

Consciousness, Philosophical Issues about

Ned Block
New York University

I. The Hard Problem

There are a number of different matters that come under the heading of ‘consciousness’. One of them is *phenomenality*, the feeling of say a sensation of red or a pain, that is what it is like to have such a sensation or other experience. Another is *reflection* on phenomenality. Imagine two infants, both of which have pain, but only one of which has a thought about that pain. Both would have phenomenal states, but only the latter would have a state of reflexive consciousness. This entry will start with phenomenality, moving later to reflexivity and then to one other kind of consciousness.

The Hard Problem of consciousness is how to explain a state of consciousness in terms of its neurological basis. If neural state N is the neural basis of the sensation of red, why is N the basis of that experience rather than some other experience or none at all? Chalmers (1996) distinguishes between the Hard Problem and “easy” problems that concern the function of consciousness. The Hard Problem (though not under that name) was identified by Nagel (1974) and further analyzed in Levine (1983).

There are two reasons for thinking that the Hard Problem has no solution.

1. **Actual Failure.** In fact, no one has been able to think of even a highly speculative answer.
2. **Principled Failure.** The materials we have available seem ill suited to providing an answer. As Nagel says, an answer to this question would seem to require an objective account that necessarily leaves out the subjectivity of what it is trying to explain. We don’t even know what would *count* as such an explanation.

II. Perspectives on the Hard Problem

There are many perspectives on the Hard Problem but I will mention only the four that comport with a naturalistic framework

1. *Eliminativism*, the view that consciousness as understood above simply does not exist (Dennett, 1979; Rey, 1997). So there is nothing for the Hard Problem to be about.
2. *Philosophical Reductionism* or *Deflationism*. Deflationists (for example, Dennett 1991) move closer to common sense by allowing that consciousness exists, but they “deflate” this commitment—again on philosophical grounds—taking it to amount to less than meets the eye (as Dennett might put it). One prominent form of deflationism in this sense makes a conceptual reductionist claim: that consciousness can be conceptually analyzed in non-phenomenal terms. The main varieties of analyses are behavioral, functional, representational and cognitive.
3. *Phenomenal Realism*, or *Inflationism* the view that consciousness is a substantial property that cannot be conceptually reduced or otherwise *philosophically* reduced in non-phenomenal terms. Logical behaviorists thought that we could analyze the concept of pain in terms of certain kinds of behavior, but inflationists reject all such analyses of phenomenal concepts in

non-phenomenal terms. According to most contemporary inflationists, consciousness plays a causal role and its nature may be found empirically as the sciences of consciousness advance. Inflationism is compatible with the *empirical scientific* reduction of consciousness to neurological or computational properties of the brain, just as heat was scientifically but not philosophically reduced to molecular kinetic energy. (It is not a conceptual truth that heat is molecular kinetic energy.) Inflationism (which is the view of the author) accepts the Hard Problem but aims for an empirical solution to it. (Block, 1995, Flanagan, 1992, Loar, 1997, Nagel, 1974.) One inflationist, McGinn, 1991, argues that an empirical reduction is possible but that we can't find or understand it. Another inflationist, Searle, 1992 endorses a roughly naturalistic point of view and rejects philosophical reduction of phenomenality, but he also rejects any empirical reduction of phenomenal properties.

4. *Dualistic Naturalism* In this catch-all category, I include Chalmers' (1996) view that standard materialism is false but that there are naturalistic alternatives to Cartesian dualism such as pan-psychism. Nagel (2000) proposes that there is a deeper level of reality that is the naturalistic basis both of consciousness and of neuroscience.

Here are some examples of deflationism in this sense. Pitcher (1971) and Armstrong (1968) can be interpreted as analyzing consciousness in terms of beliefs. One type of prototypical conscious experience as of blue is a matter of an inclination (perhaps suppressed) to believe that there is a blue object in plain view. (See Jackson, 1977 for a convincing refutation.) A different analysis appeals to higher order thought or higher order perception. These theorists take the concept of a conscious pain to be the concept of a pain that is accompanied by another state that is about that pain. A pain that is not so accompanied is not a conscious state. (Armstrong, 1968, Carruthers, 1992, Lycan, 1990) (Rosenthal, 1997 advocates a higher order thought view as an empirical identity rather than as a conceptual analysis.) Another deflationist view that is compatible with the analyses in terms of beliefs concerns not the states themselves but their contents. Representationism holds that it can be established philosophically that the phenomenal character of experience is its representational content. Many representationists reject conceptual analysis, but still their accounts do not depend on details of the science of mind; if any science is involved, it is evolutionary theory. (See Harman 1990, 1996; Dretske, 1995; Lycan 1996; McDowell, 1994, Tye 1995.) (Shoemaker 1994 mixes phenomenal realism with representationism in an interesting way.) Conceptual functionalists say that the concept of consciousness is analyzable functionally (Lewis, 1994).

The inflationist regards all of these deflationist accounts as leaving out the phenomenon. Phenomenality has a function and represents the world, but it is something over and above that function or representation. Something might function like a phenomenal state but be an ersatz phenomenal state with no real phenomenal character. The phenomenal character that represents red in you might represent green in me.

Phenomenal character does represent but it also goes beyond what it represents. Pain may represent damage but that is not what makes pain awful.

According to deflationism, there is such a thing as consciousness, but there is no Hard Problem, that is, there are no mysteries concerning the physical basis of consciousness that differ in kind from scientific problems about for example the physical/functional basis of liquidity, inheritance or computation.

III. Dissolution of the Hard Problem?

Suppose (to move away from the neurologically ridiculous example of c-fiber stimulation in favor of a view championed as a theory of visual experience by Crick and Koch, 1990) that cortico-thalamic oscillation (of a certain frequency) is the neural basis of an experience with phenomenal quality Q. Now there is a simple (over-simple) solution to the Hard Problem from a physicalist point of view. Phenomenal quality Q = cortico thalamic oscillation (of a certain sort). Here's a statement of the solution (or rather dissolution):

The Hard Problem is illusory. One might as well ask why H₂O is the chemical basis of water rather than gasoline or nothing at all. Just as water *is* its chemical basis, so Q is its neural basis, and that shows the original question is wrongheaded..

This point is correct as far as it goes, but it doesn't go far enough. It just puts the Hard Problem under a different part of the rug. For now we want to know: how could one property *be* both phenomenal property Q *and* cortico-thalamic oscillation. How is it possible for something subjective to be something objective or for something first-personal to be something third-personal? I am not suggesting that the problem is that we cannot find an explanation for this identity. There are no explanations for any identities. (Block, 1978a, Block and Stalnaker, 1999) The problem rather is that the claim that a phenomenal property is a neural property seems just as mysterious—maybe even more mysterious--than the claim that a phenomenal property has a certain neural basis. I would distinguish between explaining an identity and explaining how an identity can be true, where the latter involves removing a sense of puzzlement. Further, we can even see the same reasons given before as reappearing:

1. **Actual Failure** No one has even a wildly speculative answer to the question of how something objective can be something subjective.
2. **Principled Failure.** As before, actual failure doesn't seem accidental. The objective seems to necessarily leave out the subjective. The third-personal seems to necessarily leave out the first personal. Further, as McGinn (1991) notes, neural phenomenal are spatial, but the phenomenal is *prima facie* non-spatial.

The reasons that I mentioned above for thinking that the explanatory gap resists closing seem to surface in a slightly different form. However, as we shall see, in this form, they are more tractable.

IV. How to Begin Solving the Hard Problem

The main element in the approach I shall be suggesting to the Hard Problem is a distinction that is widely appealed to in discussions of Jackson's (1982) famous "Mary" example. Jackson imagined a neuroscientist of the distant future who is raised in a black and white room and who knows everything scientific there is to know about color and the experience of it. But when she steps outside the room for the first time, she learns what it is like to see red. Jackson argued that since the scientific facts don't encompass the new fact that Mary learns, dualism is true.

The line of response to Jackson that I think wins the day (and which derives from Loar, 1990/1997) involves appeal to the distinction between a property and a concept of that property. (See Churchland 1989; Loar, 1990/1997; Lycan 1990b; van Gulick, 1993; Sturgeon, 1994; Tye, 1999; Perry, forthcoming).

A concept is a thought element in a way of thinking, a kind of representation. If you like, you can take concepts in my sense to be interpreted symbols in a "language of thought". This usage contrasts with a common philosophical usage in which a concept is something like a meaning. Concepts in my sense are individuated in part by meanings; x and y are instances of the same concept if and only if they are instances of the same representation and have the same meaning. 'Water' and 'H₂O' are instances of different representations, so the concept *water* is distinct from the concept *H₂O*.

Someone could believe that *this color* is useful for painting pots but that red is not, even if *this color* = red. Our experiential concept of red differs from our linguistic concept of red. An experiential concept involves a phenomenal element, a phenomenal way of thinking, for example a mental image that is in some sense *of* red or an ability to produce such a mental image or at least an ability to recognize red—which arguably, could not be done without some phenomenal mental element.

Importantly, we can have an experiential concept *of an experience* (which we can call a phenomenal concept, phenomenal concepts being a subclass of experiential concepts) as well as an experiential concept of a color. And the very same mental image may be involved in both concepts. The difference between the phenomenal concept of the experience and the experiential concept of the color lies in the *rest* of the concept—in particular, the way the phenomenal element functions in the concept. This can be a matter of further concepts explicitly invoked in the concept—the concept of a color in one case and the concept of an experience in the other. One type of experiential concept (of a color or of an experience) involves a demonstrative plus a mental image plus a language-like representation, e.g. "that [attention to a mental image] color" or "that [attention to a mental image] experience" where the bracket notation is supposed to indicate the use of attention to a non-descriptive element, a mental image, in fixing the demonstrative reference. Loar (1990/97) gives an example involving two concepts in which something like a mental image of a cramp feeling is used to pick out a *muscular knot* in one concept and in the other concept, the cramp *experience* itself. The two concepts in my notation would be "that [attention to a mental image] cramp" and "that [attention to a mental image] cramp experience".

An experiential concept uses a phenomenal property to pick out a related property. For example, an experiential concept of a color can use an aspect of the experience of that color to pick out the color. A phenomenal concept uses a phenomenal property to pick out a phenomenal property. The phenomenal property used in the concept needn't be the same as the one picked out. For example, one could have a phenomenal concept of the experience of the missing shade of blue whose phenomenal elements are the flanking color experiences. Or a phenomenal element involved in one's perception of a color could be used to pick out the experience of the complementary color. Importantly, the phenomenal element in a phenomenal concept needn't be and cannot in general be conceptualized, at least if the conceptualization is supposed to be itself phenomenal. For if a phenomenal concept had to make use of a phenomenal element via a distinct phenomenal concept of that element, there could be no phenomenal concepts. Thus we can define a phenomenal concept as a concept of a phenomenal property that uses a phenomenal property to pick out a phenomenal property but not necessarily under a concept of the phenomenal property used.

In these terms, then, Mary acquired a new concept of a property she was already acquainted with via a different concept. In the room, Mary knew about the subjective experience of red via physical concepts. After she left the room, she acquired a phenomenal concept of the same property. So Mary did not learn a new fact. She learned a new concept of an old fact. She already had a third person understanding of the fact of what it is like to see red. What she gained was a first person understanding of the very same fact. She knew already that cortico-thalamic oscillation of a certain frequency is what it is like to see red. What she learned is that this[attention to a mental image] is what it is like to see red. So the case provides no argument that there are facts that go beyond the physical facts.

Recall that there is a principled reason why mind-body identity seemed impossible: that a first person subjective property could not be identical to a third person objective property. But the concept/property distinction allows us to see that the subjective/objective distinction in this use of it and the first person/third person distinction *are distinctions between kinds of concepts, not kinds of properties*. There is no reason why a subjective concept and an objective concept cannot pick out the same property. *Thus we can substitute a dualism of concepts for a dualism of properties*.

There is another way in which the concept/property distinction helps with the Hard Problem. We can blame the explanatory gap and the Hard Problem on our inadequate concepts rather than on dualism. To use a variant on Nagel's (1974) example, we are like pre-Socratics who have no way of understanding how it is possible that heat = mean molecular kinetic energy, lacking the concepts required to frame both sides of the equation. (Heat was not clearly distinguished from temperature until the 17th Century.) What is needed is a concept of heat and a concept of kinetic energy that makes it conceivable that there is a causal chain of the referential sort leading from the one magnitude to each concept. Or rather, since the phenomenal concept includes a sample of the relevant phenomenal property (on the Humean simplification I am using), there is

no mystery about the mental side of the equation. The mystery is how the physical concept picks out that phenomenal property. This is the remaining part of the explanatory gap that will be closed if at all by science. Is there a principled reason to think it cannot be? The Hard Problem itself does not contain such a reason. Perhaps our conceptual inadequacy is temporary, as Nagel sometimes appears to suppose, or perhaps it is permanent as McGinn (1991) supposes.

V. The Paradox of Recent Empirical Findings About Consciousness

I will now switch gears to a discussion of recent empirical findings on consciousness. The most exciting line of experimental investigation of consciousness in recent years uses phenomena in which perception changes independently of the stimulus. One such paradigm uses binocular rivalry. If two different stimuli—e.g. horizontal and vertical stripes—are presented to each of one's eyes, one does not see a blend, but rather *first* horizontal stripes that fill the whole visual field and then vertical stripes, that fill the whole field. Logothetis and his colleagues (Logothetis, 1998) trained monkeys to pull different levers for different patterns. They then presented different patterns to the monkeys' two eyes, and observed that with monkeys as with people, the monkeys switched back and forth between the two levers even though the sensory input remained the same. Logothetis recorded the firings of various neurons in the monkeys' visual systems. In the lower visual areas (e.g. V1), 80% of the neurons did not shift with the percept. But further along the occipital-temporal pathway, 90% shifted with the percept. So it seems that later areas in the occipital-temporal pathway—let's call it the "ventral stream"—are more dominantly part of the neural basis of (visual) consciousness than early areas. Recent work using imaging has extended and refined these findings. Kanwisher (2001) notes that "...neural correlates of perceptual experience, an exotic and elusive quarry just a few years ago, have suddenly become almost commonplace findings". And she backs this up with impressive correlations between neural activation on the one hand and indications of perceptual experiences of faces, houses, motion, letters, objects, words, speech on the other. Although the neural correlates of, say, faces and houses are somewhat different, both are in the ventral stream, mainly in the higher areas. So we have an amazing success: identification of the neural basis of visual consciousness in the ventral stream.

Paradoxically, what has also become commonplace is activation of the *very same ventral stream pathways without "awareness"*. Damage to the inferior parietal and frontal lobes has long been known to cause visual extinction—in which subjects appear to lose subjective experience of certain stimuli on one side, viz., when there are stimuli on both sides) Extinction is associated with visual neglect in which subjects don't notice stimuli on one side. For example, neglect patients often don't eat the food on the left side of the plate. Although subjects say they don't see the extinguished stimulus, the nature of the stimulus has all sorts of visual effects. For example, if the subject's task is to decide whether a letter string (e.g. 'saucer' or 'spiger') is a word, the subject is faster for 'saucer' if there is a picture of a cup or the word 'cup' in the neglected field, even though they are at chance in guessing what the picture depicts (McGlinchey.-Berroth, et. al, 1996). So the stimulus of which the subject is in some sense unaware is processed semantically.

Driver and Vuilleumier (2001) point out that the ventral stream is activated for extinguished stimuli (i.e., which the subject claims not to see). Rees et. al. 2000 report studies of a left sided neglect and extinction patient on face and house stimuli. Stimuli presented just on the left side are clearly seen by the patient, but when there are stimuli on both sides, the subject says he sees only the stimulus on the right. However, the “unseen” stimuli show activation of the ventral pathway that is the same in location and temporal course (though lower in activation) as the seen stimuli. Further, studies in monkeys have shown that a classic “blindness” syndrome is caused by massive cortical ablation that spares most of the ventral stream but not inferior parietal and frontal lobes (Lumer and Rees, 1999). Kanwisher (2001) notes that dynamic visual gratings alternating with a gray field showed activation for V1, V2, V3A, V4v and MT/MST despite the subjects saying they saw only a uniform gray field.

Is the difference between conscious and unconscious activation of the ventral pathway just a matter of the degree of activation? Zeki and Ffytch, 1998 hypothesize that this is so. But Kanwisher (2001) mentions that evidence from ERP studies using the attentional blink paradigm show that neural activation of meaning is no less when the word is blinked than when it isn't, suggesting that it is not lower neural activation strength that accounts for lack of awareness. Dehaene and Naccache (2001) note that in a study of neglect patients, it was shown that there is the same amount of semantic priming from both hemifields, despite the lack of awareness of stimuli in the left field, again suggesting that it is not activation strength that makes the difference.

The paradox then is that our amazing success in identifying the neural correlate of visual experience in normal vision has led to the peculiar result that in masking and neglect, that very neural correlate occurs without, apparently, subjective experience.

What is the missing ingredient, X, which, added to ventral activation, constitutes conscious experience? Kanwisher (2001) and Driver and Vuilleumier (2001) offer pretty much the same proposal as to the nature of X, that the missing ingredient is binding perceptual attributes with a time and a place, a token event. Rees, et. al. 2000 make two suggestions as to what X is. One is that the difference between conscious and unconscious activation is a matter of neural synchrony at fine timescales. This idea is supported by the finding that ERP components P1 and N1 revealed differences between left sided “unseen” stimuli and left sided seen stimuli. Their second suggestion is that the difference between seen and “unseen” stimuli might be a matter of interaction between the classic visual stream and the areas of parietal and frontal cortex that control attention.

Whether or not any of these proposals are right, the search for X seems to me the most exciting current direction for consciousness research. For purposes of the discussion below, I will assume that X = neural synchrony.

VI. Physicalism and Functionalism

There is a very different approach to the nature of consciousness. Dennett 1994 postulates that consciousness is “cerebral celebrity”. What it is for a representation to be

conscious is for it to be widely available in the brain. Dehaene and Naccache (2001) say consciousness is being broadcast in a global neuronal workspace.

The theory that consciousness is ventral stream activation plus neural synchrony, and the theory that consciousness is broadcasting in the global neuronal workspace are instances of the two major approaches to consciousness in the philosophical literature, *physicalism* and *functionalism*. The difference is that the functionalist says that consciousness is a role, whereas the physicalist says that consciousness is a physical or biological state that *implements* that role. To see the distinction, note that red may play the role of warning of danger. But green might also have played that role. The picture of consciousness as role could be characterized as *computational*—as contrasted with the *biological* picture of consciousness as implementer of the role.

Although functionalists are free to add restrictions, functionalism in its pure form is *implementation independent*. Consciousness is defined as global accessibility, and although its human implementation depends on our biochemistry, silicon creatures without our biochemistry could implement the same computational relations. Functionalism and physicalism are incompatible doctrines, since a non-biological implementation of the functional organization of consciousness would be regarded as uncontroversially conscious by the functionalist but not by the physicalist. The big question for functionalists is this: “How do you know that it is broadcasting in the global workspace that makes a representation conscious as opposed to something about the *human biological realization* of that broadcasting?”

The problem for functionalists could be put like this: the specifically human biochemical realization of global availability may be necessary to consciousness—other realizations of global availability being “ersatz” realizations. The typical response to this “ersatz realization problem” (Lycan, 1981) is that we can preserve functionalism by simply bringing lower level causes and effects into our functional characterizations, e.g. causes and effects at the level of biochemistry. But the utility of this technique runs out as one descends the hierarchy of sciences because the lowest level of all, that of basic level physics, is vulnerable to the same point. Putting the point for simplicity in terms of the physics of forty years ago, the causal role of electrons is the same as that of anti-electrons. If you formulate a functional role for an electron, an anti-electron will realize it. Thus an anti-electron is an ersatz realizer of the functional definition of *electron*. Physics is characterized by symmetries that allow ersatz realizations. (Block, 1978).

We have been talking about the two approaches of functionalism and physicalism as rivals, but we can instead see them as answers to different questions. The question that motivates the physicalist proposal of ventral activation plus X is: What is the neural basis of experience? The question that motivates the global broadcasting type is: What makes neuronal representations available for thought, decision, reporting and action? To attach names to the concepts, the former is a theory of *phenomenal* consciousness (what I have been calling phenomenality), and the latter *access-consciousness*. (Theorists will differ of course on whether access-consciousness is really a type of consciousness. See Burge, 1997.) We can try to force a unity by *postulating* that it is a condition on X that it

promote access, but that is merely a verbal maneuver that only throws smoke over the difference between the different concepts and questions. Alternatively, we could, *hypothesize* rather than postulate that X as a matter of fact is the neural basis of global neuronal broadcasting. Note, however, that the neural basis of global neuronal broadcasting might obtain but the normal channels of broadcasting nonetheless be blocked or cut, again opening daylight between phenomenality and access, and showing that we cannot think of the two as one. (An analogy: rest mass and relativistic mass are importantly different from a theoretical point of view despite coinciding for all practical purposes at terrestrial velocities. Failure of coincidence even if rare is theoretical dynamite if what you are after is the scientific nature of consciousness.)

Many of us have had the experience of suddenly noticing a sound (say a jackhammer during an intense conversation) at the same time realizing that the sound has been going on for some time even though one was not attending to it. If the subject did have a phenomenal state before the sound was noticed, that state was not broadcast in the global neuronal workspace *until it was noticed*. If this is right, there was a period of *phenomenality without broadcasting*. Of course, this is anecdotal evidence. But the starting point for work on consciousness is introspection and we would be foolish to ignore it.

If we take seriously the idea of phenomenality without global accessibility, there is a theoretical option that should be on the table—that ventral stream activation is visual phenomenality and the search for X is the search for the neural basis of what makes visual phenomenality *accessible*. The idea would be that the claims of extinction patients not to see extinguished stimuli are in a sense wrong—they really do have phenomenal experience of these stimuli without knowing it. A similar issue will arise in the section to follow in which I will focus on the relation between phenomenality and a special case of global accessibility, reflexive or introspective consciousness, in which the subject not only has a phenomenal state but also has another state that is about the phenomenal state, say a thought to the effect that he has a phenomenal state.

In this section, we have seen a distinction between two concepts of consciousness, phenomenality and global accessibility. In the following section, we add a third, reflexivity.

VII. Phenomenality and Reflexivity

Consider the “false recognition” paradigm of Jacoby and Whitehouse (1989). Subjects are given a study list of 126 words presented for half a second each. They are then presented with a masked word, word₁, and an unmasked word, word₂. Their task is to report whether word₂ was old (i.e. on the study list) or new (not on the study list). The variable was whether word₁ was lightly or heavily masked, the former presentations being thought of as “conscious” and the latter as “unconscious”. The result: confining our attention just to cases in which word₁ = word₂, subjects were much more likely to mistakenly report word₂ as old when word₁ was unconsciously presented than when word₁ was consciously presented. The explanation would appear to be that when word₁ was consciously presented, the subjects were able to use an internal monologue of the

following sort (though perhaps not quite as explicit): “The reason ‘reason’ (word₂) looks familiar is that I just saw it (as word₁),” thereby explaining away the familiarity of word₂. But when word₁ was *unconsciously* presented, the subjects were not able to indulge in this monologue and consequently blamed the familiarity of word₂ on its appearance in the study list.

Any monologue that can reasonably be attributed to the subject in this paradigm concerns the subject thinking about why a word (word₂) *looks familiar* to the subject. For it is only by *explaining away* the familiarity of word₂ that the subject is able to decide that word₂ was not on the study list. Thus in the “conscious” case, the subject must have a state that is *about the subject’s own perceptual experience* (looking familiar) and thus conscious in what might be termed a “reflexive” sense. An experience is conscious in this sense just in case it is the object of another of the subject’s states; for example, one has a thought to the effect that one has that experience. The reflexive sense of ‘consciousness’ contrasts with phenomenality, which perhaps attaches to some states which are not the objects of other mental states. Reflexive consciousness might better be called ‘awareness’ than ‘consciousness’. Reflexivity is phenomenality plus something else (reflection) and that opens up the possibility in principle for phenomenality without reflection. For example, it is at least conceptually possible for there to be two people in pain, one of whom is introspecting the pain the other not. (Perhaps infants or animals can have pain but don’t introspect it.) The first is reflexively conscious of the pain, but both have phenomenally conscious states, since pain is by its very nature a phenomenally conscious state.

What is the relation between reflexivity and the notion of global accessibility discussed in the last section? Global accessibility does not logically require reflexivity, since global accessibility only requires access to all response modes that the organism actually has. (Perhaps a dog or a cat does not have the capacity for reflection.) Reflexivity is a special kind of access, one that requires intellectual resources that may not be available to every being that can have conscious experience.

There is another aspect to the experimental paradigm just discussed which motivates taking seriously the hypothesis that the reflexively *unconscious* case might possibly be phenomenally *conscious*. Consider a variant of the exclusion paradigm reported by Debner and Jacoby (1994). Subjects were presented with pairs of words flanked by digits, e.g. ‘4reason5’, and then given stems consisting of the first three letters of the word (‘rea___’) to complete. There were two conditions. In the “conscious” condition, they were told to ignore the digits. In the “unconscious” condition, they were told to report the sum of the digits before completing the stem. The results were that in the “conscious” condition, the subjects were much more likely than baseline to follow the instructions and complete the stem with a word other than ‘reason’, whereas with “unconscious” presentations, subjects were much more likely than baseline to violate the exclusion instructions, completing the stem with ‘reason’ Merikle and Joordens (1997) report corresponding results for the false recognition paradigm with divided attention substituted for heavy masking.. Consider the hypothesis that there was a fleeting

phenomenal consciousness of ‘reason’ as the subject’s eyes moved from the ‘4’ to the ‘5’ in ‘4reason5’.

What are the theoretical options that deserve the most consideration?

1. The “unconscious perceptions” are *both* phenomenally and reflexively unconscious. (In this case, the exclusion and false recognition paradigms are about consciousness in both senses.)
2. The “unconscious perceptions” are fleetingly phenomenally conscious but reflexively unconscious.

A third option, that they are phenomenally unconscious but “reflexively conscious” seems less likely because the reflective consciousness would be “false”—that is subjects would have a state “about” a phenomenal state without the phenomenal state itself. That hypothesis would require some extra causal factor that produced the false recognition and would thus be less simple. One argument in favor of 2 is that subjects in experiments with near-threshold stimuli often report a mess of partial perceptions that they can’t hang on to. Some critics have disparaged the idea of fleeting phenomenal consciousness (e.g. Dennett, 1991). But what they owe us is a reason to think that 1 is the default view.

But it might seem that there is a principled argument that we could *never* find out about phenomenality in the absence of reflexive consciousness, for don’t we require the subject’s testimony about phenomenality and doesn’t that require the subject to have a state that is about the phenomenal state? We can see what is wrong with this reasoning by attention to some potential lines of evidence for phenomenality in the absence of reflexive consciousness.

Liss (1968) contrasted subjects’ responses to brief unmasked stimuli (1-4 letters) with their responses to longer lightly masked stimuli. He asked for judgements of brightness, sharpness and contrast as well as what letters they saw. He found that lightly masked 40 msec stimuli were judged as brighter and sharper than unmasked 9 msec stimuli, even though the subjects could report 3 of 4 of the letters in the unmasked stimuli and only 1 of 4 in the masked cases. He says: “The Ss commented spontaneously that, despite the high contrast of the letters presented under backward masking, they seemed to appear for such brief duration that there was very little time to identify them before the mask appeared. Although letters presented for only 7 msec with no masking appeared weak and fuzzy, their duration seemed longer than letters presented for 70 msec followed by a mask.” (p.329)

Perhaps the subjects were phenomenally conscious of all the masked letter shapes, but could not apply the letter concepts required for reflexive consciousness of them. (The subjects could apply the concepts of sharpness, brightness and contrast to the letters, so they did have reflexive consciousness of those features, even if they did not have reflexive consciousness of the shapes themselves. There are two kinds of shape concepts that could have provided—but apparently didn’t provide—reflexive consciousness of the letters: the letter concepts that we all learned in grade school, and shape concepts of the sort we have for unfamiliar shapes.) The underlying view is that phenomenal experience of shapes does not require shape concepts but reflexive consciousness being

an intentional state does require shape concepts, concepts that the subjects seem unable to access in these meager attentional circumstances. Alternatively, perhaps the phenomenal experience of shapes does involve shape concepts of some sort but the use of those shape representations in reflexive consciousness requires more attentional resources than were available to these subjects.

There is another hypothesis—that the contents of both the subject's phenomenal states and their reflective states are the same and include the features sharp, high contrast, bright and letter-like without any specific shape representation. On this hypothesis, there is no gap between phenomenal and reflexive consciousness. Both hypotheses have to be taken seriously, but the first is superior in one respect. Anyone who has been a subject in this or in Sperling's (1960) similar experiment will feel that the last hypothesis does not really capture the experience, which is one of being able to see more letters than one can categorize.

Rosenthal (1997) defines reflexive consciousness as follows: S is a reflexively conscious state of mine \leftrightarrow S is accompanied by a thought--arrived at non-inferentially and non-observationally-- to the effect that I am in S. He offers this "higher order thought" (HOT) theory as a theory of phenomenal consciousness. It is obvious that phenomenal consciousness without HOT and HOT without phenomenal consciousness are both *conceptually* possible. For examples, perhaps dogs and infants have phenomenally conscious pains without higher order thoughts about them. For the converse case, imagine that by bio-feedback and imaging techniques of the distant future, I learn to detect the state in myself of having the Freudian unconscious thought that it would be nice to kill my father and marry my mother. I could come to know—non-inferentially and non-observationally—that I have this Freudian thought even though the thought is not phenomenally conscious. Since there are conceptually possible counterexamples in both directions, the issue is the one discussed above of whether reflexivity and phenomenality come to the same thing empirically.

If there are no actual counterexamples, the question arises of why. Is it supposed to be a basic law of nature that phenomenality and reflexivity co-occur? That would be a very adventurous claim, one that no one is in a position to make. Is it a contingent fact about us but not other phenomenally conscious creatures? Then reflexivity would not provide a basic account of phenomenality. Well then, is it supposed to be a contingent fact that phenomenality and reflexivity are coextensive in *all* creatures? What would be the evidence for such a far reaching claim? Further, if reflexivity and coextensivity are correlated, then there must be a mechanism that explains the correlation, as the fact that both heat and electricity are carried by free electrons explains the correlation of electrical and thermal conductivity. But any mechanism breaks down under extreme conditions, as does the correlation of electrical and thermal conductivity at extremely high temperatures. So the correlation between phenomenality and reflexivity would break down too, showing that reflexivity does not yield the fundamental scientific nature of phenomenality.

Another option: it might be said that ‘consciousness’ is a “natural kind” term like ‘water’ or ‘heat’ or ‘light’. We know that water is H₂O, as a matter of empirical fact without having to ask whether there might be a substance in another solar system that has a very different chemical constitution but nonetheless behaves exactly like water. The reason is that as a matter of the *semantics* of the word ‘water’, the question of whether there could be a substance that behaved exactly like water but that had a completely different chemical constitution doesn’t matter since it would be wrong to *call* it ‘water’. But this is a poor analogy for two reasons. First, if there are beings in another solar system who have something that feels like phenomenality but without reflexivity, then reflexivity does not capture the fundamental nature of phenomenality, whatever we *call* what the aliens have. Second, we call anything that feels like phenomenality ‘phenomenality’, so phenomenality is not a natural kind concept in the sense that ‘water’ is.

Rosenthal’s definition of reflexivity has a number of ad hoc features. “Non-observationally” is required to rule out (e.g.) a case in which I know about a thought I have repressed by observing my own behavior. “Non-inferentially” is needed to avoid a somewhat different case in which I appreciate (non-observationally) my own psychic pain and infer a repressed thought from it. But why should the consciousness of a state depend on its causal history? Further, Rosenthal’s definition involves a stipulation that the possessor of the reflexively conscious state is the same as the thinker of the thought—otherwise *my* thinking about *your* pain would make it a conscious pain. All these ad hoc features can be eliminated by moving to the following definition of reflexivity: S is a reflexively conscious state ↔ S is phenomenally presented in a thought about S. This definition uses the notion of phenomenality, but this is no disadvantage unless one holds that there is no such thing apart from reflexivity itself. The new definition of reflexivity, requiring phenomenality as it does, has the additional advantage of making it clear that reflexivity is a kind of *consciousness*. (See Burge’s 1997 critique of my definition of access-consciousness as constituting a kind of consciousness.)

We have seen three concepts of consciousness, phenomenality, reflexivity and global accessibility. The Hard Problem arises only for phenomenality. The imaging work on consciousness engages phenomenality and accessibility. But many psychological experimental paradigms mainly engage reflexivity. So empirical investigators of “consciousness” may sometimes be talking past each other.

Glossary

Phenomenality: what it is like to have an experience, e.g. the feature of the taste of chocolate that one enjoys.

Reflexivity: A reflexively conscious state is one that is phenomenally presented in a thought about that state. Alternatively, S is a reflexively conscious state of mine ↔ S is accompanied by a thought--arrived at non-inferentially and non-observationally-- to the effect that I am in S.

Access-consciousness: global availability

Deflationism: the view that phenomenality can be conceptually reduced or otherwise philosophically reduced to function, representation or cognition.

Inflationism: phenomenal realism, the view that if phenomenality is reducible to function, representation or cognition, the type of reduction is scientific rather than conceptual or otherwise philosophical.

Qualia: a name often given to the phenomenal character of experience on an inflationist view of it. So the debate about the existence of qualia is really a debate about inflationism.

The Hard Problem: the problem of how to explain phenomenality physically.

Consciousness: Phenomenality or reflexivity or global availability

Bibliography

- Armstrong, D.M., 1968. *A Materialist Theory of the Mind*. London: Routledge & Kegan Paul
- Block, Ned, 1978a "Reductionism," *Encyclopedia of Bioethics*, Macmillan, 1978, 1419-1424
- Block, Ned, 1978b "Troubles with Functionalism". *Minnesota Studies in the Philosophy of Science* (C.W. Savage, ed.), Vol. IX, 1978,261-325. Reprinted abridged in D.M. Rosenthal (Ed.), *The Nature of Mind*, Oxford University Press: Oxford, 1991, pp.211-229
- Block, Ned, Flanagan, Owen & Güzeldere, Güven, 1997: *The Nature of Consciousness: Philosophical Debates*, Cambridge: MIT Press.
- Block, Ned and Stalnaker, Robert, "Conceptual Analysis and the Explanatory Gap", *The Philosophical Review* January, 1999
- Burge, Tyler, 1997: "Two Kinds of Consciousness" in Block, Flanagan and Güzeldere, 1997.
- Carruthers, Peter, 1992. "Consciousness and Concepts" *Proceedings of the Aristotelian Society*, supplement 66:41-59
- Chalmers, D. 1996. *The Conscious Mind*. New York: Oxford University Press.
- Churchland, Paul, 1989. "Knowing Qualia: A Reply to Jackson". From *A Neurocomputational Perspective* (Cambridge: MIT Press), 67-76
- Crick, F. and Koch, C. 1990. 'Towards a neurobiological theory of consciousness.' *Seminars in the Neuroscience*, 2:263-275.
- Dehaene, S. and Naccache, L., 2001 Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework, *Cognition* 79, 1-2. pp 1-37
- Dennett, Daniel, 1979. "On the Absence of Phenomenology" in D. Gustafson and B. Tapscott, eds., *Body, Mind and Method: Essays in Honor of Virgil Aldrich*, Dordrecht: Reidel.
- Dennett, Daniel, 1991. *Consciousness Explained*. Boston: Little Brown
- Dennett, Daniel, 1994 " "Get Real," in *Philosophical Topics*, vol. 22, no 1 &2, Spring & Fall
- Dretske, Fred, 1995. *Naturalizing the Mind* Cambridge: MIT Press
- Flanagan, O. 1992 *Consciousness Reconsidered* MIT Press: Cambridge
- Guzeldere, Guven, 1997 "The Many Faces of Consciousness: A Field Guide" in Block, Flanagan and Güzeldere, 1997, 1-67
- Harman, Gilbert, 1990. "The Intrinsic Quality of Experience" In James Tomberlin (ed.) *Philosophical Perspectives* 4, *Action Theory and Philosophy of Mind*, 31-52 Atascadero: Ridgeview

- Huxley, T. H. 1866 *Lessons in Elementary Physiology*. London: MacMillan
- Jackson, Frank, 1977. *Perception*. Cambridge University Press: Cambridge
- Jackson, Frank, 1982: "Epiphenomenal Qualia" *Philosophical Studies* 32, 127-36
- Jackson, Frank, 1986 "What Mary didn't Know". *Journal of Philosophy* 83:291-5.
- Jackson, F. 1993. "Armchair Metaphysics". In *Philosophy in Mind*, ed. J. O'Leary-Hawthorne and M. Michael, Dordrecht: Kluwer
- Kanwisher, Nancy, 2001, "Neural Events and Perceptual Awareness" *Cognition* 79, 1-2, pp 89-113
- Loar, Brian, 1990: "Phenomenal States". In James Tomberlin (ed.) *Philosophical Perspectives* 4, *Action Theory and Philosophy of Mind*, Atascadero: Ridgeview, 81-108. A much-revised version of this paper is to be found in Block, Flanagan and Guzeldere, *The Nature of Consciousness: Philosophical Debates*, Cambridge: MIT Press, 1997, 597-616
- Pitcher, George, 1971. *A Theory of Perception*. New Jersey: Princeton University Press
- Jacoby, L.L. & Whitehouse, K. 1989. "An Illusion of Memory: False Recognition influenced by Unconscious Perception". *Journal of Experimental Psychology: General* 118, 126-135.
- Levine, J. (1983) 'Materialism and qualia: the explanatory gap.' *Pacific Philosophical Quarterly* 64:354-361
- Lewis, David, 1994, "Lewis, David: Reduction of Mind" " in *Blackwell's Companion to Philosophy of Mind*, S. Guttenplan (ed). Blackwell: Oxford, 1994, 412-431
- Liss, P., 1968. "Does backward masking by visual noise stop stimulus processing?" *Perception & Psychophysics* 4, 328-330
- Logothetis, N. K. (1998). Single units and conscious vision. *Proc. R. Soc. London Ser. B.* 353, 1801-1818.
- Lumer, Erik and Rees, Geraint, 1999. "Covariation of activity in visual and prefrontal cortex associated with subjective visual perception", *Proc. Natl. Acad. Sci.* Vol 96, pp. 1669-1673
- Lycan, William, 1981. "Form, Function and Feel", *Journal of Philosophy* 78:24-50
- Lycan, William, G. 1990 "Consciousness as Internal Monitoring, *Philosophical Perspectives* 9, ed. J. Tomberlin (Atascadero: Ridgeview), 1-14. Reprinted in Block, Flanagan and Guzeldere, 1997
- Lycan, William G. 1996: *Consciousness and Experience* Cambridge: MIT Press
- McDowell, John, 1994 "The Content of Perceptual Experience" *Philosophical Quarterly* April
- McGinn, C. (1991) *The Problem of Consciousness*. Blackwell
- Nagel, T. (1974) 'What is it like to be a bat?' *Philosophical Review*
- Nagel, T., 2000. "The Psychophysical Nexus," in Paul Boghossian and Christopher Peacocke, eds., *New Essays on the A Priori*, Oxford: Oxford University Press, 2000
- McGlinchey-Berroth, R., W.P. Milberg, M. Verfaellie, L. Grande, M. D'Esposito, M. Alexander, 1996. "Semantic Processing and Orthographic Specificity in Hemispatial Neglect". *Journal of Cognitive Neuroscience* 8: 291-304
- Merikle, Philip and Joordens, Steve, 1997. "Parallels between Perception without Attention and Perception without Awareness. *Consciousness and Cognition* 6, 219-236

- Rees, Geraint, Wojciulik, Ewa, Clarke, Karen, Husain, Masud, Frith, Chris, Driver, Jon, 2000. "Unconscious Activation of Visual Cortex in the Damaged Right Hemisphere of a Parietal Patient with Extinction, *Brain*. In press
- Perry, John, 2001. *Knowledge, Possibility and Consciousness*. MIT Press: Cambridge
- Putnam, Hilary, 1967. "Psychological Predicates", later titled "The nature of mental states". In (Capitan & Merrill, eds) *Art, Mind, and Religion*. Pittsburgh University Press. Reprinted in *Mind, Language, and Reality* (Cambridge University Press, 1975).
- Rey, Georges, 1997. *Contemporary Philosophy of Mind*. Blackwell: Oxford
- Rosenthal, David, 1997. "A Theory of Consciousness", in N. Block, O. Flanagan, G. Guzeldere (eds), *Consciousness*. MIT Press, 729-754
- Searle, John 1992. *The Rediscovery of the Mind*. MIT Press:Cambridge
- Shoemaker, Sydney 1994 "Self-Knowledge and "Inner Sense; Lecture III: The Phenomenal Character of Experience". *Philosophy and Phenomenological Research* LIV, 2, June 1994, 291-314
- Smart, J.J.C., 1959, "Sensations and Brain Processes", *Philosophical Review* LXVIII, 141-156. This paper has been reprinted many times starting in 1962 in a somewhat revised form. See for example, *Materialism and the Mind-Body Problem*, D.M. Rosenthal (ed.), Prentice-Hall: Englewood Cliffs, 1971
- Sperling, George, 1960. "The information available in brief visual presentations." *Psychological Monographs* 74, 11, 1-29
- Strawson, Galen, 1994 *Mental Reality*. Cambridge: MIT Press
- Sturgeon, Scott, 1994. "The Epistemic View of Subjectivity" *Journal of Philosophy* XCI, 5, 1994
- Tye, Michael, 1995 *Ten Problems of Consciousness* Cambridge: MIT Press
- Tye, Michael, 1999 "Phenomenal Consciousness: The Explanatory Gap as a Cognitive Illusion," *Mind* 108, Number 432, 706-725
- Van Gulick, Robert, 1993 Understanding the phenomenal mind: Are we all just armadillos? In (M. Davies and G. Humphreys, eds) *Consciousness: Psychological and Philosophical Essays*. Blackwell. Reprinted in Block, et.al, 1997

Suggestions for Further Reading

Block, Flanagan and Guzeldere, 1997 contains many papers on consciousness that are not mentioned in the text. The April, 2001 issue of the journal *Cognition* (Vol. 79, numbers 1-2) is devoted to state of the art articles on consciousness. The journal *Neuropsychologia* will be publishing a state of the art issue in late 2001 or early 2002.