experiences. But that, if true, is a fact which cannot be discovered merely by *analyzing the concept of experience*" (fn. 2, p. 167; my emphasis). But what Nagel merely asserts as true has no argumentative force against certain causal-state identity theorists and some functionalists. For they take exactly the opposite of Nagel's assertion (broadly construed to include not only behavior, but also causal, functional, and intentional characterization) as a fundamental assumption.

For instance, David Lewis (1966) states: "The definitive characteristic of any (sort of) experience as such is [by analytic necessity] its causal role, its syndrome of most typical causes and effects" (p. 17). Similarly, the *concept* of a mental state for David Armstrong (1993) is that of a "state of the person apt for bringing about certain sorts of physical behavior," where he regards the mind as "an inner arena identified by its causal relations to outward act" (p. 129). As such, the relation between experiences and causal (and/or functional, intentional, etc.) characteristics *is* taken to be, contra Nagel's assumption, inherent in the concept of experience. This kind of fundamental disagreement where each side is vulnerable to the charge of question-begging against the other is a typical syndrome of the "qualia battles."

The possibility of absent qualia is closely related to the doctrine of epiphenomenalism (that phenomenal consciousness has no causal powers) and thus to the possibility of "zombies" (human replicas with all mental and behavioral attributes present save for phenomenal consciousness), as I argue below. The possibility of inverted spectrum, on the other hand, simply requires an inversion of a particular set of phenomenal qualities in some sensory domain, such as the hues in one's color space.[75]

But let us pause and ask the same question already posed about consciousness: When friends and foes of qualia disagree about whether qualia exist, are they really talking about the same thing? The ontologically rather ordinary fact that phenomenal properties of an experience exist only insofar as they belong to someone's experience (compare: geometric properties of a shadow exist only insofar as they belong to someone's shadow), when combined with the epistemologically rather extraordinary fact that experiences cannot epistemically be shared, and hence everyone can have "direct access" to only his or her qualia, seem to make it uniquely, even surprisingly difficult to investigate the ontological nature of qualia. As such, it gives rise to a wide variety of positions regarding what qualia *are*.

In "Quining Qualia," Dennett, one of the staunchest critics of the notion of qualia, tries to establish that "conscious experience has *no* properties that are special in *any* of the ways qualia have been supposed to be special." He attempts to show this by laying out what exactly it is that he wants to deny in denying the existence of qualia and sets up his target by identifying qualia with the "properties of a subject's mental states that are: 1. ineffable, 2. intrinsic, 3. private, and 4. directly or immediately apprehensible in consciousness" (Dennett 1988, pp. 43, 47). The final verdict Dennett arrives at, after an elaborate chain of "intuition pumps" designed to show that the very concept of qualia is inherently confused, is an eliminativist one: "There simply are no qualia at all" (p. 74).

In contrast to Dennett's eliminativist stance, the spectrum of other positions with respect to

qualia extends from taking qualia to be nonphysical properties that require a new ontology to reductively identifying qualia with neurophysiological properties. There are also midway, conciliatory positions. Paul and Patricia Churchland, for example, agree that when qualia are construed in the way Dennett does, the situation is indeed hopeless: "So long as introspectible qualia were thought to be ineffable, or epiphenomenal ... one can understand the functionalist's reluctance to have anything to do with them" (Churchland and Churchland 1982, p. 34). While promoting a realist attitude toward qualia, they claim that qualia will turn out to be properties intrinsic to the nervous system, such as spiking frequencies in the brain. Construed as such, qualia cease to be elusive, but their investigation also falls into the scope of disciplines other than philosophy or psychology. In the Churchlands' words: "The functionalist need not, and perhaps should not, attempt to deny the existence of qualia. Rather, he should be a realist about qualia.... [But, at the end], the nature of specific qualia will be revealed by neurophysiology, neurochemistry, and neurophysics" (Churchland and Churchland 1982, p. 31).

Owen Flanagan, who believes that an effort of triangulation involving phenomenology, psychology, and neuroscience, which he calls the "natural method," can penetrate the mystery of qualia and help dispel it, follows suit in promoting a more positive characterization of qualia: "Those who would quine qualia are bothered by the fact that they seem mysterious—essentially private, ineffable, and not subject to third-person evaluation. Qualia are none of these things." Although Flanagan does not necessarily share the Churchlands' conviction that qualia will turn out to be properties in the domain of neuroscience, he too concludes that "there are no qualia in Dennett's contentious sense, but there are qualia" (Flanagan 1992, p. 85).

A recent proposal in accounting for qualia comes from Fred Dretske's representational naturalism (Dretske 1995). According to this view, "all mental facts are representational facts" and hence, a fortiori, all facts about qualia are also representational. Dretske identifies qualia as properties that one's experience represents objects (or whatever the experience is about) as having. As such, qualia do not have to be given a functional characterization or identified with neurophysiological properties. Rather, Dretske locates qualia outside the mind, in accordance with his externalist theory of the mind. This view has the advantage of maintaining a realist stance toward qualia while remaining in a perfectly naturalistic framework.[76]

Finally, Ned Block brings the qualia issue back to the problem of "explanatory gap" and raises suspicions about the conceptual machinery of cognitive psychology to deal with qualia: "On the basis of the kind of conceptual apparatus now available in psychology, I do not know how psychology in anything like its present incarnation *could* explain qualia" (Block, 1978, p. 289). Block is neither as sure as the Churchlands about whether the answer to the nature of qualia will turn out to be in the domain of neuroscience, nor is he as optimistic as Flanagan in the promise of interdisciplinary methods to deliver a successful account of phenomenal properties. Nor is he convinced that qualia can be accounted for in a Dretskean representational framework. On the contrary, Block actually wants to raise more general doubts about the explanatory power of *any* mechanistic, functionalist, or in general physicalistic schemes to account for the presence or emergence of qualia. His worry, in other words, is about how qualia *can* be accounted for as part and parcel of any physical system, including (or rather, especially) a brain, even if one thinks that it *must* be so accounted. Block states: "*No* physical mechanism seems very intuitively plausible as a seat of qualia, least of all a *brain*.... Since we know that *we are brain-headed systems*, and that *we* have qualia, we know that brain-headed systems can have qualia. So even though we have no theory of qualia which explains how this is *possible*, we have overwhelming reason to disregard whatever *prima facie* doubt there is about the qualia of brain-headed systems" (Block 1978, p. 281).

As a consequence of this kind of general doubt about physicalism, the scenario involving beings physiologically and behaviorally similar to us, perhaps even identical down to the last molecular structure and behavioral trait, who nonetheless lack qualia altogether, is considered a genuine theoretical possibility. This step brings us to the debate on the notion of zombies and the doctrine of epiphenomenalism.

XVIII Epiphenomenalism and the Possibility of Zombies

Consciousness epiphenomenalism is the view that (phenomenal) consciousness has no causal powers and hence exhibits no effects in the world, though it may be the effect of some other cause itself. This doctrine and the possibility of zombies are closely related. If consciousness is an epiphenomenon, that is, not *essentially* linked to causal processes, or is only a recipient of but not a contributor to effects in a causal network, then there exists the possibility that the same organism that is taken to possess consciousness could be going through the very same mentations and behavior even if it had no phenomenal consciousness at all. Subtract away the consciousness, and you still get the same beliefs, desires, motives, preferences, reasoning capacities, and behavior in the organism. But what you get is a *zombie*. Its pains, tickles, and itches are all "ersatz." The zombie does not feel anything, even if it thinks and acts as if it does. Its experiences lack the qualitative feels altogether. There is nothing it is like to be *it*.[77]

Put differently, zombiehood becomes a possibility only under a view that accords with epiphenomenalism. If we maintain that consciousness has causal powers, then the absence of consciousness in my zombie twin, which is identical to me in every other respect, would make *some* difference. But by stipulation, there is no difference whatsoever between persons and their zombie twins except the fact that the latter lack consciousness. Hence, denying epiphenomenalism would also block the possibility of zombiehood. That is, if we accept that consciousness has causal powers, then my zombie twin cannot exist, even as a genuine theoretical creature.

The doctrine of epiphenomenalism has a deep-rooted history. The philosophers and the psychologists of the nineteenth century hotly debated whether consciousness was part and parcel of the causal network that was responsible for the decisions we make, actions we take, and so forth or whether it was just an idle spectator, riding along the causal processes, perhaps being caused by them, but without exerting any causal effect on those processes itself. Perhaps, the idea was, we are all automata, since all of our mental life and behavior seem to be determined by our nervous systems, in a purely mechanical framework, with no respectable place in it for consciousness.[78]

Thomas Huxley was one of the most influential advocates of such a thesis, known as the *automaton theory of consciousness*. The thesis was first formulated to apply to animals, in perfect agreement with Cartesian intuitions. Huxley put the matter as follows: "The consciousness of brutes would appear to be related to the mechanism of their body simply as a collateral product of its working, and to be as completely without any power of modifying that working as the steam whistle which accompanies the work of a locomotive engine is without influence upon its machinery" (Huxley 1901, p. 240). But, of course, the real target was human beings and the nature of human consciousness. This is where Huxley's automaton theory differed with Descartes's interactionist dualism. Huxley's account of the "brutes" was just a lead to make the same point for humans: "The argumentation which applies to brutes holds equally good of men.... It seems to me that in men, as in brutes, there is no proof that any state of consciousness is the cause of change in the motion of the matter of the organism" (pp. 243–244).[79]

In Huxley, consciousness plays no contributory role in the causal chains that take place in the nervous systems that totally determine the behavior of an organism; it only gets affected by the neural interactions. In contrast, Descartes's idea of consciousness was one of an equally causally efficacious parameter in the formula of mind-body interaction. As much as Descartes is thought to be the founder of interactionism, Huxley can be thought of as having laid out a clear foundation for epiphenomenalism with respect to the mind. The fundamental idea about epiphenomenalism remained intact until the present day, but what was then dubbed the "Automaton Theory" has been transformed into the "Problem of Zombies" in contemporary literature. Of course, it is important to note that even if we establish that it is the truth of some version of epiphenomenalism that makes zombiehood a possibility, there remain important issues about what the nature of this possibility is, for example, whether it is empirical, metaphysical, or conceptual. These are subtle issues that I cannot do justice to in the limited space here. Hence, rather than pursuing this line further, I will step back once again and examine how the background conditions for bringing the metaphysical disagreement on the possibility of zombiehood (just like the disagreement on the status of phenomenal consciousness) can be brought to a settlement.[80]

XIX Stalemate: How to Settle the Phenomenal Consciousness Dispute?

There are a variety of positions on the ontological status of phenomenal consciousness in the literature, all the way from substance dualism to property dualism, to reductionism (via some form of identity thesis), to eliminativism (usually coupled with some kind of antirealist stance), to representationalism (maintaining a naturalized realism). However, the literature does not contain any knock-down argument that would convince a "friend of phenomenal consciousness" to a reconciliatory middle ground with a "qualia skeptic." Most often, the disagreement between the two parties comes down, for each side, to the charge of begging the question against the other. The eliminativists charge the defenders of phenomenal consciousness with believing in a fiction and creating a philosophical problem out of it. In return, the eliminativists get charged with holding the most preposterous philosophical fancy for denying their opponents' characterization of qualia.

As an example, consider Daniel Dennett, who is convinced that the notion of qualia "fosters nothing but confusion, and refers in the end to no properties or features at all" (Dennett 1988, p. 49). Ned Block, as a representative of the other side of the spectrum, accuses Dennett of begging the question against (the existence of) phenomenal consciousness (Block 1993, 1995). Interestingly, the dialectic of the debate seems to be at an impossible impasse: the contention is at the fundamental level of taking for granted versus denying the existence of a feature of mentality that can at best be defined *ostensively*. Friends of qualia, as exemplified by Block, claim that there is obviously *something* in their mental life that can be theorized about under the name "phenomenal consciousness," while the qualia skeptics, as exemplified by Dennett, state that there is *no* such thing to point at in their own experience.[81]

This is unfortunately the kind of philosophical junction at which most worthy disagreements hit rock bottom. Neither side is willing to concede their own point, and moreover neither side seems to have any way of demonstrating the validity of their claim. In another statement on the side of the friends of phenomenal consciousness, John Searle satirically asks: "How, for example, would one go about refuting the view that consciousness does not exist? Should I pinch its adherents to remind them that they are conscious? Should I pinch myself and report the results in the *Journal of Philosophy*?" (Searle 1992, p. 8). In contrast,

Dennett declares: "I cannot prove that no such sort of consciousness exists. I also cannot prove that gremlins don't exist. The best I can do is to show that there is no respectable motivation for believing in it" (Dennett 1991, p. 406).

Of course, the situation on the whole (and the particular state-of-the-art philosophical understanding of the mind-body problem we have arrived at after twenty-five hundred years of pondering) is more nuanced than I have just sketched. For instance, the eliminativist position has more resourceful ways of undermining belief in qualia, and the "friends of qualia" have intuitively appealing conceptual tools on their side, such as the absent and inverted qualia puzzles and the knowledge argument. Nonetheless, neither side can help finding the other's theoretical maneuvers equally unconvincing.

The eliminativist strategy largely depends on the deconstruction of the concept of phenomenal consciousness, thus revealing theoretical tensions internal to it. In different ways, both Daniel Dennett and Richard Rorty take this approach (Dennett 1988; Rorty 1979). Dennett does this by providing a number of "intuition pumps," designed to show that our pretheoretical intuitions about phenomenal consciousness are far from being reliable and sound. On the contrary, as Dennett attempts to show, our commonsense grasp of the facts about phenomenal consciousness can result in such conceptual dilemmas that it might be a better strategy to abandon any talk about phenomenal properties altogether. Dennett is quite straightforward in this approach; he says: "I want to make it just as uncomfortable for anyone to talk of qualia—or 'raw feels' or 'phenomenal properties' or 'subjective and intrinsic properties' or 'the qualitative character' or experience—with the standard presumption that they, and everyone else, knows what on earth they are talking about" (Dennett 1988, p. 43).

If there indeed is a conceptual disarray surrounding the notion of phenomenal consciousness, it seems only fair to demand from those who take the idea of phenomenal consciousness seriously and use it as a fundamental theoretical tenet to come up with a clarified conceptual network of terms that all go along with the umbrella term of *phenomenal consciousness*. On the other hand, there may be good reasons to respect the words of the supporters of phenomenal consciousness that the only definitional way open to them is by ostension. Being unable to provide a nonostensive definition is not, by itself, sufficient reason to pronounce the notion of phenomenal consciousness as theoretically illegitimate, and thereby promote its complete abandonment. The merits or shortcomings of an ostensive definition in revealing the essence of a phenomenon have to be judged on its own ground, in virtue of its success in providing conceptual clarity and theoretical agreement in the relevant discussions.

It should be acknowledged, however, that the strategy of revealing essences by means of "pointing" has not delivered any kind of agreement with respect to phenomenal consciousness thus far.[82] The same problem appears even more acutely in thinking about the possibility of zombies. How can you tell a zombie from a nonzombie, someone who has absent qualia from someone whose qualia are intact? If zombiehood is a possibility, not only could your closest friend turn out to be a zombie, without anyone's knowledge or awareness, so could you, and not know it yourself. Zombiehood brings with it not only the problem of other minds, and thus third-person skepticism, but first-person skepticism as well. If you, the reader of these lines, suddenly turned into a zombie, no one would notice any difference, and in a significant sense of "noticing," neither would you. Remember that knowing, judging, thinking, and being aware of—in a nonphenomenal sense—are all capabilities granted to a zombie, and furthermore, "there is no need to invoke qualia in the explanation of how we ascribe mental states to ourselves [because a zombie] after all, ascribes himself the same qualia; it's just that he's wrong about it"

(Chalmers 1993; Chalmers 1996 embraces the consequences of this result under the title "the paradox of phenomenal judgement").

Thus, to the extent that "seemings" of your own phenomenal states are constituted by self-ascriptive judgments, beliefs, thoughts, memories, expectations, and so forth about those states (and no doubt there is a significant extent to which such seemings are so constituted), it would be warranted to say that your inner life would continue to *seem* the same to you, despite the fact that you would cease to have any genuine phenomenal states once you turned into a zombie. Put differently, according to the zombie hypothesis, you could now be "hallucinating" your own phenomenology. You would, *ex hypothesi*, be confidently judging that nothing changed in your inner life, and be mistaken about it, but you would never be able to find this out. Indeed, for all you know, your present existence on earth could be continuing in alternating phases of humanhood versus zombiehood, switching every other minute. Hmm....

Coming back to a distinction I introduced at the beginning of this chapter, it is also important to note that the segregationist intuition plays into the hands of epiphenomenalism and the possibility of zombiehood. Characterizing consciousness in essentially noncausal (nonfunctional, nonrepresentational) terms leaves no epistemic hook for making it possible to detect the presence or absence of phenomenal consciousness, even from a first-person perspective.

But if we are to accept the possibility that any one of us can be a zombie and not know it, that is, if any one of us can be totally lacking phenomenal consciousness while not being able to find out about it, how can we possibly expect a stalemate over the ontology of phenomenon consciousness to be resolved, while fundamentally relying on ostention for its presence?

The stalemate seems unresolvable under the proposed terms. Perhaps, then, there is something fundamentally misleading here, and it is time to start looking for ways of building an alternative conception of phenomenal consciousness based on the integrationist intuition—not one that eliminates phenomenal consciousness but not one that renders it completely inefficacious, or opaque even from the first-person perspective either. Rather, the conception should take the first-person characterization of experience seriously and support the commonsense understanding of phenomenal consciousness.

The bottom line of what seems most unacceptable here is the fact that under a framework that allows for the possibility of zombies, phenomenal consciousness is to be regarded as making no difference, in an epistemically significant sense, even in the first person. That is, a well-intended effort to promote phenomenal consciousness by conceptually separating it from all causal and representational properties actually yields a position with the opposite theoretical consequence: the demotion of phenomenal consciousness to a ghostly existence. If it is *this* sort of a property that we talk about when we consider phenomenal consciousness, would we really lose much (anything) by doing away with it?[83] And if we are committed to (internal) "pointing" as the only reliable way to verify the existence of phenomenal consciousness, the knowledge of the absence or presence of which is hidden even from the first-person perspective, that is, to the person who has it, should we perhaps not reconsider our very concept of phenomenal consciousness?

XX In Place of a Conclusion

I would like to leave the reader with the two questions I just posed above. But let me also give a brief recapitulation and try to tie some of the loose ends.

I started by noting an epistemological asymmetry in the way one has access to (the facts about) one's own experiences versus those of others. This asymmetry leads us to the notion of perspectivity, something quite unique to (the study of) consciousness, and to the distinction between first-person and third-person points of view. This duality between points of view with respect to accessing facts about experiences also manifests itself in a duality in characterizing consciousness, in causal versus phenomenal terms.

Taking these characterizations as mutually exclusive, based on the presumption that phenomenal consciousness is essentially phenomenal *and* essentially noncausal, yields what I called the *segregationist intuition*. Opposing it is the *integrationist intuition*, which maintains that phenomenal consciousness can only be characterized by means of all causal, functional, or representational elements. Given these two intuitions, I briefly argued that the former plays into the hands of the doctrine of epiphenomenalism, which, when combined with considerations from the possibility of absent qualia and zombiehood, leads us into untenable and noncommonsensical conceptions of phenomenal consciousness. This is good evidence, on the other hand, to take the latter seriously and use it as the pretheoretical basis in reexamining our notion of phenomenal consciousness.[84]

Another domain where the epistemic element of perspectivity figures in is the problem of the explanatory gap and the question of the "hard problem" of consciousness. There seems to be an unbridged gap in the explanation of how physical embodiment and conscious experience are linked. The former is in general given a causal characterization from a third-person perspective, the latter a phenomenal characterization in first-person terms. It seems that under our existing conceptual scheme, bolstered by the segregationist intuition, the "hard problem" just does not, and cannot, lend itself to a solution.

What is important to note here is that the explanatory gap, in the way it is set up, stems from an epistemological issue. The further question that remains is whether its persistence is good enough evidence to yield ontological conclusions. Some think yes; introducing an "extra ingredient" into the picture and thus augmenting one's ontology to include consciousness as a fundamental element could indeed relieve one of the nagging problem of having to bridge mechanism and experience (by emergence, reduction, elimination, and so forth) or vice versa. Others think that the epistemological nature of the explanatory gap does not warrant ontological conclusions. Although I cannot go into this debate in any further detail here, I too would like to lend my support to this latter position. True, in the presence of the explanatory gap, the link between experience and its physical underpinnings may seem arbitrary, but I think that the decision to introduce a new fundamental element into the ontology, based on the explanatory gap, seems equally arbitrary as well. At least I fail to see how the most steadfast belief in a thus-expanded new ontology would leave one *less puzzled* about just how consciousness relates to its physical underpinnings, hence diminishing the explanatory gap and explaining away the further-How question. What seems the most promising direction in reapproaching consciousness and pursuing its deep-rooted problems in the present era involves rethinking epistemology and conceptual schemes (as opposed to a priori postulation of new ontology) to yield a cross-fertilization of the first-person and third-person perspectives, which would allow theorizing about how causal efficacy figures in *how consciousness feels*, and how phenomenal quality relates to *what consciousness does*.

In any case, at present it just does not seem as if there is a way to settle the dispute decisively about the "hard problem" or the consequences of the explanatory gap. And given the troublesome stalemate over the ontological nature of phenomenal consciousness, we seem to be not quite near a satisfactory understanding of the phenomenon. If anything, the survey of the contemporary issues and current debates surrounding consciousness points to a need for a careful reexamination of our pretheoretical intuitions and conceptual foundations on which to build better accounts of consciousness. It also seems probable

that an entirely satisfactory understanding of consciousness will be possible, if at all, only when the constitutive elements of a more comprehensive framework, in which consciousness needs to be theoretically situated, are themselves better understood. And these elements include nothing less than causality, representation, indexicality, and personhood, and especially the deep-rooted dichotomies between mental and physical, and subjective and objective. As such, it is probably reasonable to assume, as Jerry Fodor likes to prognosticate regarding a complete account of rationality, that "no such theory will be available by this time next week."[85]

This being said, I conclude on a more positive note. Presently, there is an impressive rising tide of interest in the study of consciousness, and thanks to recent advances in interdisciplinary research, we are now in a better position to penetrate the mysteries of this great intellectual frontier. By integrating methodologies and perspectives from psychology, philosophy, neuroscience, cognitive science, and other disciplines and by keeping a mindful eye on the successes and failures of the past, we should be able to reach a higher vantage point and to see more broadly and more deeply than has ever before been possible. These are very exciting times for thinking about consciousness.

Acknowledgments

I would like to thank my coeditors, Ned Block and Owen Flanagan, for many helpful suggestions. Special thanks to Fred Dretske, John Perry, Roger Shepard, Brian C. Smith, and again to Owen Flanagan, for teaching me what I needed to know to write this chapter, and for their support throughout the preparation of this book.

Notes

1. The scope of my work has been limited to the philosophical and scientific paradigms rooted in Western intellectual history and, more specifically, in the analytic tradition. Obviously, there is a wealth of fascinating issues, questions, and approaches concerning consciousness that lie outside this limited scope. This chapter should not be taken as an attempt to give an exhaustive survey of all aspects of consciousness even within its own scope, and certainly not as embodying the grander ambition of covering all paradigms of the study of consciousness.

2. The results of this effort were discussed as part of a workshop attended by anesthesiologists, neurophysiologists, psychologists, and medicolegal experts in Cardiff, United Kingdom, in 1986. Rosen and Lunn (1987) is an outcome of this workshop.

3. Michenfelder makes the same point in another way, in concluding, "Thus there are a variety of end points one might choose to answer the question 'When is the brain anesthetized?' and there is no obvious basis for selecting one over another" (Michenfelder 1988, p. 36). However, cf. Nikolinakos (1994) for an optimistic outlook on the role of consciousness in anesthesia research. See, also, Flohr (1995) for an information-theoretic model of anesthesia where the "threshold of consciousness" is determined in terms of the brain's representational activity.

4. If we go down the phylogenetic ladder—for instance, from humans all the way to amoebae—where are we to cut the line and determine the bounds of the *charmed circle* to which only those who possess consciousness can belong? (The metaphor of the "charmed circle" is from Dennett 1987, p. 161.) Chimps, dogs, spiders? What about infants, fetuses, or comatose patients? On the other hand, if we insist on experience of sensations, itches, and tingles as necessary components of consciousness, is there any principled reason for stoping short of requiring something further, such as a conceptual overlay that makes possible one's situated awareness of one's own place and relations with others (not to talk of the Cartesian *res cogitans*), as essential to the nature of consciousness? The answers to these questions are all up for theoretical grabs.

It is also sociologically interesting to look at patterns in the common sense attributions of various mental abilities to various organisms. In contrast to widely dissenting opinions on the attribution of consciousness to others, there does not seem to be such a significant variation in pretheoretic intuitions with regard to attributions of intelligence, or perceptual capabilities. A preliminary survey study conducted on approximately one hundred Stanford students and faculty, based on a two-dimensional matrix of mental attributes (e.g., the ability to perceive, the ability to learn, intelligence, consciousness) versus kinds of organisms (e.g., protozoa, spiders, chimpanzees, humans) seemed to indicate a bias in our attributions toward reserving consciousness most exclusively for ourselves, while being more generous with the attributions of other mental abilities. (Güzeldere 1993).

5. This is, it turns out, a very tricky question. On the one hand, patients with only cortical brain damage make a striking contrast with those who further lack a functional brainstem. For instance, a report on the diagnosis of death, prepared by the President's Commission, makes the following statement:

> The startling contrast between bodies lacking *all* brain functions and patients with intact brain stems (despite severe neocortical damage) manifests [a tremendous difference with respect to responsiveness, and hence the attribute of life]. The former lie with fixed pupils, motionless except for the chest movements produced by their respirators. The latter can not only breathe, metabolize, maintain temperature and blood pressure, and so forth, *on their own*, but also sigh, yawn, track light with their eyes, and react to pain or reflex stimulation.

On the other hand, the commission shies away from reaching any conclusion with respect to the absence or presence of consciousness in patients of either kind: "It is not known which portions of the brain are responsible for cognition and consciousness; what little is known points to substantial interconnection among the brain stem, subcortical structures, and the neocortex" (President's Commission 1981, quoted in Capron 1988, pp. 161, 160, respectively).

Perhaps it is altogether misleading to think of the presence of consciousness in a binary fashion. It might be necessary to talk about degrees of consciousness, which could allow one to say that normal human beings are "more conscious" than those with brain damage, the brain-damaged patients more than those without a brainstem, and so forth.

6. Even though McGinn (1989) cites Julian Huxley as the author of this by now very popular, colorful quote (with no source), the credit belongs to T. H. Huxley (Julian Huxley's grandfather). In full, it reads: "But what consciousness is, we know not; and how it is that anything so remarkable as a state of consciousness comes about as the result of irritating nervous tissue, is just as unaccountable as the appearance of the Djin when Aladdin rubbed his lamp in the story, or as any other ultimate fact of nature" (Huxley 1866, 193). Interestingly, Huxley seems to have removed the reference to the Djin (as well as Aladdin's lamp) in the later editions of this book. For instance, in the 1876 edition, the same passage appears as: "But what consciousness is, we know not; and how it is that anything so remarkable as a state of consciousness comes about as the result of *irritating nervous tissue, is just as unaccountable as any* other ultimate fact of nature" (p. 188). Too bad the Djin is no longer around, for we could perhaps have wished from it to tell us if we would ever be able to solve the mind-body problem.

7. There is a longer passage in Tyndall's "Scientific Realism" where he addresses the mind-body problem in length, and concludes, in agreement with McGinn (though in 1868) that it is "as insoluble, in its modern form, as it was in the prescientific ages." Because the points Tyndall touches upon are so remarkably close to the contemporary formulations of the issues debated in the literature under the labels "explanatory gap" and the "hard problem" (e.g., the nature of the explanation between mind and body as opposed to other physical phenomena; the prospects for the mind-body problem upon reaching a fully advanced neuroscientific understanding of the brain; and the status of possible correlation-based accounts of consciousness), I quote this passage here in its entirety. (A more detailed discussion follows in section XIV.)

> The relation of physics to consciousness being thus invariable, it follows that, given the state of the brain, the corresponding thought or feeling might be inferred: or, given the thought or feeling, the corresponding state of the brain might be inferred. But how inferred? It would be at bottom not a case of logical inference at all, but of empirical association. You may reply, that many of the inferences of science are of this character—the inference, for example, that an electric current, of a given direction, will deflect a magnetic needle in a definite way. But the cases differ in this, that the passage from the current to the needle, if not demonstrable, is conceivable, and that we entertain no doubt as to the final mechanical solution of the problem. But the passage from the physics of the brain to the corresponding facts of consciousness is inconceivable as a result of mechanics.

> Granted that a definite thought, and a definite molecular action in the brain, occur simultaneously; we do not possess the intellectual organ, nor apparently any rudiment of the organ, which would enable us to pass, by a process of reasoning, from the one to the other. They appear together, but we do not know why. Were our minds and senses so expanded, strengthened, and illuminated, as to enable us to see and feel the very molecules of the brain; were we capable of following all their motions, all their groupings, all their electrical discharges, if such there be; and were we intimately acquainted with the corresponding states of thought and feeling, we should be as far as ever from the solution of the problem, 'How are these physical processes connected with the facts of consciousness?'. The chasm between the two classes of phenomena would still remain intellectually impassable.
>
> Let the consciousness of love, for example, be associated with a right-handed spiral motion of the molecules of the brain, and the consciousness of hate with a left-handed spiral motion. We should then know, when we love, that the motion is in one direction, and, when we hate, that the motion is in the other; but the 'WHY?' would remain as unanswerable as before. (Tyndall 1868, pp. 86–87)

It is worth noting that equally dissenting opinions were also available at the time. For instance, Huxley (1901) states: "I hold, with the Materialist, that the human body, like all living bodies, is a machine, all operations of which, sooner or later, be explained on physical principles. I believe that we shall, sooner or later, arrive at a mechanical equivalent of consciousness, just as we have arrived at a mechanical equivalent of heat" (p. 191).

8. For a materialist response to Adams, see Lewis (1995).

9. Antecedents of McGinn's view can be found in Nagel's work in several places, although Nagel only points out the possibility of McGinn's position without committing himself to it. For instance, he says:

> We cannot directly see a necessary connection, if there is one, between phenomenological pain and a physiologically described brain state any more than we can directly see the necessary connection between increase in temperature and pressure of a gas at a constant volume. In the latter case the necessity of the connection becomes clear only when we descend to the level of molecular description: till then it appears as a contingent correlation. In the psychophysical case we have no idea whether there is such a deeper level or what it could be; but even if there is, the possibility that pain might be necessarily connected with a brain state at this deeper level does not permit us to conclude that pain might be directly analyzable in physical or even topic-neutral terms.... *Even if such a deeper level existed, we might be permanently blocked from a general understanding of it.* (Nagel 1986, pp. 48–49; my emphasis).

The possibility of a permanent cognitive closure in humans with regard to the understanding of the mind-body relation is an intriguing idea. But when it comes to taking this possibility as a statement of certainty, as McGinn does, it seems fair to question what warrants this conviction, especially in the absence of empirically grounded reasons. In particular, there is a curious tension between McGinn's confidence, on one hand, of his own cognitive ability to *assert* such a sweeping verdict on behalf of all human beings (at present as well as in the foreseeable future), and the aim of his argument, on the other hand, which ultimately strives to *attenuate* confidence in the powers of human cognitive abilities to solve the mind-body problem.

Put differently, McGinn wants his readers to simultaneously believe, as a result of his largely a priori reasoning, both that our cognitive abilities are limited to forever fall short of bringing a solution to the mind-body problem, *and* that they are nonetheless sufficiently powerful to foresee the exploratory limits of the human mind vis-à-vis the very same issue. By the same kind of reasoning, isn't there equally good reason to think, one wonders, that the opposite claim might rather be true—that it is more likely (or, at least not less likely) that we will someday come to a satisfactory understanding of the nature of the mind-body relation than it is that we will ever be able to determine how far the human understanding will extend? (For a thorough critique of McGinn's position, see Flanagan (1992), chap. 6.)

10. It is actually not a straightforward matter to give a precise definition of naturalism, and I will not attempt one here. Roughly speaking, I take "naturalism" to denote the view that everything is composed of fundamental entities recognized by the natural sciences (ontological dimension), and possibly that the acceptable methods of theorizing about these entities are those commensurable with methods employed in the natural sciences (methodological dimension). For a comprehensive overview of naturalistic approaches in epistemology and philosophy of science in the twentieth century, see Kitcher (1992); for a recent analysis of the present status of naturalism, see Stroud (1996).

In philosophy mind, naturalism is often regarded as a close ally of two related but not identical views, materialism and physicalism, but there are exceptions. See, for instance, Post (1987) for a detailed attempt to lay out a fully naturalist but nonreductive metaphysics that is committed to a monism of entities with a pluralism of irreducible "emergent" properties. See also Chalmers (1996) for a somewhat similar view he calls "naturalistic dualism." For a thorough examination of the physicalist program, compare Poland (1994); for contemporary objections to physicalism, see Robinson (1993).

11. As it happens, Miller is not the first one to think about placing a ban on "consciousness" in order to help sort out the tangles in related terminologies. Here is a quote from the turn of the century, by philosopher and William James scholar Ralph Barton Perry (1904): "Were the use of the term 'consciousness' to be forbidden for a season, contemporary thought would be set for the wholesome task of discovering more definite terms with which to replace it, and a very considerable amount of convenient mystery would be dissipated. There is no philosophical term at once so popular and so devoid of standard meaning.... Consciousness comprises everything that is, and indefinitely much more. It is small wonder that the definition of it is little attempted" (p. 282). Hence, according to Perry, it is not (only) the ubiquitous familiarity with consciousness that renders attempts to give it a precise characterization or definition unnecessary; the reason is rather the difficulty of the analytic task involved in doing so.

12. A related term in Latin was *conscientia*, which can literally be translated as "knowledge with," which appeared in English and in French as "conscience." "Conscience" also had, and still has, a significant ethical aspect, which is reflected in another *OED* entry: "Internal knowledge, especially of one's own innocence, guilt, deficiencies, etc." (Cf. Baldwin 1901). Nonetheless, *consciousness* and *conscience* have been separate words with quite distinct meanings in English, at least since the time of Locke. In contrast, there is only one word in Romance languages like French (*conscience*) and Italian (*coscienza*) that carries both meanings. Outside the boundaries of Indo-European languages, the term corresponding to consciousness in, say, Turkish—an Altaic language—carries a political, but not really ethical, connotation, in addition to the common psychological usage (*bilinç* or *şuur*, as opposed to *vicdan*). In any case, this little linguistic excursion gives no evidence of a semantic taxonomy that systematically relates to one based on language families.

13. The following quotation, taken from the circles where the concern with consciousness has to do primarily with the social rather than the psychological sense highlights this distinction quite eloquently:

> When you speak of "consciousness," you do not refer to the moral conscious: the very rigor of your methods ensures that you do not leave the strictly scientific domain which belong to you. What you have in mind exclusively is the faculty of perceiving and of reacting to perception, that is to say, the psychological concept which constitutes one of the accepted meanings of the word "consciousness." (Pope Paul VI, addressing a gathering of scientists for the conference *Brain and Conscious Experience* in Rome in 1964; quoted in Kanellakos and Lukas 1974, p. i)

14. This distinction is not uncontested, however. Some, especially in continental philosophy, think that there is no intransitive sense of consciousness: all consciousness is consciousness of. Jean-Paul Sartre, for example, is a typical representative of this view: "We establish the necessity for consciousness to be consciousness *of* something. In fact it is by means of that of which it is conscious that consciousness distinguishes itself in its own eyes and that it can be self-consciousness; a consciousness which would not be consciousness (of) something would be consciousness (of) nothing. (Sartre 1956, p. 173). The origins of this kind of an essentially intentional construal of consciousness goes back to Edmund Husserl's work from which Sartre adopted his view, most likely *Ideas* (Husserl 1913). (I thank Ron Brady for this pointer.)

15. See Güzeldere (1996) for an analysis of how the creature and state senses of consciousness can be connected by means of the "Introspective Link Principle," yielding various "higher-order monitoring" conceptions of consciousness. For other recent attempts to distinguish different senses of consciousness and sort

out some definitional issues, see Lycan (1987, preface; 1997b, chap. 1), Goldman (1993), and Natsoulas (1983, 1986).

16. My decision to address problems that bear only on the psychological sense of consciousness should not be taken to imply that the two subcategories are not related in interesting ways. In fact, it seems a philosophically significant task to investigate the nature of the relation between the social and the psychological senses of consciousness—is it something more like a genus-species relation, or one of family resemblance, or something completely unique? It can also be questioned whether one can fruitfully give an analysis of one of these halves while eschewing the other. Nonetheless, for the purposes of this chapter, I opt for focusing solely on the psychological sense of consciousness. Even if it may be impossible to fully understand the social sense of consciousness without referring to the psychological sense, or vice versa, due to the conceptual disarray surrounding the term "consciousness" such an analytic strategy seems essential as a first step.

17. For an exploration of the distinction between epistemological and ontological considerations, as well as the question of whether a set of criteria to distinguish the *mental* in general from the *physical* can be coherently formulated, see among others Rorty (1970a, 1970b) and Kim (1972).

18. The status of the mind-body problem in ancient Greek philosophy is also worth a visit. It is generally argued that there is no single term in ancient Greek that reflects the counterpart of the Cartesian/Lockean conception of consciousness, and that nothing like the contemporary debates on the mind-body problem or the problem of consciousness was ever in their horizon. For instance, Matson (1966) claims that "the Greeks had no mind-body problem" (p. 101), and Wilkes (1995) argues that "[Aristotle] paid absolutely no attention to consciousness *per se*" (p. 122). Similarly, Hamlyn (1968a) states: "There is an almost total neglect of any problem arising from psycho-physical dualism and the facts of consciousness. Such problems do not seem to arise for him. The reason appears to be that concepts like that of consciousness do not figure in his conceptual scheme at all; they play no part in his analysis of perception, thought, etc. (Nor do they play any significant role in Greek thought in general.) It is this perhaps that gives his definition of the soul itself a certain inadequacy for the modern reader" (p. xiii). See also Kahn (1966), Hamlyn (1968b), and Wilkes (1988) for similar views.

However, this view is not uncontested. For instance, Alastair Hannay (1990) suggests that in Greek philosophy one can find, contrary to the skepticism expressed above, something like a distinction between the social and psychological senses of consciousness. According to Hannay, Greeks distinguished between *syneidesis* (primarily ethical individual or shared knowledge) and *synaesthesia* (Aristotle's variation of the "unity of apperception"), in much the same way as modern philosophy proceeded in the seventeenth and eighteenth centuries. In alliance, Hardie (1976) argues against Matson, Kahn, and Hamlyn, and states that "it is ... paradoxical to suggest that Aristotle was unaware of the mind-body problem" (p. 410). According to Hardie, Aristotle was "the first psychologist, and for him psychology without the conscious *psuchē* would have been Hamlet without the Prince" (p. 405). Ostenfeld (1987) goes a step further, and claims that both Plato and Aristotle were dealing with the mind-body problem in much the same sense Descartes did and we are. This debate, so far as I can see, is far from resolved at present.

19. The debate about whether consciousness consists in the higher-order awareness of first order mental states is very much alive in the contemporary literature. Among those who defend this view, some take the higher-order representation to be some form of perception (for example, Armstrong 1980, Churchland 1988, Lycan 1997), and others as some form of thought (for example, Rosenthal 1986, 1997; Carruthers 1989, 1996). For critiques, see Dretske (1993, 1995) and Shoemaker (1994). Despite the fact that such higher-order awareness accounts of consciousness have many promising aspects, I have to stop short of giving a proper exposition here. (A more detailed treatment of this approach can be found in Güzeldere 1995b.) I will also leave the discussion of "self-consciousness" (which is sometimes underwritten by such higher-order accounts) out of the scope of this chapter.

20. For Malebranche, although we can have a "clear idea" of our bodies, we cannot, unlike what Descartes believed, have a clear idea of our souls or minds. Put differently, we cannot know our minds through a clear idea; rather we know them "only through consciousness or inner sensation" such as "pain, heat, color, and all other sensible qualities" (Malebranche 1923, Elucidation 11: Knowledge of the Soul, pp. 86–87). As such, Malebranche gives qualitative aspects of the mind a much more central place in his theory, in contrast to the Cartesian view. See Schmaltz (1996) for a thorough account of Malebranche's philosophy of mind. (I thank Tad Schmaltz for the relevant material and helpful discussion on this issue.)

21. Labeling the seemingly opposing views of James of 1890 and James of 1904 simply a "change of mind" is probably too superficial a conclusion in terms of historical scholarship, and not quite fair to James either. It is important to note that James's denouncement in the latter work is of "consciousness as an entity" rather than the reality of "conscious states." Regarding consciousness as an "entity" has close connotations to Cartesian substance dualism. Even though there is no straightforward advancement of such a metaphysical position in the *Principles of Psychology*, James's position with respect to the ontology of consciousness is not entirely clear there. Hence it might be better to characterize his 1904 article as marking merely the abandonment of consciousness as a nonmaterial entity, not consciousness per se as a subject matter. This interpretation is supported by James's own remark that he means to "deny that the word [*consciousness*] stands for an entity, but to insist most emphatically that it does stand for a function" (James 1971, p. 4).

Even though this much seems quite straightforward, we are by no means faced with an unproblematic account of consciousness. In fact, James never quite works out the metaphysical presuppositions and consequences of his view of characterizing consciousness as a function, as opposed to an entity. Moreover, the ontological turn he takes toward "radical empiricism" at around the same period as the publication of his "Does Consciousness Exist?" complicates matters. It is probably well warranted to remark that William James never held a long-standing metaphysical position with respect to consciousness void of internal tensions. At a certain stage in his life, roughly midway between the publication of the two above mentioned works, he went so far as to defend the plausibility of the immortality of consciousness in an article titled "Human Immortality," in the following words: "And when finally a brain stops acting altogether, or decays, that special stream of consciousness which it subserved will vanish entirely from the natural world. But the sphere of being that supplied the consciousness would still be intact; and in that more real world with which, even whilst here, it was continuous, the consciousness might, in ways unknown to us, continue still" (James 1956, pp. 17–18). Perhaps the historical fact of the matter regarding James's attitude toward the metaphysics of consciousness is reflected most accurately in Gerald Myers's following remark, from his extensive study of James's life and thought: "James wanted to hold that in one way consciousness does not exist, but that in another way it does; yet he was never able, even to his own satisfaction, to define the two ways clearly enough to show that they are consistent rather than contradictory" (Myers 1986, p. 64). For related work, see among others Dewey (1940), Lovejoy (1963), and Reck (1972). (I thank Denis Phillips, Imants Barušs, and Eugene Taylor for helpful pointers and discussion on William James's views on consciousness.)

22. The claim of introspectionism's being well established here refers not as much to the soundness of its methodology and theoretical grounding as to its pervasiveness and preeminence in the field of psychology as a whole. To see this, one only needs to survey the monolithic psychology literature of the few decades roughly between the end of nineteenth and the beginning of the twentieth centuries: all major psychology journals are edited by the protagonists of the introspectionist school, all articles report studies involving introspection as their primary method, and so on. Ironically, the same observation holds of the period that immediately follows (roughly from late 1910s to early 1960s), except with behaviorism substituted for introspectionism.

It would be interesting to pursue the question of whether the fluctuation in James's life with respect to consciousness occurred as a result of, or was influenced by, the general air of dissatisfaction with the internal conflicts of the introspectionist school toward the end of its tenure, which led to behaviorism's rapid rise and takeover of the intellectual landscape. Or was the influence in the opposite direction? These are all intriguing questions, but unfortunately they lie outside the scope of this chapter.

23. This is only one (as it happens, also historically the most significant) use of the word *introspection*. A number of different phenomena have passed under the same name. For instance, toward the end of the last century, Brentano and Comte argued that introspection, as a second-order mental act that gathers information about first-order sensations, was misconstrued. Mill and James agreed and proposed a model of introspection as retrospection: the examination of one's own mental happenings retrospectively, through the medium of memory of the immediate past. (For details, see Lyons 1986, chap 1.) A second, separate phenomenon that

made its way to the cognitive psychology literature in the 1970s under the name *introspection* was the phenomenon of reasoning about the causes of one's own behavior, *in terms of one's beliefs, desires, motivations,* and so forth. (For a seminal article that piqued most of the initial interest in this literature, see Nisbett and Wilson 1977.) In any case, my analysis deals with introspection only in the former sense.

24. Edwin Boring notes that no subject left Wilhelm Wundt's laboratory without having provided 10,000 data points (Boring 1953, p. 172). William James humorously observes that if it had not been for the sustained patience and the inability to get bored of the leaders of introspectionism who came from the Germanic part of the continental Europe, the enterprise of introspectionism could have never endured. "They mean business," James remarks, "not chivalry" (James 1950, pp. 192–193).

25. The nature of the relation and the degree of influence between positivism and behaviorism are not uncontroversial. Even though it is generally taken for granted that the two movements enjoyed a genuine ally status, the details of this received view have recently been contested by Laurence Smith. Smith claims, "With their common intellectual background and orientation, behaviorism and logical positivism were naturally disposed to form some sort of alliance. But only after both movements were well under way was there any significant interaction between them" (Smith 1986, p. 5; cf. the rest of his book for further details).

26. In all fairness I should add that behaviorism did manage to bring in fresh air to psychology of the late nineteenth and early twentieth centuries at a time when an uncomfortable sense of containment within the rigid introspectionist paradigm was rapidly growing. The realization that psychology could employ nonhuman subjects and pursue research without being solely dependent on the linguistic data to be provided by trained introspectionists seemed, rightly, to open up new horizons. This should also explain, in part, the rather immediate success and popularity of behaviorism and the symmetrically rapid fall of introspectionism. Unfortunately, as I will detail below, behaviorism turned out to constrict psychology into an even more rigid cast in comparison to its predecessor.

27. John Watson would not miss the chance to put a nail in introspectionism's coffin by alluding to this controversy: "Psychology, as it is generally thought of, has something esoteric in its methods. If you fail to reproduce my findings, it is not due to some fault in your apparatus or in the control of your stimulus, but it is due to the fact that your introspection is untrained. The attack *is made upon the observer and* not upon the experimental setting" (Watson 1913, p. 163).

Interestingly, approximately two hundred years earlier, a similar debate had taken place between two empiricist philosophers, Locke and Berkeley, on almost exactly the same issue. The question was whether there were any "abstract ideas": ideas that are not of particular things but of universals—"types" of particular things. In the following quotation, notice that Berkeley rests his challenge of Locke's position on this question on exactly the same grounds that Titchener challenged Külpe: personal experience based on introspection.

> If any man has the faculty of framing in his mind such an idea of a triangle as is here described, it is in vain to pretend to dispute him out of it, nor would I go about it. All I desire is that the reader would fully and certainly inform himself whether he has such an idea or no. And this, methinks, can be no hard task for anyone to perform. What more easy than for anyone to look a little into his own thoughts, and *there try whether he has, or can attain to have,* an idea that shall correspond with the description that is here given of the general idea of a triangle, which is "neither oblique nor rectangle, equilateral, equicrural or scalenon, but all and none of these at once"? (Berkeley 1977, pp. 13–14)

28. There is of course a third, unmentioned but important school of psychology that emerged during the period of transition from introspectionism to behaviorism: Gestalt psychology. The fact that a separate account of Gestalt psychology is not being provided here is certainly not because it is intellectually unworthy of consideration. Quite the opposite, Gestaltists were very keen about the reasons for introspectionism's failure, and they brought a fresh new perspective on the basis of which a large number of facts in the psychophysics of perception could be fruitfully reinterpreted. Nonetheless, Gestalt psychology shared many of the same ontological assumptions with respect to consciousness and the role of phenomenology in studying the mind with *introspectionism. As a result, as far as the history of* consciousness in psychology research is concerned, it does not constitute the sort of sharp contrast that behaviorism provides. Hence, the brief treatment.

Finally, the emerging clinical wing of psychology, the psychoanalytic school, also had its disagreements with

introspectionism, and it constituted the third distinct angle of attack alongside with behaviorism and Gestalt psychology. Unfortunately, I cannot go into a detailed analysis concerning these three movements here. For a well-documented historical account of introspectionism and the debates and movements that surrounded it, cf. the section on "Modern Experimental Psychology" in Boring (1929), as well as chapters 1 and 2 of Lyons (1986).

29. So much so that it is very rare, even today, to come *across "consciousness" or "introspection" in any* psychology or cognitive science textbook, or even psychology dictionaries. See, for example, Corsini (1984) or Stillings et al. (1987), which contain no entries for "consciousness," "awareness," or "introspection."

30. Naturally, there were internal disagreements, and thus different schools, within Behaviorism, and not each brand of the doctrine was as hardheaded. Most notably, the *analytical* (logical) behaviorists (who were mostly philosophers, e.g., Hempel 1949) were interested in analyzing meanings of mental terms in a purely behavioral vocabulary, whereas the *methodological* behaviorists (who were mostly psychologists) wanted merely to restrict their research to the study of publicly observable behavior without having to attempt any conceptual analysis or even deny the reality of the publicly unobservable mental phenomena.

For example, according to Edwin Holt (1914), "the true criterion of consciousness is not introspection, but specific responsiveness" (p. 206). Since making behavior the *criterion* of consciousness is not quite the same as *identifying* the two, consciousness thus becomes "externalized" by means of a publicly observable measure, but the metaphysical question of identity is left open. As such, the two phenomena could be said to be coexistent, as Holt (1915) acknowledges in a later work: "When one is conscious of a thing, one's movements are readjusted to it, and to precisely those features of it of which one is conscious. The two domains are coterminous" (p. 172). Edward Tolman's position in his "A Behaviorist's Definition of Consciousness" (1927) is also similar to Holt's in stopping short of advancing a metaphysical claim: "Whenever an organism at a given moment of stimulation shifts and there from being ready to respond in some relatively differentiated way to being ready to respond in some relatively more differentiated way, there is consciousness" (p. 435). (See also Tolman 1967 and note 43 for his position with respect to the study of "raw feels" in psychology.)

In contrast, the *ontological* behaviorists were in favor of doing away with consciousness, or any aspect of the mind, by *identifying* it with some piece of behavior. For example, Lashley (1923) maintained the following thesis: "The conception of consciousness here advanced is, then, that of a complex integration and succession of bodily activities which are closely related to or involve the verbal and gestural mechanisms and hence most frequently come to social expression" (p. 341).

Although the assumptions of these three schools are, *by and large, logically independent of one another,* Watson (1913, 1970), an indoctrinated behaviorist, seems to have believed in all of them, arguing that the time was ripe for psychology to discard all reference to consciousness. It is no doubt that a Watsonian universe would make life much easier for philosophers and psychologists. It would, for instance, remove the epistemic *duality in the study of consciousness by collapsing* the distinction between the first-person and third-person perspectives. Furthermore, by making consciousness ultimately an operationalized parameter in the domain of behavior, it would allow a set of behavioral criteria to settle questions about who or what possesses creature consciousness. But, as is evident from the history of *psychology, life is never easy in the domain of mind.* Questions about consciousness remained a nagging issue during behaviorism's tenure, and they eventually led its prominent figures like B. F. Skinner to not only acknowledge the existence of the phenomenon, but also adopt a conciliatory position in his later works. For example, Skinner (1974), after stating that the common conception of behaviorism as a school of thought that "ignored consciousness, feelings, and states of mind" was all wrong, concedes that the "early behaviorists wasted a good deal of time, and confused an important central issue, by attacking the introspective study of mental life" (pp. 3–5).

Undoubtedly, consciousness was not the only factor that brought the demise of behaviorism. A different line of attack, for example, came from the quarters of newborn modern linguistics on the issue of explanation of verbal behavior. In particular, Noam Chomsky's famous review (1959) of Skinner (1957) is a milestone that shook behaviorism (in psychology) in its foundations. For an influential critique of logical behaviorism (in philosophy), see Putnam (1963).

31. Of course, there were a few exceptions who spoke up while the reign of behaviorism was still tight and proved to be visionaries. Worth mentioning here is a

lengthy discussion Miller gave on consciousness in his excellent survey of psychology as early as 1962. It is possible to recount even earlier attempts to break the silence, and directly or indirectly talk about consciousness, especially in the fields of attention, learning, and cybernetics. Cf. Hebb (1949), Abramson (1951–55), Hilgard (1956), and Broadbent (1958). See also Hilgard's remarks on this issue in his lucid survey, "Consciousness in Contemporary Psychology" (1980). For an account of the "cognitive revolution" in psychology, see Baars (1985) and Hilgard (1987), chap. 7.

32. Years later, Alan Baddeley (1993), a prominent psychologist who has devoted his career to the investigation of memory, validates Mandler's insight in the following words: "I am rather surprised to find myself writing about consciousness.... There are very good reasons why the study of consciousness has been discretely ignored by cognitive psychology during its early years of development.... Why, then, have I changed my mind? In my own case, the strongest reason has come from the pressure of empirical evidence; I am an experimental psychologist who uses empirical data to drive theory, and it has become increasingly difficult to have a model of memory that is at all complete, without directly or indirectly including assumptions about consciousness" (pp. 11–13). (Note, by the way, that Baddeley 1990—in many respects a very thorough book on memory—contains no references to consciousness.)

33. The information-processing models of consciousness, although not the only game in town, are still very much alive today, in psychology as well as in philosophy. For instance, Dennett's central claim in his most recent *Consciousness Explained* is that "conscious human minds are more-or-less serial virtual machines implemented—inefficiently—on the parallel hardware that evolution has provided for us" (Dennett 1991, p. 218). See also Hofstadter (1979), Harnad (1982), and Sommerhof (1990, 1996) for theorizing about consciousness in computational and systems-analysis terms. However, information-processing models of the mind (and, a fortiori, of consciousness in particular) have not always been everyone's favorite. For example Hubert Dreyfus, in his well-known critique of the research program and methodology of artificial intelligence, brought the whole information-processing approach under severe criticism (Dreyfus 1979, 1992, esp. chap. 4, "The Psychological Assumption"). Another line of attack was developed from neighboring quarters by philosopher John Searle. In an essay that later became known as the "Chinese room argument," Searle argued that no amount of information processing could alone provide a system with original (as opposed to derivative, assigned, etc.) semantics (Searle 1980). Dreyfus's critique never focused on consciousness per se, but Searle, in a newer work, deals exclusively with the problem of consciousness in cognitive science, and in general computational paradigms (Searle 1992). In contrast, a rival account of consciousness built entirely on computational ideas can be found in Jackendoff's *Consciousness and the Computational Mind* (1987). For a predecessor of the information-processing accounts of consciousness, see Donald Hebb's *The Organization of Behavior* (1949), a work that came out of the behaviorist era but anticipated what was ahead with foresight: Hebb argues to identify consciousness "theoretically with a certain degree of complexity of phase sequence in which both central and sensory facilitations merge, the central acting to reinforce now one class of sensory stimulations, now another" (p. 145).

34. Note, however, that a curious passage in Descartes's *Principles of Philosophy* suggests a theoretical commitment to something very much like Freud's unconscious, which does not sit squarely with his explicit commitment to the transparency of the mind:

> The strange aversions of certain people that make them unable to bear the smell of roses, the presence of a cat, or the like, can readily be recognized as resulting simply from their having been greatly upset by some such object in the early years of their life.... And the smell of roses may have caused severe headache in a child when he was still in the cradle, or a cat may have terrified him without anyone noticing and without any memory of it remaining afterwards; and yet the idea of an aversion he then felt for the roses or for the cat will remain imprinted on his brain till the end of his life (Descartes 1992), p. 195: *Principles of Philosophy*, pt. I, §9, AT, 429)

Unfortunately, I have not been able to find any further elaboration of this idea in Descartes's writings, which would surely be relevant in better understanding the nature of what seems to be an apparent theoretical tension.

35. The controversy over the status of unconscious mental states is multifaceted. Another staunch critic of the unconscious, though for reasons different from Titcheners' (that have to do with his stance against panpsychic views of consciousness), was William James (1950a). He fretfully remarks: "The distinction ... *between the unconscious and the conscious being of the mental state* ... is the sovereign means for believing what one likes in psychology, and of turning what might become a science into a tumbling-ground for whimsies" (p. 163). As discussed in section XI, a different line of objection is also raised, this time in a Cartesian spirit, by Searle (1992) and Strawson (1994).

Note also that in Titchener, talk about the *unconscious* has switched to talk about the *subconscious*, but there is enough reason to think that nothing theoretically significant hangs on this implicit substitution. This terminological variation stems from the fact that Freud and his contemporary, Pierre Janet, had an initial disagreement that left them with two different terms (*unconscious* and *subconscious*), and each one adopted and perpetually owned his own term with a vengeance. But this was more a result of personal quarrels between the two personalities than a genuine theoretical dissonance on the nature and structure of *that which is not conscious*. And so far as I can tell, there is no evidence that Titchener's use of Janet's term, *subconscious*, rather than Freud's *unconscious* is the result of a "conscious decision" and a theoretical commitment. For an illuminating account of the relation between Freud and Janet, see Perry and Laurence (1984).

36. For an interesting discussion of the question of whether the Freudian unconscious is a "theoretical construct" on a par with scientific theoretical entities, see Dilman (1972).

37. Note that over time, Freud grew dissatisfied with his tripartite structure and eventually introduced the new elements of the id, the ego, and the superego into the picture:

> In the further course of psycho-analytic work, however, these distinctions (i.e., conscious, preconscious, and unconscious) have proved to be inadequate, and for practical purposes, insufficient. This has been clear in more ways than one; but the decisive instance is as follows. We have formed the idea that in each individual there is a coherent organization of mental processes; and we call this *ego*. (Freud 1962, p. 7)

Later Freud (1964) gives a schematic depiction of the structure of consciousness, with the id, the ego, and the superego being "superimposed" on the classical tripartite division of the conscious, the preconscious, and the unconscious.

38. Cf. John Kihlstrom's work for a cognitivist overview of the various forms of the unconscious (e.g., Kihlstrom 1984, 1987).

39. For a thoughtful discussion of the theoretical issues involved, see Reingold and Merikle (1990).

40. Another paradigm in contemporary psychology that makes use of the conscious-unconscious distinction is that of implicit learning and implicit memory, as well as implicit perception. The focus of interest in this paradigm is on measuring the amount of learning and memory possible in the absence of subjects' awareness of the stimuli presented to them. A certain branch of this work became sensationalized in the media under the title "subliminal perception" in the 1970s. For a thorough and sympathetic account of the nature of this phenomenon, as well as the history of related research, see Dixon's *Subliminal Perception: The Nature of a Controversy* (1971) and his later *Preconscious Processing* (1981). For possibly the most influential recent work in this area (especially in masking studies), see Marcel (1983a, 1983b).

Naturally, there are also skeptics. For instance, Eriksen stated quite early on, "At present there is no convincing evidence that the human organism can discriminate or differentially respond to external stimuli that are at an intensity level too low to elicit discriminated verbal report. In other words, a verbal report is as sensitive an indicator of perception as any other response that has been studied" (Eriksen 1960, p. 298).

More recently, Holender (1986) presented a negative and rather controversial statement on subliminal perception, which also included a comprehensive survey of the field. For a collection of contemporary position papers in this paradigm, see the special issue of *Mind and Language* on "Approaches to Consciousness and Intention" (Spring 1990).

41. Searle also has an explanation to offer regarding the motivations underlying the sort of separationist view that Fodor promotes with respect to consciousness and intentionality:

> There has been in recent decades a fairly systematic effort to separate consciousness from intentionality. The connection between the two is being gradually lost, not only in cognitive science, but in linguistics and philosophy as well. I think the underlying—and perhaps unconscious—motivation for this urge to

> separate intentionality from consciousness is that we do not know how to explain consciousness, and we would like to get a theory of the mind that will not be discredited by the fact that it lacks a theory of consciousness. (Searle 1992, p. 153)

Perhaps a piece of careful Freudian psychoanalysis would resolve this issue for good. Lacking such expertise, I choose to leave the question open.

42. Panpsychism is a deep-rooted idea that can probably be traced, in one form or another, back to Thales and other ancient Greek philosophers. Nagel (1988) presents a contemporary discussion of panpsychism, characterizing it as the view that "the basic physical constituents of the universe have mental properties, whether or not they are parts of living organisms" (p. 181). Panpsychism was quite popular as a metaphysical doctrine among the psychologists (in particular, the psychophysicists) of the nineteenth century, including such prominent figures as Gustav Fechner and Hermann Lotze. William James, in contrast, was never sympathetic to this view; chapter VI of James (1950a) contains a cogent critique of panpsychism (under the title "Mind-Stuff Theory"). For recent discussions of panpsychism in the context of ongoing consciousness debates, see Seager (1995) and Hut and Shepard (1996).

43. The origins of the term *raw feels* goes back, so far as I can trace, to the work of behaviorist psychologist Edward Tolman. In outlining what falls outside the scope of "scientific psychology," Tolman (1967) characterizes raw feels (from his opponents' perspective) as follows: "Sensations, says the orthodox mentalist, are more than discriminanda-expectations, whether indicated by verbal introspection or by discrimination-box experiments. They are in addition immediate mental givens, 'raw feels'. They are unique subjective suffusions in the mind" (Tolman 1967, pp. 250–251). But it is probably Herbert Feigl (1967) who is responsible for the introduction and wide acceptance of *raw feels* in the philosophical terminology.

44. In a short passage in the *Tractatus*, Wittgenstein makes a similar point regarding the self (or subjecthood):

> *5.633* Where in the world is a metaphysical subject to be found?
> You will say that this is exactly like the case of the eye and the visual field. But really you do *not* see the eye.
> And nothing *in the visual field* allows you to infer that it is seen by an eye.
> *5.6331* For the form of the visual field is surely not like this:

(Wittgenstein 1974, p. 57)

Keith Gunderson (1970) also discusses this issue under the title, "The Investigational Asymmetries Problem" and makes the similar point that "just as the eye does not, cannot, see itself in its own visual field, so too, the self will never, in its inventory-taking of the world, find itself in the world in the manner in which it finds other people and things" (p. 127).

Again, the point raised is well taken for consciousness so far as one's own selfhood is involved in it, but it is not obvious just how it generalizes into a difficulty (much less an impossibility) with the study of consciousness in general by (other) conscious beings.

45. This observation is intended to be ontologically neutral. The emphasis here is on the "mode of access" part and not on the "facts" themselves. In particular, it does not entail the existence of a special class of facts, "first-person facts," on the basis of an assumption of ontological difference between facts of one's own consciousness and those of others.

46. Of course, technically speaking, it is not possible to digest food in someone else's stomach either, but "digestive epistemology" just does not seem to be a fashionable topic these days.

47. Perhaps the most succinct expression of this difference is given in Sydney Shoemaker's question: "If what I want when I drink fine wine is information about its chemical properties, why don't I just read the label?" (quoted in Dennett 1991, p. 383). There is a ready-made answer to this question: It indeed *is* information about the chemical properties of the wine that a connoisseur is interested in, but only if that information can be accessed in a certain sensory modality—gustatorily, not visually. Put in Fregean (1892) terms, reading the label and sipping the wine would provide access to the same referent via different "modes of presentation." Note, however, that Shoemaker's question remains not fully addressed until this answer is supplemented by a satisfactory account of something akin to modes of presentation regarding qualia.

48. Of course, not all publicly observable properties of an experience are intrinsic. There are often a great many extrinsic properties that determine what the experience is about that are equally accessible to the experiencer and the observer. Some think that all important properties of experiences, including those that determine an experience's phenomenal character, are extrinsic. See Dretske (1995).

49. There is a spectrum of positions with respect to these dichotomies that yield deep differences in the metaphysics of consciousness. Let me mention a few exemplary positions. Nagel (1979, 1986) takes the distinction between subjective and objective points of view as fundamental to important philosophical problems, such as personal identity, free will, and the mind-body problem. Velmans (1991) posits that first-person and third-person accounts of consciousness are complementary, but not reducible, to one another. In contrast, Dretske (1995) argues that as a "result of thinking about the mind in naturalistic terms, subjectivity becomes part of the objective order. For materialists, this is as it should be" (p. 65). This is in accord with an earlier statement by Lashley (1923), who claims that "the subjective and objective descriptions are not descriptions from two essentially different points of view, or descriptions of two different aspects, but simply descriptions of the same thing with different degrees of accuracy and detail" (p. 338). Papineau (1993) argues that it is a mistake to think that first-person and third-person thoughts refer to different entities on the basis of an epistemic difference, and calls it the "antipathetic fallacy." Finally, Perry (1979, 1993) examines the status of the first-person in relation to the role of indexicality in mind and language. Also, for two alternative approaches, see Hut and Shepard (1996) for a prioritization of the first-person over the third-person, and Smith (forthcoming) on how to get to the third-person from the first-person. For a scrutiny of the metaphysical foundations of such dichotomies as objectivity versus subjectivity, see among others Goodman (1978), Rorty (1979), Putnam (1981), and Smith (1996).

50. For a lucid analysis of James's account of the structure of "fringe consciousness," see Mangan (1993). Regarding works on the structure of phenomenal consciousness, continental Europe was certainly more of a center than James's Cambridge. See, for instance, Brentano's chapter, "On the Unity of Consciousness," in his *Psychology from an Empirical Standpoint* (Brentano 1874), a work that slightly precedes James's *Principles of Psychology* (1890). A more detailed analysis of this sort was later given by Husserl. See, among others, his *Ideas: General Introduction to Pure Phenomenology* (1913) and *The Phenomenology of Internal Time-Consciousness* (1928). A more recent attempt along these lines, which comes from the analytic tradition, can be found in Searle (1992, chap. 6).

51. The idea of finding systematical bridging relations between the "mental" and the "physical" in order to establish explanatory hooks on consciousness was also the driving factor behind the emergence of psychophysics as a research program in the nineteenth century. This is exemplified in, for instance, Gustav Fechner's work where he sought ways of formalizing a logarithmic relation between the intensity of physical stimuli (measured in "physical units"), and the magnitude of felt sensory experience (measured in "psychological units") as reported by the subject. (Cf. Fechner 1966, see also Boring 1942 and Hilgard 1987, chap. 4.) The same idea was also operative in Gestalt psychology in the hypothesizing of a relation of isomorphism between "the structural characteristics of brain processes and of related phenomenal events" (Köhler 1971, p. 81; see also Köhler 1980 and Boring 1929, chap. 22).

52. John Searle presents a major attack on behaviorist theories of consciousness in his *Rediscovery of the Mind*, arguing for what he dubs "the principle of the independence of consciousness and behavior." His thesis is that "the capacity of the brain to cause consciousness is conceptually distinct from its capacity to cause motor behavior." One consequence he draws from this thesis is that "a system could have consciousness without behavior." Under certain qualifications, I find this view plausible. However, Searle goes further to claim that "*ontologically speaking, behavior, functional role, and causal relations are irrelevant to the existence of conscious mental phenomena*" (Searle 1992, p. 69; emphasis in original). This further and more encompassing claim does not directly follow from the weaker one. Moreover, it opens up a path the logical conclusion of which may turn into "epiphenomenalism": the view that consciousness plays no causal role itself, though it may be the causal effect of other phenomena. I find Searle's second thesis untenable and its consequence very

undesirable. I will come back to this issue in my discussion of epiphenomenalism and the possibility of zombies.

53. There are also scientifically documented cases where the relation between consciousness and externally observable behavior breaks down. I briefly discussed the phenomenon of "gaining consciousness' while under general anesthesia as one example. There are also several diseases of the nervous system that fall in the category of demyelinating neuropathies (diseases that result from loss of conduction of nerve impulses due to the lack of formation of myelin, a fatty substance essential to the insulation of axons in neurons), which result in the patient's gradual loss of reflexes and muscular strength, and hence behavior, while not resulting in substantial sensory changes (e.g., the Guillain-Barré syndrome). The ultimate state of such a patient involves very little outward behavior with no loss of consciousness, defying the behaviorist dogma. (Cf. Reeves 1981.) A moving account of a somewhat related nervous system disorder, encephalitis lethargica (commonly known as the sleeping sickness) was given in Oliver Sacks's popular book, *Awakenings*. Sacks's description of the victims of encephalitis lethargica is worth quoting at least for its literary value:

> Patients who suffered but survived an extremely severe somnolent/insomniac attack of [encephalitis lethargica] often failed to recover their original aliveness. They would be conscious and aware—yet not fully awake; they would sit motionless and speechless all day in their chairs, totally lacking energy, impetus, initiative, motive, appetite, affect or desire; they registered what went on about them with profound indifference. They neither conveyed nor felt the feeling of life; they were as insubstantial as ghosts, and as passive as zombies: von Economo compared them to extinct volcanoes. [However, . . .] one thing, and one alone, was (usually) spared amid the ravages of this otherwise engulfing disease: the "higher faculties"—intelligence, imagination, judgement, and humour. These were exempted—for better or worse. Thus these patients, some of whom had been thrust into the remotest or strangest extremities of human possibility, experienced their states with unsparing perspicacity, and retained the power to remember, to compare, to dissect, and to testify. Their fate, so to speak, was to become unique witnesses to a unique catastrophe. (Sacks 1974, pp. 9, 12)

54. Let me also mention that some philosophers think that the only legitimate sense of consciousness is phenomenal consciousness (e.g., Searle 1992, Flanagan 1992), while others believe only in access consciousness (e.g., Dennett 1991), and still others believe in phenomenal consciousness but try to account for it in causal, functional, or representational (i.e. "access-related") terms (e.g., Van Gulick 1988, Tye 1992, Dretske 1995).

55. Some of these ideas are briefly explored in Güzeldere and Aydede (forthcoming).

56. I will discuss absent qualia, an offspring of philosophical imagination that was conceived as a result of taking the "hard problem" (perhaps too) seriously, in discussing zombies below. Regarding the explanatory gap, here is a surprisingly contemporary expression of the problem from the nineteenth-century philosopher-psychologist Charles Mercier: "The change of consciousness never takes place without the change in the brain; the change in the brain never ... without the change in consciousness. But *why* the two occur together, or what the link is which connects them, we do not know, and most authorities believe that we never shall and can never know" (Mercier 1888, p. 11).

Note that the point Mercier is raising is very similar to the one expressed by John Tyndall in section I. A similar but more recent statement, though in a more determinedly pessimist tone, can be found in a rather unlikely source. Here is Freud on the "hard problem":

> We know two things concerning what we call our psyche or mental life: firstly, its bodily organ and scene of action, the brain (or nervous system), and secondly, our acts of consciousness, which are immediate data and cannot be more fully explained by any kind of description. Everything that lies between these two terminal points is unknown to us and, so far as we are aware, there is no direct relation between them. If it existed, it would at the most afford an exact localization of the processes of consciousness and would give us no help toward understanding them. (Freud 1949, pp. 13–14)

In contemporary philosophy of mind, Nagel's formulation of this problem has been most influential. The difficulties Nagel raises with respect to "bridging the explanatory gap" between things physiological and things phenomenal are also reflected in Kripke's attack against identity theory. Even though the latter follows a different path, using tools from philosophy of language, they arrive at very similar conclusions. (Cf. Nagel 1974, 1979, 1986, and Kripke 1980.) The following quotation eloquently summarizes Nagel's (and presumably, Kripke's) position:

> We cannot directly see a necessary connection, if there is one, between phenomenological pain and a physiologically described brain state any more than we can directly see the necessary connection between increase in temperature and pressure of a gas at a constant volume. In the latter case the necessity of the connection becomes clear only when we descend to the level of molecular description: till then it appears as a contingent correlation. In the psychophysical case we have no idea whether there is such a deeper level or what it could be; but even if there is, the possibility that pain might be necessarily connected with a brain state at this deeper level does not permit us to conclude that pain might be directly analyzable in physical or even topic-neutral terms. (Nagel 1986, pp. 48–49)

57. Searle's position is not as straightforward as Nagel's, however. Although Searle talks about subjectivity as an irreducible ontological property unique to consciousness, he also maintains the following position, in a somewhat puzzling way, in the same book: "Consciousness is, thus, a biological feature of certain organisms, in *exactly the same sense* of 'biological' in which photosynthesis, mitosis, digestion, and reproduction are biological features of organisms. . . . One of the main aims of this book is to try to remove that obstacle, to bring consciousness back into the subject matter of science as a biological phenomenon *like any other*" (Searle 1992, pp. 93, 95; emphasis added).

58. As Wittgenstein somewhat sarcastically remarks:

> The feeling of an unbridgeable gulf between consciousness and brain-process: how does it come about that this does not come into the considerations of our ordinary life? This idea of a difference in kind is accompanied by slight giddiness. . . . When does this feeling occur in the present case? It is when I, for example, turn my attention in a particular way on to my own consciousness, and, astonished, say to myself: THIS is supposed to be produced by a process in the brain!—as it were clutching my forehead. But what can it mean to speak of "turning my attention on to my own consciousness"? This is surely the queerest thing there could be! (Wittgenstein 1958, §412, p. 124e)

59. Searle says: "For any artefact that we might build which had mental states equivalent to human mental states, the implementation of a computer program would not by itself be sufficient. Rather, the artefact would have to have powers equivalent to the powers of the human brain" (Searle 1984, p. 41).

Searle's argument is against functionalist accounts of consciousness. Even though he should not be taken to commit himself to a single specific underlying substance (i.e., neuronal structures), he nonetheless seems to think that consciousness is not medium-independent, at least so far as the causal powers of the medium go. This view seems to be based on the implicit assumption that there are causal powers that cannot be captured by functional organization, but it unfortunately leaves the central notion of causal power unexplicated.

60. Here is a noteworthy historical fact: Over a hundred years before Crick and Koch presented their findings on the 40–70 Hz phenomena, in 1879, Payton Spence published an essay in which he argued, on purely metaphysical grounds, that the basic form of consciousness consists of a constant alteration of conscious and unconscious states. But the alteration is so rapid that the subject never becomes aware of the discrete nature of her consciousness; she is under the illusion of having a continuous stream. Spence then speculated that there must be an underlying mechanism in the brain that is responsible for this alteration—something like a very rapid oscillation of neural tissue. In his own words:

> The simplest form of consciousness, or mental life, must consist in an alteration of a state of consciousness with a state of unconsciousness—a regular rhythmical revelation of the Affirmation, consciousness, by its Negation, unconsciousness, and *vice versa*. . . . Perhaps it would be safer, for the present, to call it a pulsation, or an undulation in the brain, or a vibration of the molecules of the brain, paralleled in consciousness. This pulsation or vibration is, of course, very rapid; otherwise, we would not have to infer its existence, but would know it by perceiving the alterations of one state with another. (Spence 1879, p. 345)

The interesting part comes when M. M. Garver, a neurophysiologist of the same era, finds the idea plausible and follows up on it on experimental grounds. In particular, he investigates the neural basis of voluntary action (often associated with or regarded as an aspect of

consciousness in those times, by psychologists including William James) and publishes his results in the *American Journal of Science* in 1880. According to Garver, mental activity is subserved by a cerebral oscillatory mechanism with a frequency range of 36–60 Hz. Garver hypothesizes that the change in the frequency of the neural oscillations correlates with minimum and maximum levels of mentation, which results in voluntary action. Garver formulates his hypothesis as follows: "The cerebral portion of the nervous system is continually varying in its activity, waxing and waning between certain limits, periods of maximum activity following periods of minimum activity at the rate of 36 to 60 times per second" (Garver 1880, p. 190).

At the end of his article, Garver claims that this pattern can be extended to accommodate Spence's hypothesis of alternating states of consciousness and suggests that the lower and upper limits of the oscillation frequency can be taken as the correlates of consciousness and unconsciousness, respectively.

While Spence and Garver cannot perhaps be said to be in pursuit of a solution to the binding problem, and thus have anticipated the Crick and Koch hypothesis over a hundred years ahead of its time, I find the similarity in the basic idea of seeking a neural oscillatory basis for consciousness fascinating. Except a passing remark by William James in his discussion of the continuity of consciousness (James 1950, p. 220, footnote), Spence and Garver's work seems to have gone, so far as I could trace, unnoticed to date.

61. The idea of localized functions in the brain precedes James's work, and goes back at least to the once-too-popular phrenology of Franz Gall in early 1800s (see, for instance, Ackerknecht and Vallois 1956). Jean Baptiste Bouillaud, in 1825, proposed a hemispheric asymmetry in brain function, but it was Carl Wernicke and Paul Broca who made the greatest contribution to the idea of modularity in the brain. Wernicke hypothesized that two particular areas in the left hemisphere of the brain (roughly, the left frontal lobe and the posterior cortex), which later became known as Broca's and Wernicke's areas, were responsible for language production and language understanding, respectively. For an elemental neuropsychological account of aphasias that result from damage to these areas and other related matters, see Kolb and Whishaw (1990).

62. For a comprehensive survey of similar neuropsychological disorders, see Farah (1995). I discuss some philosophical issues involved in the phenomenon of blindsight against the background of Block's access versus phenomenal consciousness distinction in Güzeldere (1995e).

63. The precursors to Dennett's *Cartesian Theater* metaphor can be found in the writings of Gilbert Ryle and U. T. Place. Ryle (1949) characterized and criticized the Cartesian notion of the mind as "a secondary theater in which the episodes enacted enjoy the supposed status of 'the mental'" (p. 158). Similarly, Place (1956) called it a mistake to suppose that "when the subject describes his experience, when he describes how things look, sound, smell, taste, or feel to him, he is describing the literal properties of objects and events on a peculiar sort of internal cinema or television screen, usually referred to in the modern psychological literature as the 'phenomenal field'" (p. 107). (The view that Ryle and Place criticize also constitutes a particular family of *sense-data* theories of perception that were quite popular at the time.)

Interestingly, while the metaphor of mind as an inner theater never occurs explicitly in the writings of Descartes (so far as I could tell), it can be found in a vivid passage in Hume (in his discussion of personal identity): "The mind is a kind of theatre, where several perceptions successively make their appearance; pass, re-pass, glide away, and mingle in an infinite variety of postures and situations." But Hume is also careful not to endorse, in virtue of using this metaphor, the kind of ontological conclusion Descartes is criticized as holding: "The comparison of the theatre must not mislead us. They are the successive perceptions only, that constitute the mind; nor have we the most distant notion of the place, where these scenes are represented, or of the materials, of which it is compos'd." (Hume 1955, Book I, IV:V, 85).

64. Another issue related to the Where question has to do with lateralization of brain function in light of the "split-brain" research of the last few decades. The performance of commissurotomy on humans (proposed and initiated by surgeon Joseph Bogen in 1960) to control interhemispheric spread of epilepsy produced a number of patients in whom individual investigation of specialized hemispheric capabilities became possible. The research on such patients (initially pursued by psychobiologist Roger Sperry and his collaborators) revealed a number of interesting facts about hemispheric specialization and resulted in a sizeable scientific literature, as well as a huge corpus of popular psychology writing on the so-called left-brain versus right-brain

distinction with regard to personality types, social behavior, and so on. One of the major results that came out of the commissurotomy research is the hypothesis proposed and defended by (among others) Michael Gazzaniga (1993), that human cognition as well as consciousness (in the sense of awareness of experience) are subserved by special brain circuitry normally located in the left hemisphere. For an account of the early work on commissurotomy, see Gazzaniga (1970). Galin (1974) explores the implications of hemispheric specialization for psychiatry. For a comprehensive collection of current research results in human neuropsychology, including articles on modularity of mental function and hemispheric specialization, see Gazzaniga et al. (1995).

65. It is interesting to note that this is a junction at which some upholders of the "pro-qualia intuition" meet on common ground with the most indoctrinated qualia skeptics, such as the behaviorist psychologists for whom exorcising qualia out of the scope of psychology was a primary goal. Notice, for instance, the similarity between Jackson's (1982) position (who characterizes himself as a "qualia freak," p. 127), and the position defended by Edward Tolman: "[Regarding visual perception in others] we never learn whether it 'feels' like our 'red' or our 'green' or our 'gray', or whether, indeed, its 'feel' is perhaps sui generis and unlike any of our own.... Whether your 'raw feels' are or are not like mine, you and I shall never discover. Your color 'feels' may be the exact complementaries of mine, but, if so, neither of us will ever find it out, provided only that your discriminations and my discriminations agree.... If there be 'raw feels' correlated with such discriminanda-expectations, these 'raw feels' are by very definition 'private' and not capable of scientific treatment. And we may leave the question as to whether they exist, and what to do about them, if they do exist, to other disciplines than psychology—for example, to logic, epistemology, and metaphysics. And whatever the answers of these other disciplines, we, as mere psychologists, need not be concerned" (Tolman 1967, pp. 252–253).

Whether this surprising "meeting of minds" between such arch-opponents as behaviorist psychologists and a certain brand of qualia defenders on presumably the very point of contention between them speaks in favor of the former or the latter party (if either), I leave open to the judgment of the reader.

66. William James (1950a), who characterizes consciousness as a "fighter for ends," makes the following statement (largely in agreement with Herbert Spencer 1891, 1898) in support of the causal construal and the evolutionary relevance of consciousness:

> *It is a well-known fact that pleasures are generally associated with beneficial, pains with detrimental experiences....* These coincidences are due, not to any pre-established harmony, but to the mere action of natural selection which would certainly kill off in the long-run any breed of creatures to whom the fundamentally noxious experience seemed enjoyable. An animal that should take pleasure in a feeling of suffocation would, if that pleasure were efficacious enough to make him immerse his head in water, enjoy a longevity of four or five minutes. But if the pleasures and pains have no efficacy, one does not see why the most noxious acts, such as burning, might not give thrills of delight, and the most necessary ones, such as breathing, cause agony.... The conclusion that [consciousness] is useful is ... quite justifiable. But, if it is useful, it must be so through its causal efficaciousness.

James's conclusion is that "the study *a posteriori* of the *distribution* of consciousness shows it to be exactly such as we might expect in an organ added for the sake of steering a nervous system grown too complex to regulate itself" (pp. 141, 143–144).

67. See also Kitcher (1979) and Revonsuo (1994) for related points.

68. Precursors to Nagel's thinking on this issue can be found in the writings of B. A. Farrell and Timothy Sprigge. Even though it was made famous by Nagel, the original formulation of the question "what is it like to be a bat?" goes back to Farrell's somewhat neglected essay, "Experience" (1950). In discussing the issue of experiential knowledge, Farrell imagines a Martian visitor about whose sensory capacities we obtain all the information there is. According to Farrell, "We would probably *still* want to say: 'I wonder what it would be like to be a Martian.'" He continues: "There *is* something more to be learned about the Martian, and that is what his experience is like." Farrell then extends the question to babies and mice, as well as an opium smoker, and finally a bat: "I wonder what it would be like to be, and hear like, a bat" (pp. 34–35).

The lessons Farrell draws out of his ruminations are quite the opposite of Nagel's conclusions, however. It is rather Sprigge (1971) who makes the connection between the "what it is like" aspect of experience and

physicalism's difficulty with accommodating it in the particular way Nagel problematizes the issue: "When one imagines another's conscious state, there is no conclusive way of checking up whether one has done so correctly or not.... Presuming that the object (that is, at least normally, the organism) with which one is concerned, is indeed conscious, then *being that organism* will have a certain definite complex quality at every waking moment.... Physical science makes no reference to qualities of this kind. Thus consciousness is that which one characterises when one tries to answer the question what it is or might be like to *be* a certain object in a certain situation" (p. 168).

69. Lycan (1997b) also makes a similar point: "The phrase 'what it's like' is more sinning than sinned against; nothing whatever is clarified or explained by reference to it, and it itself is not only badly in need of explanation, in general, but at least three-ways ambiguous in particular" (p. 176).

70. See Nagel (1974, 1979, 1986) for a full range of problems that involve subjectivity. See Lycan (1990), Biro (1991), Akins (1993), and Dretske (1995), among others, for deflationary responses. See also Nagel (1983) for a discussion of how subjectivity figures in the problem of self, without ever touching on qualia, and Perry (forthcoming) for a penetrating analysis and critique of Nagel's account.

71. For Jackson's formulation, see Jackson (1982, 1986). For various critiques, see Nemirow (1980), Churchland (1989), Lewis (1990), Dennett (1991, chap. 10), Van Gulick (1993), Loar (1990), Harman (1993b), Dretske (1995, chap. 3), and Perry (1995). For a related empirical study on the conceptual representation of colors in the blind and the color-blind, see Shepard and Cooper (1992).

One of the earlier formulations of the knowledge argument can be found in C. D. Broad's thought experiment about the archangel who knows all about chemistry but lacks the sense of smell. Broad sets up the problem as follows:

> Would there be any theoretical limit to the deduction of the properties of chemical elements and compounds if a mechanistic theory of chemistry were true? Yes. Take any ordinary statement, such as we find in chemistry books; e.g., "Nitrogen and Hydrogen combine when an electric discharge is passed through a mixture of the two. The resulting compound contains three atoms of Hydrogen to one of Nitrogen; it is a gas readily soluble in water, and possessed of a pungent and characteristic smell." If the mechanistic theory be true the archangel could deduce from his knowledge of the microscopic structure of atoms all these facts but the last. He would know exactly what the microscopic structure of ammonia must be; but he would be totally unable to predict that a substance with this structure must smell as ammonia does when it gets into the human nose. The utmost that he could predict on this subject would be *that* certain changes would take place in the mucous membrane, the olfactory nerves and so on. But he could not possibly know that these changes would be accompanied by the appearance of a smell in general or of the peculiar smell of ammonia in particular, unless someone told him so or he had smelled it for himself. (Broad 1962, p. 71)

Similarly, an early formulation of the knowledge argument, as well as an antecedent of the Nemirow-Lewis critique, appears in Feigl's discussion of "cognitive roles of acquaintance." Feigl asks: "What is it that the blind man cannot know concerning color qualities?" and proposes the following answer:

> If we assume complete physical predictability of human behavior, i.e., as much predictability as the best developed physical science of the future could conceivably provide, then it is clear that the blind man or the Martian would lack only *acquaintance* and *knowledge by acquaintance* in certain areas of the realm of qualia. Lacking acquaintance means not having those experiential qualia; and the consequent lack of knowledge by acquaintance simply amounts to being unable to label the qualia with terms used previously by the subject (or by some other subject) when confronted with their occurrence in direct experience. Now, mere *having* or *living through* is not *knowledge* in any sense. "Knowledge by acquaintance," however, as we understand it here, is propositional, it does make truth claims.

Feigl then goes on to suggest, anticipating many of the present-day critiques of the knowledge argument, that the blind person (or Jackson's fictional color scientist, Mary) does not lack any knowledge per se; all he or she lacks is a particular mode of knowing the same facts as do normally sighted people: "What one person *has* and *knows by acquaintance* may be identical with what someone else knows by description. The color experiences of the man who can see are known to him by acquaintance, but the blind man can have inferential knowledge, or knowledge by description *about* those same experiences" (Feigl 1967, p. 68). A related puzzle, based again on a thought-experiment, was posed as a "jocose problem" to John Locke by William Molyneux, an amateur philosopher, in a letter dated 1693: "Suppose a man *born* blind, and now adult, and taught by his *touch* to distinguish between a cube and a sphere of the same metal, and nighly of the same bigness, so as to tell, when he felt one and the other, which is the cube, which the sphere. Suppose then the cube and sphere placed on a table, and the blind man be made to see: *quaere*, whether *by his sight, before he touched them*, he could now distinguish and tell which is the globe, which the cube?" (Quoted in Locke (1959), Book II, Chapter IX, pp. 186–187.) As such, Molyneux's question transforms the inquiry of whether non-experiential facts can yield knowledge of experiential facts (in the knowledge argument) into a puzzle about *intersensory* translation—whether tactile facts can yield knowledge of visual facts ("facts" taken broadly). Locke's negative answer to this question was in agreement with Molyneux's opinion. But others disagreed, and Molyneux's question became one of the central topics of contention among such philosophers as Berkeley, Leibniz, Reid, Diderot, and Voltaire, in the greater context of the controversies over innateness and abstract ideas. See Morgan (1977) and Sanford (1983) for further exposition and discussion of the related issues.

72. For a classical treatment of the identity theory, see Sellars (1964). For recent arguments in its defense, see Enç (1983) and Hill (1991). A thus far unmentioned but related concept that has played a significant role in the philosophy of mind over the past few decades is *supervenience* (presumably imported from Rom Hare's work in ethics [1952] by Donald Davidson 1970). The thesis that the mental supervenes on the physical is put forth as a better materialist solution to the mind-body problem than the identity thesis. The supervenience thesis is roughly that the mental character of a state or event is wholly determined by its physical profile, such that there cannot be a change in the former without a change in the latter. Put differently, sameness in the (subvening) physical properties is hypothesized to guarantee sameness in the (supervening) mental properties.

But the status of the supervenience thesis vis-à-vis the mind-body problem, especially with regard to phenomenal consciousness, remains controversial. Are supervenient properties ontologically distinct (autonomously emergent) from physical properties, for instance? Or (otherwise) does supervenience boil down to old-fashioned identity? How are we to explain the *nature* of the supervenience relation itself? Given the wide variety of possible supervenience relations (for example, weak versus strong, local versus global), there is a vast and technically complicated literature in this area, but there exist no clear-cut received views that are taken to unanimously answer all these questions. For a systematic and thorough exploration of the supervenience thesis in all its different characterizations, see Jaegwon Kim's essays collected in Kim (1993). Two other useful collections that contain representatives of contemporary theorizing on supervenience are Beckermann, Flohr, and Kim (1992), and Savellos and Yalçın (1995). See also McLaughlin (1989) for a discussion that relates supervenience to the question of epiphenomenalism, and McLaughlin (1992) for a thorough exposition of the thought of British emergentists in which one can find numerous clues for the present supervenience debates. In addition to philosophy of mind, notions of supervenience, emergence, and different kinds of reduction have been central to discussions in philosophy of biology, particularly during the first half of this century. See Brandon (1996) for an exposition of these issues that also ties them to debates in contemporary biology, and Harris (1993) for an exploration of the relations between the natures and the study of mind and life in philosophy, psychology, and biology.

73. It is interesting to note, by the way, that the word *consciousness* was hardly ever present in the philosophical literature around the time of functionalism, which was instead brewing with the term *qualia* and, to a lesser extent, terms like *raw feels* and *phenomenal aspect of consciousness*. In the 1990s, in contrast, many of the same old problems have gained new interest and impetus (perhaps from slightly but essentially similar perspectives) within a terminology populated with the magic word *consciousness* and its derivatives. The reasons for this terminological change probably lie partly outside philosophy, for example in the wide acceptance of the term *consciousness* into other fields with which philosophy interacts.

74. There is a sizeable literature regarding the absent qualia argument. Some of the most influential thought experiments that support the possibility of absent qualia are due to Block (1978, 1980b) and Block and Fodor (1980). See Shepard (1993) for a recent discussion in the

context of color vision and evolutionary theory. Shoemaker (1975, 1981a) presents an eloquent defense of functionalism against the absent qualia arguments. See also Dretske (1995) who raises the possibility of absent qualia in an externalist context in relation to Donald Davidson's (1987) "swampman" argument, and Tye (1995) for a representationalist critique of absent qualia. For further discussion, see Lycan (1981), Levin (1985), Graham and Stephens (1985), White (1986), Fox (1989), Levine (1989), Horgan (1987), and Hardcastle (1996).

One of the most commonly cited absent qualia arguments is based on Block's (1978) "Chinese Nation" scenario, designed to "embarrass all versions of functionalism" by showing that functionalism is guilty of "classifying systems that lack mentality as having mentality" (p. 275). Block asks us to imagine the functional simulation of a human brain by the Chinese nation by connecting each of the billion inhabitants of China in appropriate ways through radio links, and having them communicate from a distance like neurons in a brain and thereby animate an artificial body for a certain period of time. According to Block, while this China-body system is "nomologically possible" and "it could be functionally equivalent to [a human being] for a short time," it is doubtful "whether it has any mental states at all—especially whether it has ... 'qualitative states', 'raw feels', or 'immediate phenomenological qualities'" (pp. 276–278). Block's point is to establish the shortcomings of functional characterizations of qualia, by appealing to intuitions that he takes as common sensical, such as the intuition that such "distributed minds" are absurd.

A similar intuition was commonly employed in discussions regarding the unity of mind versus the divisibility of matter with the aim of embarrassing all forms of materialism in early modern philosophy. For example, the eighteenth century English theologian Samuel Clarke appeals to the absurdity of the distributed-minds intuition in a piece of hypothetical reasoning, similar to Block's, to make a case for the "immateriality and natural immortality of the soul" as follows: "That the soul cannot possibly be material is moreover demonstrable from the single consideration even of bare sense and consciousness itself. For suppose three, or three hundred, particles of matter, at a mile, or at any given distance, one from another; is it possible that all those separate parts should in that state be one individual conscious being?" But Clarke then takes his argument a step further to apply it to human beings: "Suppose then all these particles brought together into one system, so as to touch one another; will they thereby, or by any motion or composition whatsoever, become any whit less truly distinct Beings, than they were at the greatest distance? How then can their being disposed in any possible system, make them one individual conscious being?" (Clarke 1707, p. 82)

75. The origins of this problem go back indeed to a puzzle about visual experience, described by John Locke a few centuries ago:

> Neither would it carry any imputation of falsehood to our simple ideas, if by the different structure of our organs it were so ordered, that the *same object should produce in several men's minds different ideas* at the same time; e.g. if the idea that a violet produced in one man's mind by his eyes were the same that a marigold produced in another man's, and *vice versa*. For, since this could never be known, because one man's mind could not pass into another man's body, to perceive what appearances were produced by those organs: neither the ideas hereby, nor the names, would be at all confounded, or any falsehood be in either. For all things that had the texture of a violet, producing constantly the idea that he called blue, and those which had the texture of a marigold, producing constantly the idea which he constantly called yellow, whatever those appearances were in his mind; he would be able as regularly to distinguish things for his use by those appearances, and understand and signify those distinctions marked by the name blue and yellow, as if the appearances or ideas in his mind received from those two flowers were exactly the same with the ideas in other men's minds (Locke 1959, bk. II, chap. 32, §15, p. 520).

There are also several positions with respect to the inverted spectrum argument. Here is a simple set: Block (1978) and Block and Fodor (1980) raise the possibility of inverted spectrum against functionalism, and Block (1990) presents an original twist on the same problem as a reply to Harman (1990), who argues against inverted spectrum on externalist grounds. Shoemaker (1975, 1981b, 1991) is more lenient toward accepting the possibility of inverted spectrum compared to his rejection of the possibility of absent qualia, but he presents an argument on how to accommodate qualia inversion within a broadly functionalist framework. However, see also Levine (1988), who argues that absent qualia and inverted spectrum stand or fall together as logical possibilities against functionalism. Dretske (1995) regards inverted spectrum as a problem for functionalism but not for representationalism (and hence, not for materialism in general). Tye (1995) argues for the conclusion that is somewhat similar to Shoemaker's: spectrum inversion is possible in narrow functional duplicates, but this does not constitute a problem for wide functionalism (that Tye defends).

76. A similarly externalist theory is put forth by Gilbert Harman (1990) where qualia are identified with intentional properties, as well as by Michael Tye (1995). David Armstrong had earlier suggested identifying qualia (what he called "secondary qualities") with properties of physical objects in his discussion of "Realist Reductionism" (Armstrong 1993, chap. 12, 270–290). Clues for these positions can be found in Elizabeth Anscombe's discussion of the intentional nature of sensations (Anscombe 1965).

77. Perhaps a caveat about the particular brand of epiphenomenalism and zombiehood I am referring to is in order here. Epiphenomenalism about phenomenal consciousness is, of course, different from epiphenomenalism about the mind in general. Discussions about epiphenomenalism earlier this century generally assumed the latter kind (see, for instance, Broad 1962). In the contemporary literature, however, the focus has somewhat shifted. Probably largely due to the advent of functionalist and computational-representational theories of mind in a materialist framework, many today take intentional states, such as beliefs and judgments, as contentful internal structures in the brain (see, for example, Fodor 1987). As such, no one thinks that beliefs, *qua* such physical structures, lack causal properties. The controversy is rather on whether their content (semantics) has a causal role in the explanation of behavior (cf. Dretske 1988).

But the real hot spot of the epiphenomenalism debate has *to do with phenomenal consciousness—whether* qualia play any role in the otherwise causally characterizable economy of our mental lives. Note, after all, that while there is a vast literature on the possibility of absent qualia, no one seems to be worrying about the possibility of "absent beliefs" or "absent judgments." There seems to be a crucial difference between beliefs and pains: while it is considered legitimate to *attribute* beliefs to someone who behaves in ways that can be explained by belief-attributions of the relevant sort, it is considered very problematic to so attribute pains, because the essence of pains, the reasoning goes, is not attributable (by a third party) but rather *accessible* in a privileged way (through the first-person perspective). There is something it is like to have pains, but there is nothing it is like to believe that there is no greatest prime number (or even that one is in pain). It is this difference that warrants epiphenomenalism, in the present literature, as a possibility with respect to pains, but not beliefs (about prime numbers, one's pains, or anything else).

Accordingly, under the stipulation of appropriate environmental and historical conditions, it is generally regarded as a possibility that a physical replica of a human being can lack all qualia, while not lacking beliefs or judgments (or other such intentional states). This is the possibility of the modern zombie that has center stage in debates about phenomenal consciousness. Ned Block (1995) calls such replicas *phenomenal zombies*: "the familiar ... robots that think but don't feel" (p. 234). Or, as David Chalmers (1996) describes: "My zombie twin ... will be psychologically identical to me. He will be perceiving the trees outside, in the functional sense, and tasting the chocolate, in the psychological sense [similar to its ordinary twin *modulo* qualia]. ... He will be awake, able to report the contents of his internal states, able to focus attention in various places, and so on. It is just that none of this functioning will be accompanied by any real conscious experience. There will be no phenomenal feel. There is nothing it is like to be a zombie" (p. 95).

It is important to notice that zombies, construed as such, are taken to be in possession of all sorts of beliefs, thoughts, and judgments that their human twins typically have, including the self-ascribed ones. As Chalmers (1993) states: "Zombie Dave's beliefs may not be colored by the usual phenomenological tinges, but it seems reasonable to say that they are nevertheless beliefs. Beliefs, unlike qualia, seem to be characterized primarily by the role that they play in the mind's causal economy." Accordingly, the zombie twin, too, takes aspirin because he *thinks* he has a headache, wants anesthetics at the dentist chair because he *believes* the root canal will hurt, and so on. It is just that all his beliefs and judgments about his own qualia are systematically false. Nothing hurts in him, even if he sincerely believes he has a splitting headache. Accordingly, we should not be motivated to put him under anaesthesia when his tooth is being drilled, despite all his screams, if it is the pain quale that matters.

Put differently, what distinguishes this kind of a zombie from its human twin is the *stipulated* everpresence of a gap between the "appearance" and "reality" of the zombie's qualitative states—what qualia he judges himself to have versus what qualia he really has—and nothing much else. (Note that something like an "appearance-reality" distinction is required in order to coherently conceptualize the possibility of a zombie.) As such, there is a psychologically significant and explanatorily important sense in which things *seem* (i.e., are judged, thought, believed, expected, noticed,... to be) a certain way to the zombie twin.

Consequently, to the extent that nonqualitative intentional states, (including self-ascribed beliefs about one's own qualitative states) are constitutive of a first-person perspective (cf. Chisholm 1981), the zombie can be said to have such a perspective (albeit a systematically misguided one). A zombie's life is not, after all, completely devoid of all mental elements. He only lacks an important component of an otherwise intact epistemic perspective. After all, it is in virtue of having such a perspective that receiving anaesthetics at the dentist chair *seems* to matter to the zombie twin, even if his preferences are entirely on the basis of false self-ascribed beliefs (about his own non-existent pain qualia). It is this characterization of zombiehood that is typically invoked in contemporary debates, and I will confine my discussion accordingly throughout the rest of this chapter.

78. Notice that this view, as such, does not deny that we are conscious. It comes close, however, in positing that our being conscious, in itself, makes no difference—hence the path to zombiehood.

79. James (1879) is a vigorous response to Huxley. See also Capek (1954) for a commentary on this exchange.

80. As far as I could trace, the term *zombie* enters the philosophical vocabulary with Kirk (1974) in an argument against materialism. For other arguments that defend the intuition in favor of the possibility of zombiehood, see Block (1978), Block and Fodor (1980), Searle (1992), and Chalmers (1996). For counterarguments, see Lewis (1972) and Shoemaker (1975), as well as Kirk (1994). For an expression of philosophical intolerance for epiphenomenalism and the possibility of zombies, see Dennett (1991, chp. 10). Güzeldere (1995d) distinguishes among physiological, functional, and merely behavioral zombies, and briefly examines their respective underlying metaphysical assumptions. (See *Journal of Consciousness Studies* (2:4, 1995) for a symposium on zombies based on Moody 1994).

81. Here is Dennett (1979) on describing the phenomenology of one's own experience: "We are all, I take it, unshakably sure that we are each in a special position to report, or to know, or to witness or experience a set of something-or-others we may call, as neutrally as possible, *elements of our own conscious experience*.... Propositional episodes ... comprise our streams of consciousness by embodying our semantic intentions of the moment, by being the standards against which we correct, or would correct, any failures of execution were we to utter anything at the time.... These are ... thinkings that p. ... I call them judgments.... Such judgments *exhaust* our immediate consciousness, that our individual streams of consciousness consist of nothing but such propositional episodes. My view, put bluntly, is that there is no phenomenological manifold in any such relation to our reports. There are the public reports we issue, and then there are the episodes of our propositional awareness, our judgments, and then there is—so far as *introspection* is concerned—darkness" (pp. 93–95).

This quotation probably represents a position at a far end of the spectrum of views on the nature of inner experience. There are, of course, many midway views between the positions that occupy the two endpoints. For instance, one can agree with "friends of qualia" that there *is* something crucial to theorize about under the term "phenomenal consciousness" distinct from "episodes of propositional awareness and judgments," but maintain that the way they choose to theorize about it is misguided.

82. The disagreement is not only between those who believe in phenomenal consciousness and those who deny it; it pervades the community of the supporters of this very notion. For instance, Block claims that Searle, in an attempt to point to consciousness, confounds too many senses of the term. As Block says, "It is important to point properly." But who has *the* omniscient pointer?

83. Jaegwon Kim (1996) makes a similar point with respect to mind in general: "Saving mentality while losing causality doesn't seem to amount to saving anything worth saving. For what good is the mind if it has no causal powers?" (p. 237). Or, to transform a point Fred Dretske (1988) makes with respect to the explanatory role of content into one about phenomenal consciousness (with apologies), "if having a mind is having *this* kind of qualia (which don't *do* anything), one may as well not have a mind" (p. 80).

However, this position should be distinguished from eliminativism about consciousness, defended, for instance, by Rey (1988). Rey's suggestion is to do away with the largely folk-theoretic notion of consciousness because it contains Cartesian elements and plays no useful causal-explanatory role. I am suggesting (and probably Kim and Dretske would agree), in contrast, that because we do *not* want to do away with our common sense notion of consciousness (or, at least, a significant part of it), we need to seek to secure a genuine role for it in the causal web of the world.

84. Put differently, I am essentially in agreement with David Lewis (1980) on a point he makes regarding the status of pain that I take as an objection to the *segregationist intuition*: "Only if you believe on independent grounds that considerations of causal role and physical realization have no bearing on whether a state is pain should you say that they have no bearing on how a state feels" (p. 222). Along these lines, see Humphrey (1992); see also Hardin (1987, 1988) for an important attempt to deflate the explanatory gap between the causal and phenomenal aspects of consciousness in the case of visual perception and color qualia.

85. The Fodor quote is from Fodor 1987, p. 156. Fodor holds a much more pessimistic opinion regarding the prospects for a theory of consciousness in comparison to a theory of rationality, however. Regarding rationality, he thinks that "certain residual technical difficulties" notwithstanding, "we are (maybe) on the verge of solving a great mystery about the mind: *How is rationality mechanically possible?*" (Fodor 1987, pp. 156, 21). Regarding consciousness, here is what he says: "Nobody has the slightest idea how anything material could be conscious. Nobody even knows what it would be like to have the slightest idea about how anything material could be conscious. So much for the philosophy of consciousness" (Fodor 1992, p. 5).

Perhaps then, one thinks, Fodor should consider starting to think about consciousness and give those working in the field a helping hand. But similarly gloomy sentiments are expressed by Block (1994), a prominent figure in the consciousness literature, as well: "The notable fact is that in the case of thought, we actually have more than one substantive research programme, and their proponents are busy fighting it out, comparing which research program handles which phenomena best. But in the case of consciousness, we have nothing—zilch—worthy of being called a research program, nor are there any substantive proposals about how to go about starting one.... Researchers are *stumped*.... No one has yet come up with a theoretical perspective that uses these data to narrow the explanatory gap, even a little bit" (p. 211).

My ultimate conclusion, expressed in the last paragraph of this chapter and hopefully substantiated by the exposition presented thus far, differs from both.

References to the Introduction appear on pages 807–816.