Introduction: What Is "The" Problem of Consciousness?

Both in philosophy and in psychology "the problem of consciousness" is supposed to be very special. It is not just the mind-body problem; few theorists question the eventual truth of materialism in some form, but many see a deep principled difficulty for the materialist in giving a plausible account of "consciousness." Nor is it just the problem of intentionality, or mental aboutness, in particular, since intentional states need not be conscious in any sense at all.\(^1\) It has to do with the internal or subjective character of experience, paradigmatically sensory experience, and how such a thing can be accommodated in, or even tolerated by, a materialist theory of the mind.

And it is a conceptual problem, not merely an empirical one—it is a "How could ... possibly ..." question, not merely a "How does..." question. Scientifically, there is felt to be a systematic, if not insuperable, obstacle to psychological research, and philosophically there is felt to be a conceptual tension between materialism and the phenomenal or subjective character of experience.

Now, if we are to understand "the" problem, we must have a clear sense of what is meant by "consciousness," "subjective," "phenomenal," and the rest. And until recently, such a sense has been sadly lacking. But even in the past few
years, I will argue, it has been in a way even more garishly lacking. Namely, although any number of useful clarifications and distinctions have emerged in the literature, *they have only continued to be trampled*. From time to time philosophers have acknowledged and even articulated a multiplicity of meanings, especially of the term "consciousness," but the fact of the multiplicity has never properly been taken to heart. Both psychologists and some philosophers still use the word univocally and without explication, as if it had one clear meaning and we all knew what it meant; many sharply distinct phenomena are still being lumped together under that heading and its commonly but falsely presumed synonyms, such as "subjectivity," "phenomenal character," and "qualia." The main purpose of this book is to argue that once we enforce the distinctions, we will be able to divide and conquer in a most promising manner.

1 CANDIDATES

Here are some importantly different uses of the word "conscious."

(i) Organism consciousness. A thing is a conscious being as opposed to a nonconscious being if and only if (iff) it has the capacity for thought, sensation, feeling, etc. (whether or not that capacity is ever exercised).

(2) Control consciousness. A creature is conscious as opposed to unconscious iff it is awake, has occurrent mental states, and is in control of its actions in a way that is consonant with those mental states.

N.b., a person who is "unconscious" in this "control" sense can still be conscious of things, as in vivid dreams, and can also have qualia in any going sense of this term (which will be multiply disambiguated in chapter 4). Control consciousness is not too different from Ned Block's (1993, 1995) "access" consciousness, from David Rosenthal's (1991b) "creature" consciousness, or from Roger Shepard's (1991) "objective" consciousness. This is a rough category, for there are lots of odd fringe cases here, e.g., that of somnambulism.

(3) Consciousness of. A creature is conscious (or aware) of this or that. The object of such consciousness may be external, abstract, physical, internal, somatic, mental, or whatever.

Note that this is a stronger notion than that of consciousness/awareness that something is the case, for one can be aware that *x* is *F* without being conscious/aware of *x* at the time (Dretske 1993).

(4) State/event consciousness. A state of a subject, or an event occurring within the subject, is a conscious state or event, as opposed to an unconscious or subconscious state or event, iff the subject is aware of being in the state or hosting the event.

There are at least three subcases of un- or subconsciousness in sense (4): simple distraction or other inattention; inaccessibility of processing, as in language understanding or pattern recognition; repression or some other Freudian mechanism.

(5) Reportability. In one useful sense, one is conscious of all and only those items on which one can readily issue a verbal report.

Putnam (1960), Rorty (1970), and Dennett (1978a) all mobilize versions of this notion. Though reportability can be accepted as a sometimes handy stipulation, I think it is not an ordinary sense of the term "consciousness," because it is both too easy and too hard to satisfy. For example, Putnam himself described a very dull automaton that, whenever it is in a particular state *A*, is simply caused then to print out the
expression "I am in state A," but this is not an example of consciousness in any standard sense. Conversely, a creature might be conscious in almost any sense without having speech organs or even a module capable of driving such organs.\(^3\)

(6) **Introspective consciousness.** This is "perception" by Lockean "inner sense," i.e., by focusing one's attention on the internal character of one's experience itself.

There are light and casual degrees of this, more deliberate degrees, and even very determined degrees amounting to serious phenomenological investigation. (Note how entirely independent sense (6) is of (i) and (2); beings could be conscious in both of those senses but lack introspection entirely.) I should mention—what will be explained at more length in the next chapter—that the metaphor of "inner perception" is only that, and is to be cashed in terms of empirically discoverable attention mechanisms that output representations of some of one's own psychological states.\(^4\)

Very likely, introspective consciousness is a special case of (i), consciousness of, and arguably it explains (4), state/event consciousness (see chapter 2).

(7) **Subjective consciousness.** This is (metaphorically speaking) having a "point of view." A subject's consciousness in this sense is "what it is like" for the subject to be in whatever mental states it is in. A tighter characterization might be: what can be described, if at all, only in the first person.

I will argue in chapter 3 that this sort of consciousness is completely explained by (6), introspective consciousness.

(8) **Self-consciousness.** This is, at least, having a sense of oneself as an individual separate from other individuals and the rest of one's surroundings. In a stronger sense, it is consciousness of one's self.

---

**CHAPTER ONE**

(Whatever this might be, and if there is any, I would not touch this sense for a free week on Maui with champagne thrown in.)\(^5\)

There are more uses of "conscious." (And, of course, there are even more specialized and topic-tied uses: "altered states of consciousness," "false consciousness," social abstract-noun phrases such as "feminist consciousness" or "the religious consciousness of the age" and the like.)

Which uses are problematic in the seemingly special way? Uses (1) (organism consciousness) and (2) (control consciousness) involve no more than the mind-body problem itself. Uses (3) (consciousness of) and (4) (state/event consciousness) seem to be just special cases of intentionality. Use (5) (reportability) is not especially interesting, as its proponents have been concerned to point out. Use (6) (introspective consciousness) is no more troublesome than are ordinary modes of perception or attention mechanisms, however empirically difficult these might be (remember that "the" problem is not primarily an empirical one).

So we will be dealing primarily with consciousness in sense (7). (As I said, I refuse to touch sense (8).) But to have narrowed the topic down this far is, comparatively, to have narrowed it hardly at all, for further distinctions proliferate. Here are some possible candidates for "the" problem of consciousness (sense (7)).

(A) **The subject/object distinction**, i.e., being in a mental state oneself versus observing someone else's brain while he or she is in that state. The felt mutual incongruity of these two conditions can psychologically lead to doubt about materialism.

I argued in *Consciousness* that one must guard against the "stereoptic [or stereoscopic]" fallacy of supposing that, because to have a vivid perceptual experience oneself is nothing like observing the brain of someone else who is having it,

**INTRODUCTION: WHAT IS "THE" PROBLEM OF CONSCIOUSNESS?**
having such an experience must be entirely different in na­
ture from any goings-on in the brain that underlie one's hav­
ing it. (Even to call this inference a "fallacy" is to flatter it.)

(B) Immediate or at least privileged access. Almost no one
would deny that I can both experience and know the con­
tents of my own mind in a way that you can neither experi­
ence nor know the contents of my mind.

But this is not a problem; it is simply a fact, which-must be
explained, or at least accommodated, by any adequate the­
ory of the mind.

(C) Temporal and other empirical anomalies. There are
fascinating empirical but philosophically shocking puzzles
about the relations between observationally detectable
events and subjects' awareness of these events. (These are
discussed at length by Dennett and Kinsbourne [1991a] and
Dennett [1991]: see chapter 2. below.) Work on one version
or another of the "binding problem" in brain science is also
sometimes labeled as research on "consciousness" per se, as
for example by Crick and Koch (1990).

(D) How/why did consciousness evolve, and what is it
[1991], and especially Flanagan [1992] have all offered con­
jectures on this.)

(E) Epistemology, e.g., of "inverted spectrum." How
could we possibly tell whether someone's internal color
spectrum is inverted, or whether a food sample tastes
the same to one person who likes it and to another who
does not?

Candidates (C) through (E) are interesting and perhaps in­
tractable problems for psychophysical research, but unless
one is a verificationist, one will not see great philosophical
significance in them.

(F) "Inverted qualia" and "absent qualia," or zombie, ar­
guments. To some philosophers it has seemed obvious that
the conditions of this or that materialist theory of conscious
experience could be fulfilled and yet the organism in ques­
tion have either inverted phenomenal contents or no con­
scious or phenomenal experience at all.

Such arguments are most often aimed against functionalist
versions of materialism. (The issues here are metaphysical as
opposed to epistemological, as in (E).) In chapters 3 and 5
of Consciousness I accepted a few of these arguments, aimed
against simpleminded materialist theories that I do not hold,
but put the rest down to their proponents' hyperactive imagi­
nation. I will revisit some of these issues in chapter 6 below.

2 MORE SERIOUS BUT STILL DIVERSE CANDIDATES

Candidates (A) and (B) are nonproblems. Candidates (C),
(D), and (E) are at least not "the" big conceptual problem.
Candidate (F) comprehends a family of parallel arguments
by counterexample against various materialist theories of
the mind (as I said, I believe that some of the counterexam­
ples work and some do not), but the arguments are purely
negative, and even where they do work, they do little by
themselves to bring out what "the" big problem is supposed
to be.

Now focus introspectively on the phenomenal character
of your inner experience, say on the rich cyan color presented
to you by the background on your new state-of-the-art color
monitor, and some bigger conceptual difficulties will bare
their teeth.

(G) Homogeneity or grainlessness. Some theorists, most
notably Sellars (1962, 1965, 1971) have seen a severe in­
congruity between the smooth continuous character of a
phenomenal quality such as color—phenomenal pink is simply "pink through and through," in Sellars's phrase—and the discrete, particulate nature of the material of which we are made.

This can be written off by materialists as a threshold effect, simply a matter of the subject's failure to perceive the gaps that make for grain. As Armstrong (1968a) would say, we mistake our failure to perceive gaps for a successful perceiving that there are no gaps. But the matter of homogeneity is extraordinarily complicated, and, somewhat arbitrarily, I choose not to engage it in this book.6

(H) The monadic, first-order qualia of apparent phenomenal objects. I contend that to a striking extent Russell was right: one's visual field, for example, does take the form of a mosaic, an array of irregularly shaped color patches fitting together, and we can talk extensively about each patch and its qualitative and relational properties.7

In particular, not only the cyan color presented by your monitor but also the yellow of Russell's famous second candle flame seem to be the colors of phenomenal things. If we are not simply to give up materialism in favor of a sense-datum metaphysics, this sort of qualia talk must be explicated in a naturalistically acceptable way, and that is no easy task. I argued in Consciousness that the objects are intentional objects, often unreal intentional inexistents; I will defend this theory further in chapter 4 below. And as we will see, a new and more rarefied use of the term "quale" has emerged in recent years, posing a new problem for the theory (see chapter 6).

To anticipate an important theme of this book, note that there is no obviously important relation between qualia in the strict sense of (H) and consciousness in any of our original senses of this term, save the uninteresting sense (1) (organism consciousness). In particular, one cannot simply assume that if a subject's state involves a quale, then the subject must be conscious or aware of being in that state or being acquainted with that quale.8 One might go on to argue a connection-between qualia in the present sense and some version of subjectivity or "what it's like," but here too one would have to argue it. I will say more of this in chapter 4.

(I) The intrinsic perspectival, point-of-view and/or first-person aspect of experience, as discussed by Nagel (1974) This aspect is not just (A), the mere fact that, e.g., seeing cyan oneself is not like watching the gray cheesy brain of someone else who is seeing cyan, nor is it just the inevitable but entirely objective filtering of external information.

(J) Funny facts, as revealed by the "knowledge arguments" given by Nagel (1974) and Jackson (1982). These facts cannot be scientific facts, for they can be known only from particular points of view, some of which may be inaccessible to human beings.

Candidates (I) and (J) are closely related, though not precisely the same problem. I addressed both in chapter 7 of Consciousness; I will do considerably better in chapter 3 below.

(K) Ineffability. One often cannot express in words what it is like to have a particular sensation, except in comparative terms that can be understood only by someone who has had a closely related experience.

(L) The "explanatory gap," called to our attention by Levine (1983): even if God were to assure us that, say, the type-identity theory of mind is true and that such-and-such a conscious experience is strictly identical with a firing of certain neural fibers, we would still lack an explanation of why those fiber firings feel to their subjects in the distinctive way in which they do.

For example, granted that M-fiber firings do simply constitute their subject's being appeared to mauvely, why does this
experience have the qualitative character of phenomenal mauve rather than some other such as phenomenal yellow?

Candidates (I) through (L) do call for some positive account, and I gave none in *Consciousness*. I will argue here that a good one is furnished by Locke's "inner sense" picture of introspection, coupled with the analogy with indexical utterances in natural languages that I will develop in chapter 3.

As I said in my preface, my strategy will be to divide and conquer. Once we keep (G) through (L) ruthlessly separate and tackle each set of issues on its own terms, we will see that each is tractable. I do not say that in each case my solution will be immune to objection; I do claim that my solutions are plausible and that they cannot easily be refuted.

I should emphasize that my goals are modestly limited to those sketched in the preceding paragraph: in sum, to show how it is possible and even plausible that a purely physical organism might exemplify unusual features of the sort that figure in (G) through (L), not just individually but all at once. I will not claim, here or ever, to "explain consciousness" (in the title phrase of Dennett's well-known work). For on my own view, that would be to explain each of any number of different things, a set of Herculean empirical and philosophical tasks. Also, as Joseph Levine has observed to me, it would also court a fallacy of composition, since the set of explanations of all the various phenomena might not yield a master explanation of their organic sum.

Some defenders of materialism like to start from a heavily fortified position, by conceding almost nothing to their consciousness- or subjectivity- or qualia-driven critics; they simply deny the early premises of their critics' arguments. I incline, temperamentally, in just the opposite direction. I would rather give the critics lots of rope, grant their early premises, and then, after making many concessions, show that the arguments fail anyway and that the hostile premises can triumphantly be welcomed within a master theory of consciousness, subjectivity, qualia, and the rest. In particular, I will concede, indeed insist, that the phenomenal characters of (some) mental states are

- real,
- internal,
- specially accessible to/by me,
- ineffable,
- intrinsically perspectival,
- in one sense, inaccessible to science,
- in one sense, inexplicable.

As we will see, each of these features is nicely predicted by the Lockean picture I have mentioned, with no implication that consciousness is a surd in nature or in any other way an obstacle to functionalist theories of mind.

I will also have another go at the subproject of establishing what I call the *hegemony of representation*. As in *Consciousness* (though I was insufficiently explicit about it at the time), I am concerned to maintain a weak version of Brentano's doctrine that the mental and the intentional are one and the same. Weak, because I am not sure that intentional suffices for representation, but my claim is strong enough: the mind has no special properties that are not exhausted by its representational properties, along with or in combination with the functional organization of its components. It would follow that once representation itself is (eventually) understood, then not only consciousness in our present sense but subjectivity, qualia, "what it's like," and every other aspect of the mental will be explicable in terms of representation together with the underlying functionally organized neurophysiology, without our positing any other ingredient not already well understood from the naturalistic point of view.

I do not think there will be any "problem of consciousness" left.