Eric Lormand, "Steps Toward a Science of Consciousness?" (1998)

"Beats the heck out of me! I have some prejudices, but no idea of how to begin to look for a defensible answer. And neither does anyone else." That's the discussion of conscious experience offered by one of our most brilliant and readable psychologists, in his new 650-page book, modestly titled How the Mind Works. There is no widely accepted scientific program for researching consciousness. Speculation on the subject has been considered safe, careerwise, mainly for moonlighting physicists or physiologists whose Nobel Prizes and similar credentials are long since safely stored away. This essay describes some recent efforts of philosophers of mind who have stepped into the breach. Some argue that the puzzle of consciousness is impossible to solve, and some argue that with certain confusions removed there's no distinctive puzzle at all. I write from the standpoint of a third group who think the puzzle is difficult but tractable, and who get involved under the pretext that "philosophy is what you do to a problem until it's clear enough to do science to".

Some preliminary distinctions

In that spirit, let's consider a few different uses of the word "conscious". Sometimes we use it to describe things with minds: human beings or perhaps animals or other subjects. For example, we speak of creatures as being (or not being) conscious of objects in their environments, or conscious that things are a certain way, or self-conscious, or, conscious period (e.g., awake or maybe dreaming). None of these various uses are the focus of this essay. The central issues surround the use of "conscious" to describe things within minds: thoughts, feelings, and other mental "states" (events, objects, processes, etc.). We never speak of mental states as themselves being self-conscious or conscious about anything, but only as being (or not being) conscious, period. By most accounts there are two important considerations according to which we call mental states conscious: whether a state is "introspected" and whether it has "phenomenal" character. These bits of jargon take some getting used to, but I will try to give some shape to them briefly.

Introspection is supposed to be a way each of us has to find out about our own mental states without inferring their existence from observations of our bodies or surroundings (e.g., without watching our overt actions, and without consulting a psychoanalyst). With barely any hesitation you can report on many of your mental states—are you sleepy, or feeling a tickle, or aware of what day it is? Other cases take a bit more "soul-searching", but you can pull them off with eyes closed—do

you have a memory of the 7's times table, or a resentment toward your neighbor? There is widespread agreement that introspection does not make us omniscient about our own minds. Some states of your mind may be quite inaccessible to introspection—exactly what makes you forget or dream or comprehend the things you do? Most famously, psychoanalysts appeal to deeply nonintrospectible attitudes and emotions to explain otherwise bizarre dreams, associations among concepts, apparent slips of the tongue, emotional disorders, neurotic physiological reactions, and so on. Even without commitment to psychoanalytic methods, psychological research reaches parallel conclusions about nonintrospectible mentation, as in cases of subliminal perception. A spectacular case is "blindsight", which appears in subjects with damage to certain neural pathways connecting portions of the retina to the parts of the brain controlling vision. They sincerely deny that they have visual experiences in the affected parts of their visual field. Yet in some sense they have perceptual states sensitive to stimuli presented there. When begged enough times to guess whether an "X" or an "O" is presented, they reluctantly oblige, and surprisingly guess correctly 80-90% of the time. When asked to reach for objects in blindsight regions, also, some subjects reflexively pre-orient their hand and fingers in ways suited to the specific shapes of the objects. Some can even catch projectiles they swear they cannot see! Something mental must be going on, without associated introspectibility.

Now turn to phenomenal character, the other main factor we appeal to in calling a mental state "conscious". Consider these four kinds of introspectible states: perceptual experiences, such as tastings and seeings; bodily-sensational experiences, such as those of pains, tickles and itches; imaginative experiences, such as those of one's own actions or perceptions; and streams of thought, as in the experience of thinking "in words" or "in images". All these states have features that make up "what it is like" (or "seems like" or "feels like") for one to undergo them. We sometimes try to describe these features, for example, by saying that a given pain is "sharp" or "throbbing" to some degree, or that a given visual image is "blurry" or "moving". These specific features are called "phenomenal properties", or sometimes "qualitative properties", "sensational qualities", "raw feels", or "qualia", more or less interchangeably. Let's say that a mental state is "phenomenal" just in case there is something or other it's like for one to have it. We can also call them "experiences".

To clarify the apparent difference between being phenomenal and bring introspected, consider whether a mental state can have one but not the other feature. Can there be nothing it's like to have a state, even when one is introspectively aware of it? Take as a test case the philosopher's favorite example

of a mental state: the belief that snow is white. Usually one's belief that snow is white lies dormant and unintrospected, though one can raise it to introspection easily. When one introspects the belief, is there something having the belief is like? A "yes" answer is tempting, but on a careful look it isn't clear that the phenomenal character attaches to the belief itself. When we try to describe what having the belief "is like", we seem to rely on what it's like to have experiences accompanying the belief, such as auditory imaginings of asserting the words "snow is white" (or "I believe snow is white", or "Mon Dieu! La neige! Blanche!"), or visual imaginings of some fictitious white expanse of snow, together with feelings or imaginings of moving eyeballs, eyelids, brow, breath, jaw-muscles, etc. as one thinks. Pending evidence of further aspects of what it's like to have the belief, this illustrates how there can be something it's like when one has an introspected state, although the state itself has no phenomenal character.

This suggestion about the philosopher's favorite belief generalizes to other beliefs and to other "attitudes" such as desiring, wondering, and hating. Sometimes one's hatreds are resistant to introspection, and persist for weeks or years, even when one is distracted from them or fast asleep. There is something it's like when one introspects them, but this seems best described in terms of the phenomenal characters of associated states, not of the hatred itself: images of hateful speech or of misfortune to the hated, feelings of clinched fists, etc. These can come and go while the hatred itself remains constant. Similarly for mental character traits such as forgetfulness or for moods such as elation; when we introspect them the phenomenal character seems attached to various perceptions, bodily sensations, imaginings, and thoughts that are merely symptomatic of the trait or mood. So we can distinguish between mental states that do have phenomenality (at least when introspected)—e.g., perceptions, bodily sensations, imaginings, and verbal or imagistic thoughts—and the kind that do not have phenomenality (even when introspected)—e.g., beliefs and other attitudes, along with traits and moods. Nonphenomenal mental states may help determine how other things seem to us, but only phenomenal mental states themselves seem some way to us.

Doubts about scientific explanation

Phenomenal character has seemed more troublesome to philosophers than introspection. It inspires several charming arguments for the scientific inexplicability of consciousness, and even for mind/body dualism. Here are three. The argument from "phenomenal objects" involves active reader participation. Step One: form in your mind a purple, round afterimage (using a light bulb and a white wall). I will wait … Step Two: rigorously examine every physical entity in your

brain, body, and (causally relevant) environment, searching for the afterimage. Let me know when you are done … Step Three: discover that nothing physical in or around you has the right features to be the afterimage, since nothing in your brain or surroundings is purple and round like the afterimage. Finally, Step Four: give up the search, concluding that nothing physical is the afterimage, i.e., that the afterimage is not physical. There is nothing special about afterimages here: you can reach similar conclusions by repeating the steps for other phenomenal states, e.g., forming a curved and yellow mental image of a banana, or forming a throbbing pain in your big toe, or thinking in soft, medium-pitched sentences. Since the afterimage, the banana image, the pain, and the words are not to be found among the physical things in your brain, body, and neighborhood, they must be among the nonphysical things in your soul, presumably beyond the reach of scientific understanding.

For a second route from the phenomenal to the nonphysical, imagine this:

A super-scientist, Mary, has never seen anything colored, because she lives her life in a black-and-white room, and is even herself painted black-and-white. From a black-and-white terminal in this room, she learns all the objectively specifiable physical (and causal or "functional") facts in the world, a huge list. In particular she learns about wavelengths of light and their detailed impact on eyes and brains and behavior. When she finally leaves the room and first sees color, she is delighted, and exclaims "Oh! It’s like this to see red!"

Mary seems to learn a new fact about the nature of phenomenal experience, one that cannot be identical to a physical or functional fact, or else it would be among the facts she already knows before leaving the room. If so, certain facts about phenomenal states are not in the physical realm, and so are also presumably outside of science’s domain. This is called the "knowledge argument".

The final argument suggests that there is a permanent "explanatory gap" preventing scientific accounts of consciousness from being as satisfying as other scientific explanations, such as the chemical account of water. How does chemistry justify the conclusion that H2O is water rather than, say, oil? In part by showing that there is a preponderance of H2O in our lakes and rivers, and that H2O boils, erodes rocks, quenches thirst, etc. But what makes these features relevant to determining what’s water? It seems we must start out with a prechemical concept of water specifying that anything with these features is water. If science establishes that H2O has the features, then using our water concept we can literally prove that H2O is water. The situation seems very different for consciousness:

For any objective, scientific account of phenomenality, one can conceive of a creature that meets all the conditions in the account but lacks phenomenal states. In the extreme case, we can conceive of a world that is an exact physical duplicate of the actual world—complete with duplicate stars, planets, rocks, plants, animals, and philosophers—but which lacks any beings enjoying what-it's-like-ness. All the human-like beings in that world would be nonphenomenal "zombies", who do not have experiences it's like anything to have. They would be able to walk and talk and react to the environment in complex ways (perhaps as in some fancier version of blindsight), pursuing various objectives according to their best laid plans, but all without any "light on" inside.

If this is conceivable, then the prescientific concept of phenomenality, of what-it's-like-ness, will never allow us to prove from any scientific premises the presence of phenomenal states. There is no mystery about why H2O is sufficient for water, given everything else we know about H2O and given our prechemical concept of water. By contrast, for whatever ingredients science specifies in a recipe for consciousness, there will always be a mystery remaining about why they would be sufficient for consciousness, no matter what else we know about them.

The threat to self-knowledge

The idea that phenomenality depends on nonphysical features raises a strange kind of skepticism about one's own phenomenal states. The dualist who believes that zombies are possible believes that two people could be alike in all nonphenomenal respects, while one has phenomenal experience and the other does not. If as I argued above beliefs (like other attitudes, and moods) are nonphenomenal states, then this means that two creatures could have all the same beliefs, and even all the same introspectively generated beliefs, while differing in zombiehood. Each could be fully convinced of his own rich, detailed phenomenal experience, but one would be utterly wrong. It is difficult to see how either could justify his belief that he is not a zombie, if zombies are possible. This would lead to the bizarre possibility, for all we know, that we, ourselves, might not after all have states it's like something to be in. There would be no introspective way for you to determine whether you are a zombie without phenomenal states (but perhaps with normal nonphenomenal states, such as the belief that you're having phenomenal states).

Dualists are not alone in facing this worry. A similar problem arises even if we try to explain phenomenality directly in biological (e.g., neurophysiological) terms. Imagine that, perhaps after discovering correlations between phenomenal experience and certain neural occurrences in the brain, we try to explain the

difference between conscious and unconscious mental states by appeal to differences in these neural features. The problem is that there is no purely introspective way to determine whether one has the relevant neural features—rather than having a differently wired brain, or maybe an electronic "brain" manufactured at M.I.T. As long as we're getting science-fictional, for all one knows introspectively, one might even have a control system that can switch between the right kind of brain and the wrong kind. One might even oscillate constantly between having and lacking "something it's like", without being able to discern the difference!

We might simply refuse to get worked up about threats to knowledge that are based on such outlandish possibilities as complete physical duplicates or hybrid bionic brains. But keep in mind that the subject matter of the threatened knowledge is very special. We are not merely imagining threats to knowing the external world or to knowing one's deeply unconscious mind. We are trying to imagine your being utterly clueless about whether you have any conscious experiences at all, e.g., about whether you're feeling severe pain, or whether there's anything at all it seems like for you, fully awake and alert, as you read these words. It would be mysterious if there were describable situations in which introspection left completely open the questions it should answer most easily and surely. Probably each of us approaches the subject ready to insist: whatever claims are negotiable about others, and whatever other claims are negotiable about me, the one thing I cannot be wrong about is my having a conscious mental life.

Making good on such brash claims about introspection requires extreme measures. In keeping track of the outside world, at best, certain mechanisms keep one's beliefs in rough accord with the facts (e.g., mechanisms of perception, reason and memory, and the persistence of facts when one is not continually checking them). Such mechanisms fail; a mechanism of this complexity that could never possibly fail would be a miracle. That is why infallibility about the outside world would be mysterious. But the same reason to expect fallibility holds for introspection: if there is the slightest mechanism correlating one's experiences with one's introspections of them, it should be breakable, and if there is no mechanism, a perfect correlation between the two would seem to be sheer luck.

Perhaps the only nonmiraculous way for introspection to be guaranteed correct about phenomenality is for introspection to constitute phenomenality. If there were further nonintrospectible requirements on having phenomenal experience—whether these requirements were spiritual, biological, neurophysiological, or hidden psychological ones—this would jeopardize one's introspective knowledge

of whether one has experience. So could introspections of phenomenality be self-fulfilling? One difficulty is that we have already seen cases in which introspection is not sufficient for phenomenality: beliefs and other attitudes, moods, and character traits can be introspected without being phenomenal. This suggests that there are (at least) two kinds of introspective processes, one which is available to nonphenomenal beliefs, moods, etc. and does not constitute phenomenality, and one which is available to phenomenal perceptions, imaginings, etc. and does constitute their phenomenality. Let's see what we find among the various available accounts of introspection, and then bring those resources to bear on the puzzles about phenomenal experience.

Theories of introspection

How should introspection be explained? The etymology suggests likening it to an inwardly directed form of perception. But since we have no literal inner eyes or ears, pointed at our brains or souls, it is unclear what the analogy between perception and introspection could be. Still, theorists of introspection have considered the analogy a useful foil, developing alternatives in the face of various objections to the idea of inner perception.

We often try to distinguish between perceiving things (the apples in front of one's face) and merely inferring beliefs about them (the apples over at the store). So if introspection is an inferential process, this might weigh against the perceptual metaphor. And in fact scientific investigations of 'confabulation' reveal the widespread presence of hidden inferences in introspective access. In identifying one's beliefs and motivations, one systematically but sincerely reports attitudes one thinks rational or statistically normal in the circumstances, even if one doesn't have them. For instance, in the 'bystander effect', increasing the number of joint witnesses to a needy person decreases the likelihood that any of them will assist. Bystanders rarely report this as a factor in their decision whether to help, however, often claiming instead to have reached a decision based solely on their own likelihood of success. Presumably we make the same kind of mistakes about other mental states, such as our moods and character traits, erring systematically in the direction of the states we judge appropriate or normal in our circumstances. So, much apparently noninferential access to attitudes and moods consists of self-directed, fallible guesses, based at best on commonsense abilities to rationalize behaviour.

Upon examination this process does not seem to be perceptual in any interesting sense. Initially it may seem noninferential, but that is simply because one fails to

introspect the intermediate inferential steps. (Contrast the nonintrospective case of finding out about yourself by inference from the testimony of a psychologist, a case in which you do notice intermediate steps such as hearing the testimony, thinking about the trustworthiness of the psychologist, etc.) A natural idea about introspection, then, is that it need not involve any special means of access, but is simply what we say we're doing when we reach beliefs about our mental states but have no idea how. All there is to introspection is the production, in any variety of hidden ways, of beliefs about one's mental states (so-called "second-order" or "higher-order" beliefs.)

However, there is room to hold out for a kind of introspection that does involve a special, noninferential means of access. The confabulation model of rationalizing or statistical guesswork does not extend easily to introspection of phenomenal experiences. For example, untutored subjects offer consistent and apparently reliable reports of stinging (rather than throbbing) pain feelings when a limb has restricted blood flow. They seem not to infer or confabulate these feelings, since no commonsense principles of rationality dictate that one should feel stinging rather than throbbing, and since the subject need know no relevant statistical information about how people feel in these circumstances. So there may be a more restricted domain of the mind in which introspection is more like inner perception and less like hidden theoretical inference.

Another objection to the inner-perception metaphor is based on the existence of perceptual illusions. Since one often suffers ordinary perceptual illusions, the more analogous introspection is to perception, the more likely it would be that one would suffer naïve introspective illusions about what one's conscious experiences are like. But it rarely if ever happens that one mistakes, say, a dull pain for a sharp pain, in the way that one mistakes a roadside cow for a horse. This may be evidence that introspection is sometimes neither inner perception nor self-directed theoretical inference, but a process with fewer breakable causal links. Some introspective access may be like one's psychologically primitive abilities to shift among mental states. Just as the transition from believing that p and q to believing that p presumably takes place without intermediate inference or inner perception, so might the transition from (say) believing that p to believing that I believe that p, or the transition from having a dull pain to believing that I have a dull pain. As one author proposes, simply, "our brains are so wired, that … one's being in a certain mental state produces in one …the belief that one is in that mental state. This is what our introspective access to our own mental states consists in." A challenge for this view is to explain lawful and systematic patterns among introspectible and nonintrospectible states, without an ad hoc assumption, for each pattern, that it

happens to be "wired" to higher-order beliefs in the right way. For example, why aren't brain states governing autonomic bodily functions introspectible? Why are perceptual experiences introspectible but not subliminal perceptions and early perceptual states? If all one needs is a wire, why does it seem easier to introspect one's fleeting thoughts than one's deeply held beliefs?

Consider another issue confronting both the higher-order belief view and the wired-belief view. If introspection should be understood in terms of these end-products, could it be seen as constitutive of or even necessary for phenomenality? A "yes" answer would help avoid the bizarre self-skepticism, but it would not sit well with the view that many species of animals can have experiences—that there is something it's like for cats and dogs to hurt or to see bright lights, for instance. It is implausible that these beings have introspective beliefs that they hurt and see. This would require having concepts of hurting and of seeing, and perhaps a self-concept, and all this would seem to involve capacities beyond the reach of most nonhuman animals—for example, the ability to conceive of others as hurting and seeing, and the ability to remember or envision oneself hurting and seeing. Defenders of an introspective belief requirement on phenomenality must either deny that animals have conscious experiences, or else somehow attempt to minimize the conceptual sophistication needed for introspective beliefs. This tension is more commonly taken to be a serious strike against requiring introspective beliefs for phenomenality, especially given that a similar tension arises in the case of human infants. Even if we conclude that newborns (say) don't yet have general concepts of pain, and so can't genuinely believe that they are in pain, this is a far cry from concluding that they can't be in pain.

Even for beings with the requisite conceptual capacities, it seems implausible that introspective beliefs must accompany each of their experiences. At any given moment one can attend only to a small proportion of the sensory stimuli one encounters. It is also difficult to attend simultaneously to the outside world and to one's experience of it. Nevertheless, plausibly, there is something many inattentive perceptions of unattended stimuli are like; experience would be quite impoverished were it not for the contributions of background noises and odors, pressures on one's feet or seat, moisture under one's tongue, peripheral vision, and so on. It is possible to maintain that one continually forms beliefs about these experiences, but this fits poorly with the difficulty of remembering these experiences (after they change, for example).

In short, if we wish to find a kind of introspection that can help to explain phenomenality, the higher-order-belief and wired-belief views are unlikely to fit

the bill. Though elusive, an inner-perception model of introspection might more plausibly yield a requirement for phenomenality. Just as one can sense a daffodil without having a concept of daffodils, or a tendency to remember the daffodil, so perhaps one can inwardly sense an experience without having a concept of experiences, or a tendency to remember the experience. Animals and babies might sense even if they cannot form beliefs; likewise, perhaps they can inwardly sense even if they cannot form higher-order beliefs. And just as there can be passive, inattentive perception, so perhaps inner-perceptual introspection need not be done intentionally or with attention. A creature's most pressing cognitive needs require mental resources to be directed at the external world, but if inner perception is normally inattentive, it might not draw resources away from attentive outer perception. That would make it more plausible that one has continual inner-perceptual access to one's experiences.

Against the idea that there is inner perception of one's perceptual experiences, it is traditional to object that this would require ghostly "sense data" interposed between physical objects and one's perceptions of them. Accepting inner perception may seem to involve accepting that one at best perceives outer objects indirectly through perceptions of phenomenal mental entities. Such a mediation theory would have difficulty explaining why introspective access to the alleged sense data would not in turn require perceiving further entities ("sense-data data") and so on, infinitely. On one counterproposal, introspection of some mental states consists of their "reflexively" representing themselves (in addition to representing other things). A perceptual experience represents itself rather than being represented by a separate introspective state.

Nevertheless, there is cause for concern about a reflexivity story given the larger aims of scientific explanation. Reflexive representation coheres poorly with more general theories of representation in philosophy of mind, which are "naturalistic" in the sense that they can be pressed into service in providing scientific accounts of the mind. For example, on "causal" theories, mental states represent certain of their ideal or standard causes or effects, and on "correlation" theories, mental states represent certain conditions that they ideally or standardly or historically correlate with. A pattern of neural firing in your brain might truly or falsely "mean" that your flowers are blooming, because that pattern tends to be caused by or to correlate with your flowers' blooming. But reflexive representation for experiences doesn't fit with causal theories, since no mental state, not even a conscious experience, causes or is caused by itself. It also doesn't fit with correlation theories, since every mental state, even a nonconscious nonexperience, correlates perfectly with itself. How can it be that less than all mental states—and more than

none—are reflexive in whatever way is allegedly relevant to phenomenal consciousness?

At any rate, the sense-data objection against inner perception seems misplaced. Inner perceptions needn't be interposed between objects and one's perceptions of them—the causal chain in perceiving a table needn't proceed from the table to an inner perception and then to a perception of the table. Rather, on a more natural view, the causal chain goes directly from the table to a perception of the table, and then (in cases in which the table-perception is introspected) to an inner perception of the perception of the table.

All of the views of introspection described thus far assume that a subject's introspective "access" to a mental state is always a matter of the state's somehow being mentally represented—by a higher-order belief or by an inner perception or by the state itself. The most vehement critics of such views deride them for positing a "Cartesian Theater" in the brain which separates full-blown conscious experiences (those mental states parading onstage) from nonconscious states (those operating backstage). For instance, they point out that no such unified "finish line" for consciousness has ever been found in the brain, and they question what it's usefulness would be. Given that perceptual mechanisms work hard to discriminate things and events in the outside world, and given that the resulting perceptual states can directly guide behavior and mental processing (as in subliminal perception or blindsight), what would be the point of "showing" the resulting perceptual states in a theater of inner perception?

Verbal reportability is the process most frequently appealed to as an alternative requirement for introspection. The idea is that a state may create an input to a system in charge of language use, and so be reportable verbally, without first causing a belief or a perception about itself. This is in keeping with ordinary and scientific reliance on reportability as a symptom of consciousness. One threat is that reportability may only seem relevant to consciousness and introspection because it correlates somewhat with inner-directed representation. Normally in reporting verbally one perceives one's reporting—hears one's speech, feels one's facial motions, etc.—and is thereby in a position to understand one's reports—to recognize one's own voice and realize which mental states one's words express. By contrast, if there is speech without any kind of self-perception, perhaps as in some forms of hypnosis or sleeptalking, this may not seem sufficient for, or even relevant to, introspection or consciousness.

If inner-perception is a kind of introspection that separates phenomenal from nonphenomenal states, this does not need to be done in a single, unified "Cartesian Theater"; there might be several little inner perceivers (or hundreds or thousands of tiny ones) instead of one big one. But why would we have any number of inner-perceptual mechanism—for what function? Perhaps by allowing one to detect certain qualities of mental representations, inner sense allows one to detect their quality: whether and how the representing is degraded (as in doubled or blurred perception), whether it is imaginative (vs. perceptual), whether it is obscured, and so on. This may facilitate behavioral or inferential "corrections," including behavior aimed at improving the quality of perception (shifting position, squinting, etc.). Or perhaps inner perception is a remnant of a primitive stage of evolution of perception. Critics wonder why, given that a creature discriminates conditions out in the environment [in outer perception], it would also need to discriminate states of its brain [in inner perception]. But from an evolutionary standpoint this question may get things backwards. Presumably there were stages in the evolution of sensory organs in which "nearby" states of the brain and body (including the nascent sense organs) were easier to discriminate reliably than "far away" environmental states (e.g., the exact location of predators or mates). Think of flies or oysters or maybe viruses here—for the most part, do they really discriminate distant conditions in the outside world, or conditions of their own organs (including their primitive "sense" organs and nascent nervous systems)? If the latter, then it may be that our own inner perception is leftover from this stage, not something added on after our ancestors became good at keeping track of the distal environment. And if inner perception is an ingredient in phenomenality, this would suggest that consciousness, far from being the icing on evolution’s cake, is widespread in the animal kingdom.

Questions about the driving forces of and constraints upon evolutionary design are usually very difficult to answer. Admittedly we have no convincing positive account of what inner perception would be good for, if anything. But what I would urge in the meantime is that the same mystery arises about phenomenality and phenomenal properties: why do we have them? The objection against inner perception could be a selling point for it as an account of phenomenality! Given how difficult it is to understand what functions phenomenal properties play, if any, it would be surprising if a philosophical theory of phenomenality appealed to a phenomenon with obvious functions.

Addressing doubts about a science of phenomenality

Let’s take stock. We have three puzzles blocking scientific explanation of consciousness—the argument from phenomenal objects (e.g., colored and shaped afterimages), the knowledge argument (what black-and-white science cannot reveal), and the explanatory-gap argument (that the presence of phenomenal states is unprovable). Among other things, these add up to a surprising threat to knowing what it’s like to have one’s mental states, or even to knowing whether there is anything it’s like at all. In order to avoid this skeptical separation between introspection and phenomenality, it’s worth exploring whether introspection is built right into phenomenality. So we’ve searched through available suggestions about introspection to find something serviceable. Despite the need for further elaboration, inner perception remains standing as an account of introspection that may also be sufficient or at least necessary for phenomenality. We’ll end by testing inner perception’s mettle against the three puzzles.

The problem about phenomenal objects is that in certain experiences one seems to be aware of little denizens of an inner mental world: e.g., colored and shaped mental "images", bodily "sensations" such as "pains" that may be throbbing or in one’s limb, and inner "speech" with "private" volume and pitch. In such experiences nothing in one’s brain or body or (causally relevant) environment is literally purple and round, literally throbbing and in a limb, or literally soft and medium-pitched. So if phenomenal objects do exist with these properties, they are not among the things in one’s brain or body or environment.

The only way to avoid dualism, then, is to be an "eliminativist" about mental entities with these properties, to deny that images and pains and inner speech with such properties exist. The challenge for the eliminativist about phenomenal objects is to explain why people are often tempted to claims of phenomenal objects, with ordinary perceptible properties. Afterimages look purple and round, pains feel dull or in motion, and inner speech seems to sound faint or high-pitched. For the eliminativist, these are best treated as illusions. An introspective account of experience, coupled with an inner-perceptual account of introspection, may help explain why we undergo these illusions. In an afterimage experience, for example, the inner-perception model posits two states: an ordinary outer-perceptual brain state representing (or misrepresenting) roundness and purpleness in the environment, and an inner-perceptual brain state representing the outer-perceptual brain state. Taken together, what do these two states "tell" the brain? That there is something purple and round, and that there something related going on in the brain (the outer-perceptual state). All that’s needed to generate the illusion is for these two bits of information to get confused, so that the brain (mis)treats its own state as purple and round, as a mental "image". And it would be natural for the brain to

confuse the two pieces of information, because the two states bearing the information go hand-in-hand; the outer-perceptual state causes the inner-perceptual state.

The knowledge argument turns on the idea that when the super-scientist Mary leaves her black-and-white room and first sees something red, she learns a new fact about experience—specifically, a fact about what it's like to see red—one that could not be on the list of scientifically describable physical facts she learns in the room. Most responses involve denying that Mary learns a new fact upon experiencing red. On some views, she learns how to do new things—to imagine experiencing redness or to recognize redness visually—without coming to know that any new fact obtains. On others, she learns that an old physical fact about experience obtains, but comes to know it in a new way—via introspective access. The inner-perception view helps to explain how this may be. Consider an analogy: Granny Smith knows there is an apple in Grandpa Jones' basket but has not perceived it yet. She can think of the apple by describing it physically—as "the smallest fruit in Jones' basket" or some such. When she finally does take it out and look at it, then she can think about the very same apple in a new way—"so this is the apple I was thinking of". She could not think of it simply as "this" before (while staring at something else). The existence of two distinct ways of thinking of apples—by "description" and by "demonstration"—does not mean there are two distinct apples thought about. Likewise, Mary inside the room knows there are facts about what it's like to see red, but she has not innerly perceived these facts yet. She can think of the facts by describing them physically—e.g., as the condition of having such-and-such brain state. When she finally gets the experience and innerly perceives it, she can think about the very same fact in a new way—"so this is what's it's like to see red". As with the apple, nothing here shows that there are two distinct facts Mary thinks about, nothing here shows that her list of facts in the room was incomplete.

Finally, what can be said about the alleged explanatory gap between the physical realm and phenomenal consciousness? The challenge is to come up with a scientific recipe for consciousness that enables us literally to prove that following it leads to genuine consciousness, just as we can prove that following the scientific recipe for water leads to genuine water. If we don't meet this standard, the objection goes, we don't have a satisfying explanation of consciousness.

At least two lines of response may be advanced against the explanatory gap argument. One concedes that we can never prove that a scientific recipe leads to consciousness, but insists that equally we cannot prove that the scientific recipe for

water leads to water. Suppose we follow the recipe—take two parts hydrogen and one part oxygen, stir until it reaches the desired consistency to erode rocks, to boil at 100 degrees, to quench thirst, to match what's in our lakes and rivers, etc.—and then ask whether what we have is water. Defenders of the explanatory gap argument say that a "yes" answer follows from our very concept of water, defined as whatever colorless liquid predominates in our lakes and rivers and quenches thirst, erodes rocks, etc. If so, then anyone who denies that the result of the recipe is water is implicitly contradicting himself. But perhaps these claims are overstated. Perhaps someone who believes in water but denies that it boils, is in lakes, etc. is making a false and bizarre claim but is not strictly contradicting himself. We have weak or strong beliefs that water is liquid at room temperature, freezes and boils, etc. But none of this is built into our concept of water. Compare: some people deny that scientific recipes for life must lead to genuine life, perhaps because they believe life requires a nonphysical "vital spirit" or "breath of life" as described in Genesis. Likewise, it is conceivable that water is "hydral spirit" as described in the yet-to-be-discovered Living Sea Scrolls. On this dualist view of water, water is not what erodes rocks and quenches thirst, but is instead a spiritual substance that hangs around physical stuff that does. The point is, scientists do not need to prove that this is impossible; they can instead justify rejecting it by appeal to simplicity, conservatism, and other familiar general grounds for comparing explanatory theories. A similar response should be available against dualists who posit "phenomenal spirit" or a "breath of consciousness", even if no scientific recipe for consciousness renders them inconceivable.

A second strategy tries to give the explanatory gap theorists what they say they want: a literal proof that certain scientific ingredients lead to genuine phenomenal cake. This strategy denies that there is an unbridgeable conceptual gap; in effect, it denies that zombies (nonphenomenal physical duplicates of phenomenal creatures) are even conceptually possible. Perhaps if we analyze the concept of phenomenality very carefully we can deduce that there is something it's like for a creature to have mental states, just from scientifically describable physical facts about the creature. What does it mean to say that a mental state is "like something" for its bearer? Compare: when we say a used car is "like new" for a customer, we mean that it appears new to the customer. Perhaps likewise, if a mental state appears some way to its bearer—as the inner-perception view of introspection suggests—then that proves that the state is "like something" for its bearer, i.e., is a phenomenally conscious state. If so, then the difficult but tractable task remaining for a science of consciousness would be to describe the mechanisms in the brain that are responsible for inner perception.

Suggestions for further reading

At the start I divided philosophers into three camps, those who think the problem of consciousness is embarassingly hard (e.g., because consciousness is nonphysical), those who think it's embarassingly easy (e.g., because ordinary and philosophical thought about consciousness is muddled), and those who think it's somewhere in between (e.g., difficult but tractable). The most extensive yet readable books from these three camps, respectively, are The Conscious Mind by David Chalmers (Oxford, 1996), Consciousness Explained by Daniel Dennett (Little Brown, 1991), and Ten Problems of Consciousness by Michael Tye (MIT, 1995). The best general collection of essays is The Nature of Consciousness edited by Ned Block et al. (MIT, 1997); see especially the papers by Georges Rey (on the threat to self-knowledge), and those in the final four sections on the explanatory gap (due to Joseph Levine), the knowledge argument (due to Frank Jackson), qualia, and inner monitoring of brain states.

While any of those four books would give a good overall sense of the debates, there are some residual issues they are not meant to cover. William Lyons provides a valuable historical survey of theories of introspection, including experimental results such as the bystander effect, in The Disappearance of Introspection (MIT, 1986). For an introduction to some relevant scientific background on blindsight, sleep, and other phenomena related to consciousness, try the papers in Consciousness in Contemporary Science edited by Anthony Marcel et al. (Oxford, 1988). Peter Carruthers defends a higher-order-belief theory of consciousness, and uses it to argue that animals do not experience pain, in The Animals Issue (Cambridge, 1992), which is mainly devoted to moral questions. Sidney Shoemaker mounts an extended attack on inner perception, and provides a rich but difficult discussion of self-knowledge, in The First-Person Perspective and Other Essays (Cambridge, 1996; his quote about "wired" introspection is on p. 222). My own defenses of inner perception, elaborations of the view, and criticisms of rival views are in the journals Philosophical Topics (1994), Nôus (June 1996), and on the internet at http://www-personal.umich.edu/~lormand/phil/cons.

A few more loose ends from the first paragraph of this paper. Despite giving up on conscious experience, Steven Pinker's How the Mind Works is an excellent introduction to the rest of cognitive science. If you're interested in what the Nobel laureates have to say, try The Remembered Present by Gerald Edelman (Basic Books, 1989) and The Astonishing Hypothesis by Francis Crick (Scribners, 1994), and for an equally eminent moonlighter, Roger Penrose's Shadows of the Mind

(Oxford, 1994). The optimistic quote about how philosophy can lead to science is Jerry Fodor’s, but I cannot find the reference.