Philosophy 340 W. Lycan
Spring, 2015

# CONSCIOUSNESS

What is consciousness? Bad question. What *are* consciousness? The term has many senses.

"Creature" consciousness: A thing is a conscious *being* or creature iff it has the capacity for mentation.

"Minimal" consciousness: A thing has that iff it is (actually) hosting any mental activity at all.

"Control" consciousness: What psychologists call the "normal waking state"; you're awake and functioning and have normal control of your actions.

"State" consciousness: A mental state is a "conscious" state when one is directly aware of being in it: "a conscious memory," "a conscious decision." [Careful, though: the term "conscious state" has been used in at least two other ways.]

Those senses are not interesting or mysterious. What people think is mysterious is *subjectivity* in one form or another. So let's turn to "state" consciousness: A mental state is a **conscious state** (in one sense of the term!) just in case *its subject is directly (from the inside) aware of being in it*. (Again, as in a conscious memory or a conscious decision.) That, I think, is the beginning of subjectivity. And it is the easiest issue in that area, so let's start with it.

<u>Awareness of one's own mental states</u>

Rosenthal contrasts two "conceptions of" (really *doctrines* regarding) consciousness in the present sense. One is Descartes': Descartes simply identified the mental with the conscious; there's no such thing as a mental state you're unaware of being in. A mental state and your awareness of it are one and the same thing. Mentality just *is* awareness.

But an opposing view was generated by two kinds of 20th-century psychology: There are representational, mental states that we are unaware of being in—repressed beliefs, desires and intentions (Freud), and/or computational information-processing states, as in your language module or in early vision (cognitive psychology).

For that matter, we sometimes just fail to notice sensations that we have, because our attention is directed elsewhere. A pain can go unfelt, if you're a soldier in combat or just playing in a fast and violent basketball game. (You realize later that the injury must have hurt, but you weren't aware of it at the time.) Armstrong gives the example of the long-distance truck driver, driving on automatic pilot while thinking about something else, who "comes to" and realizes that he hasn't been paying attention to his driving. He must have seen the stop lights, etc., or he would have crashed. But he wasn't aware of seeing a red light at the time.

The following discussion assumes that the second, antiCartesian doctrine is correct. But the doctrine remains controversial. I'll be delighted if one or more of you devotes a paper to defending Descartes on this.

So, Rosenthal and Armstrong presuppose that occurrent mental or psychological states fall roughly into three categories: those whose subjects are aware of being in them; those whose subjects are not aware of being in them, but could have been had they taken notice; and those, such as language-processing states, which are entirely subterranean and inaccessible to introspection. A theory is needed to explain these differences.

"Aware *of*" suggests intentionality, viz., that a conscious state is itself the intentional object of another state, the awareness. If intentionality is representation, that would mean that the state is represented by the subject, perhaps by another state of the subject's brain.

There are two main representational theories of conscious awareness and conscious states in the present sense. There is Rosenthal's "higher-order thought" theory, according to which what makes a mental state a conscious state is simply that the subject is having a thought about it, provided that the thought was directly caused by the state itself. There is also the Lockean "inner sense" or "higher-order perception" theory offered by Armstrong. On Armstrong's view, the representing is done quasi-perceptually, by a set of functionally specified internal attention mechanisms of some kind that scan or monitor first-order mental/brain states.

Each of those views easily explains the differences between our three categories of first-order state; a state is, or is not, or could not be a conscious state accordingly as it itself is, or is not, or psychofunctionally could not be the object of a higher-order quasi-perception or thought. But let's concentrate on the Armstrong theory.

Further motivation for a representational theory of awareness is obvious enough: When we deliberately introspect and thereby become aware of a first-order mental state that we had not realized we were in, the awareness is quasi-perceptual or at least takes the form of a mental state of some kind itself directed upon the first-order state; it feels as though we are "looking at" a particular sector of our cognitive or phenomenal field. And as we saw in class, this kind of introspective attending is under voluntary control; we can "look" around various sectors of our phenomenal fields at will.

Caveat (not an objection): This "inner sense" theory explains very little. In particular, it does not explain why sensations have the qualitative characters they do, or what it is like to have an experience of this or that sort. There is lots more to subjectivity.

Problems

Objection 1: Some philosophers have alleged a regress. If the second-order representation is to confer consciousness on the original, first-order state, it must itself be a conscious state; so there must be a third-order representation of it, and so on forever. Reply: "Inner sense" theorists reject the opening conditional premise. The second-order representation need not itself be a conscious state. (Of course, it *may* be a conscious

state, if there does happen to be a higher-order representation of it in turn.) It does not *confer* consciousness on the original state in the sense of bequeathing it. The original state is a conscious state just because you quasi-perceive it.

Objection 2: Some philosophers have complained that representational theories leave introspective beliefs too fallible and underrate the privileged access we have to our own mental states. An internal monitor is a *mechanism*, and every mechanism is fallible and works only contingently. But the objectors contend that our awareness of our own mental states is either infallible or, if not flatly infallible, constrained against unreliability. Sydney Shoemaker, for example, grants that pain can "occasionally" escape awareness, but he insists that that could not happen "as a matter of course; it may be true in Lake Wobegon that all of the children are above average, but it can't be true everywhere." Reply: The Wobegon analogy fails; nothing about introspection affords any arithmetical calculation such as that of an average. Rejoinder (Wittgenstein): But there are concepts which impose rules of the form, "X can and occasionally does happen without Y, but this couldn't be common or normal or without some special explanation"; to grasp the very concept is to know that any X without Y is a necessarily rare aberration. To have the very concept of pain is in part to know that although in special circumstances there can be pain-behavior without pain, pain-behavior is overwhelmingly good evidence of actual pain. So too with introspection. It can and occasionally does go wrong, but that has to be an aberration and it requires a special explanation; it couldn't happen as a matter of course. But (again) the "inner sense" model predicts that if you have a broken introspector, introspection *could* go wrong all the time, even systematically. If our introspectors were all to break, introspection would be completely unreliable—which idea makes no sense.

Objection 3 (Karen Neander): Like any mechanism that is a monitoring-and-reporting device, the internal scanner can be expected to issue the occasional *false positive*. Even though you are not in any pain at all, your introspector could start yelling, "Yow, terrible pain in your left elbow!!!" But what would that be like?? (a) It would be exactly like *feeling pain*; yet supposedly there is no pain at all. -But that's as near as matters to a contradiction.. (b) It would be like hearing a voice saying "Yow, terrible pain…" even though you felt no pain. –But that wouldn't really be introspection, but only having the experience of hearing a voice. (c) You would be in a weird dissociative state. –What kind of dissociative state? The alleged possibility has no clear description. Which suggests that it's *not* a possibility, contra the "inner sense" theory.

Objection 4: "Inner sense" meets methodological solipsism. Consider beliefs and desires. They themselves are in the head, but if Fodor-Putnam-Stich externalism is true, their contents are not. An internal monitor can scan what is in the brain, but it cannot scan causal-historical-or-whatever processes extending outside the head and connecting one's brain states to egg foo yung. Yet you know introspectively that what you desire is some egg foo yung. Or take one of your beliefs about water. Introspection tells you effortlessly and reliably that that belief is (indeed) about water—water, not XYZ. But an internal monitor can't know whether you're on Earth and in contact with $H_2O$ rather than

on Twin Earth and facing XYZ. Thus, introspection cannot be, or cannot simply be, the operation of an internal monitor.

Objection 5 (Georges Rey): Internal monitoring per se is cheap. Every laptop does it. Any halfway competent computer has proof-checkers and fail-safes of various kinds that monitor its own first-order computational states. So every laptop is conscious? Reply: It's only *mental* states that are made conscious by being monitored. So every laptop has mental states? Rejoinder: All right. But if we just assume that some creature or device does have first-order mental states but no conscious ones, could we make those states conscious just by adding a simple monitor and turning it on? As before, monitoring is cheap; buy a little one at the hardware store and you're in business.

Objection 6 (Wesley Sauret): Armstrong and Lycan conflate internal monitoring with attention, and Lycan has taken the position that monitoring *just is* attending in the ordinary sense, that happens to be directed inward. But no current neuroscientific theory regards attending as higher-order monitoring; rather, being attended is just a strengthening or enhancement of the first-order mental state itself. So it is gratuitous to posit the internal monitor. Reply: (Gulp), right, I concede. Though doubtless *there are* internal monitors of various kinds in the brain, they're not needed to explain state consciousness.

Shoemaker vs. inverted spectrum

Block (and Fodor, as cited by Shoemaker) had argued that an inverted spectrum argument like the one that figured in the parallel objection to Behaviorism also refutes Functionalism (of any sort): Just as Behaviorism entails that creatures which are behaviorally alike must be mentally alike, so does Functionalism entail that creatures which are functionally alike must be mentally alike. Yet, just as we can imagine spectrum inversion with respect to all actual and hypothetical behavior, we can imagine spectrum inversion w.r.t. actual and hypothetical behavior *plus internal functional organization*; so Functionalism is false, since the qualitative character of visual states can come apart from functional states.

I said "just as," but that's not quite right. In the case of Behaviorism, the imagining is backed up by a reason: We think that subjective colors (visual qualia) can invert w.r.t. behavior because we know that internal wiring could be inverted without a change in behavioral dispositions, the dispositions being due to uniform training from birth in the use of color words etc. That reason does not apply in the case of Functionalism; an inversion of internal wiring would be a functional difference. Notice also that Block's inversion hypothesis is more ambitious than the antiBehaviorist inversion hypothesis, since the qualia have to invert, not just w.r.t. behavior, but w.r.t. behavior plus all functional states.

Block argues:

1. Visual qualia can invert w.r.t. functional states.

2. If Functionalism is true, visual qualia cannot invert w.r.t. functional states (since Functionalism says that qualitative and other mental state just are functional states).

∴ 3. Functionalism is not true.

There is a standard materialist reply to this (it's akin to the one sketched by Shoemaker in fn 7, but a bit simpler). Remember the Identity Theorists' model of empirical or "a posteriori" identities (water = H2O, lightning = electrical discharge, genes = segments of DNA molecules, gold = element whose atomic number is 79): In every such case, of course we can imagine that the identity claim is false. After all, the identity had to be empirically discovered in the first place. But that does not show that the identity claim is not true, or that the original property really could come apart from the thing with which it has been scientifically identified. (We can imagine "inverted fluids": Water is really $H_2SO_4$, while sulfuric acid is $H_2O$. But of course this is not genuinely possible. What Lavoisier discovered is that water simply is $H_2O$.) Thus, imaginability does not entail genuine possibility. Some of the things we can imagine are not really possible (cf. "5:00 p.m. on the sun," also Escher drawings).

So, Block's premise 1 is ambiguous. (i) Does it mean just that we can imagine visual qualia inverted w.r.t. functional states? Then Block has given no argument to get us from that imaginability to real possibility. We can imagine water not being $H_2O$, lightning not being electrical discharge, etc., but that does nothing to show that water isn't $H_2O$ or that lightning isn't electrical discharge. But (ii) if premise 1 means that visual qualia really can invert w.r.t. functional states, it begs the question, or at least fails to advance the discussion; obviously the Functionalist is not going to grant the premise on that interpretation. (Of course Block may be right and the premise may be true, but he can't just assert it; "I'm right and you're wrong" isn't an argument.)

Notice again the difference between Functionalism and Behaviorism here. The inverted-spectrum argument against Behaviorism worked because there is at least a little argument, the crossed-wiring argument, to get us over the gap between imaginability and real possibility. That's what's missing in regard to Functionalism.

Shoemaker goes on to give what he thinks is a better response to Block's inverted-spectrum objection. (We didn't discuss it in class, because it's tedious.) He concedes, at least for the sake of argument, that inverted spectrum w.r.t. functional states is possible, but he argues, by way of the intrapersonal case (p. 401), that it would be behaviorally detectable. He then assumes (p. 402) that when qualitative similarity or qualitative difference holds between "co-conscious" experiences, that will give rise to corresponding beliefs about similarities or differences in the world. That allows him to say that qualitative similarity and qualitative difference are functionally definable, precisely in terms of producing those beliefs. He concludes that *classes* of qualitative states can be functionally defined. But he is forced to concede that particular qualitative

states cannot be functionally defined—because such particular states can (he continues to assume) invert within the functionally defined classes. So he ends up with a compromise position: It is functionally determined that a particular state is a qualitative state, and that it is a member of a certain similarity class, but what particular quale the state has is not determined.

(Elsewhere Shoemaker suggests that type-identity holds of particular qualia. He calls this view "selective parochialism": Multiple realizability fails for particular qualia, even though it holds for all other mental properties.)

Shoemaker on absent qualia

Of course Block claims that inverted qualia are not the worst of it. Not only can we imagine creatures which are functionally alike but have their qualia inverted as between each other, but we can imagine a third creature that is functionally identical to the first two but which has no qualia at all. (It is cold and dead inside; the lights are off; there's nobody home. It is a **zombie**.) That was the point of Block's homunculi-head, Chinese giant etc. examples.

The standard Functionalist reply works against this "absent qualia" objection too. But Shoemaker now gives a further positive argument that such absent qualia are not possible. He supposes toward reductio that they are, and deduces an absurdity:

1. Organism $O_1$ is in functional state $S_1$ and $S_1$ is accompanied by qualitative property Q, while organism $O_2$ is also in $S_1$ (and is otherwise functionally equivalent to $O_1$) but does not experience Q. [A schematic "absent qualia" scenario.]

2. If $O_1$ is in $S_1$-with-Q, $O_1$ can and normally does know that s/he is in $S_1$-with-Q (and indeed is aware of being so).

3. If X knows and is aware of the fact that P, then the fact that P was a cause of X's belief and awareness.

$\therefore$ 4. If $O_1$ is in $S_1$-with-Q, then that fact normally causes $O_1$ to believe and to be aware of it. [2,3]

$\therefore$ 5. $O_1$'s now being in $S_1$-with-Q causes $O_1$ to believe that s/he is in $S_1$-with-Q (indeed, to be aware of being so). [1,4]

6. For mental state $M_1$ to be a normal cause of mental state $M_2$ is a functional relation.

$\therefore$ 7. $O_2$'s being in $S_1$ causes $O_2$ to believe that s/he is in $S_1$-with-Q (indeed, to be aware of being so). [1,5,6]

∴ 8. $O_2$ does not experience Q, but believes s/he is in $S_1$-with-Q (indeed, is aware of being so).

--Which is absurd, so we must reject the "absent qualia" hypothesis 1.

Objections. Premise 2 can be denied; then the causal requirement on knowing will not come into play. And perhaps the outcome 8 is not as bad as Shoemaker supposes.

Shoemaker also raises a skeptical question: The argument shows that even if some creature Z were a zombie, if Z were functionally equivalent to you Z would believe that Z experienced qualia. So what reason have you for thinking that *you* experience qualia? But this too is absurd. (This assumes that we're talking about a zombie in the sense stipulated in class, one which has propositional attitudes but no sensory experience or qualia, not a mega-zombie that has no mentality at all.)

Nagel, Jackson, and "what it's like"

Farrell's/Nagel's bat example is supposed to show that phenomenal character is intrinsically subjective, in a way that makes it in principle inaccessible to science. The whole business of science is to remove all subjectivity, to display facts as they are in themselves regardless of anyone's particular perspective on them; but to do that to phenomenal character would be to miss the entire point of phenomenal character, to change the subject.

Nagel's claims are that (a) there is something to know about the bat, that can be known only by taking the bat's internal perspective, and that (b) that something seems to be a fact of a special kind.

Jackson's version of the argument:

(1) Before her cure, Mary knows all the scientific and other "objective" facts there are to know about color and color vision and color experience, and every other relevant fact. [Stipulation.]

(2) Upon being cured, Mary learns something, viz., she learns what it's like (w.i.l.) to experience visual redness. [Seems obvious.]

∴ (3) There is a fact, the fact of w.i.l. to experience visual redness, that Mary knows after her cure but did not know prior to it. [1,2]

(4) For any facts: if $F_1 = F_2$, then anyone who knows $F_1$ knows $F_2$. [Suppressed; assumes simple factive grammar of "know."]

∴ (5) There is a fact, that of w.i.l., that is distinct from every relevant scientific/"objective" fact. [1,3,4]

(6) If materialism is true, then every fact about color experience is identical with some physiological, functional, or otherwise scientific/"objective" fact.

---

∴ (7) Materialism is not true. [5,6]

4 is supplied because without 4, there seems no way to get 5 from 1 and 3.

Materialist responses

There are three main materialist responses to the Knowledge Argument. One is brutally to deny 2. 2 can be denied in either of two ways. First, Dennett in some moods simply rejects the idea of "w.i.l." to have a sensation over and above the sensation itself; the notion is empty. Second, even if we grant that there is such a thing, Dennett and Kathleen Akins insist that if Mary really did know *all* the scientific/"objective" facts, she could work out w.i.l. to see red, and so would know it after all. At least, Jackson has given us no argument for thinking she could not. (Dennett reminds us that fantastical science-fiction examples of this kind are dangerous because we are taken in by the immediate image and fail to see the real implications of the fantastical hypothesis.)

Jacksonian rejoinder: Yeah, OK, maybe. But it sure does seem that Mary would not know what it's like to see colors without having experienced color in some way. Also, how could she "work out" w.i.l.? --Not by deducing it from her body of scientific lore.

A second rebuttal of Jackson is to grant 2 but balk at 3, holding that Mary's acquisition is, not a fact, but a mere ability, a knowing-*how* (Lewis and Nemirow) or a mere acquaintance (Conee). The Ability theory has it that although Mary would acquire some knowledge, it would be only a skill or ability, not knowledge *that* anything, not knowledge of a fact. Jackson has done nothing to show that she has learned more than an ability to imagine colors, an ability to sort objects by color without using her spectrometer, etc.

Jacksonian rejoinder: Knowing w.i.l. is a matter of having a true belief. If I tell you that tasting Vegemite is like eating very salty shoe polish, you can get hold of some Vegemite and verify or refute my claim.

In the same vein, imagining is correct or incorrect. (If I am imagining my boyhood home as viewed from the street, I may get it right or I may be in error.) If Mary can imagine seeing a red object, this must mean imagining it correctly, not getting it

wrong by imagining seeing what is in fact a different color. And presumably her ability is to be reliably correct, not just accidentally so. The best explanation of that reliability is that she knows that that is how red objects look.

A second argument is linguistic: "Knowing wh-" locutions are true in virtue of the truth of "knowing that" locutions with referring terms in them. E.g., to know who robbed the diaper service is to know that N robbed the diaper service, for some suitable name N; to know when the bar closes is to know that the bar closes at t, where t refers to a time. So too, presumably, to know what it's like to see red is to know that it is like X to see red, for some suitable term X. Of course, it's hard, maybe impossible, to describe what it's like in English (except comparatively), but a demonstrative is natural here: If you're like me, you'll want to say, "It's like... THIS. I know what it's like; I just can't put it into words."

The third response is now fairly standard: the "perspectivalist" response. It begins by rejecting premise 4, the suppressed principle according to which (to put it another way) if someone knows that P but does not know that Q, then the fact that P and the fact that Q are different facts.

That principle may at first seem obvious. It seems to be licensed by Leibniz' Law. You'd think that if fact $F_1$ is known to Smith, and $F_1 = F_2$, then surely $F_2$ is known to Smith. But there are clear counterexamples to it: The fact of salt spilling just is the moving of lots of NaCl molecules, but someone can know that salt is spilling without knowing anything about NaCl; the fact of my being overpaid just is the fact of WGL's being overpaid, but someone (who does not know that I am WGL) can know that WGL is overpaid while having no idea whether I am overpaid.

What has gone wrong? As always and notoriously, Leibniz' Law fails for *representation-dependent* properties. (For more on this, see again the freebie handout "Descartes' Argument and Leibniz' Law.") That Oedipus wanted to marry Jocasta but did not want to marry his mother does not show that Jocasta was not his mother; the poor woman was his marriage-object under one description or mode of presentation but not under the other. And *being known to Smith* is a representation-dependent property: Whether Smith knows a given fact depends on how Smith represents that fact. She may know it under one representation but not know it under a different one. That is just what is going on in the "salt" and "overpaid" examples. One may see salt spilling but lack the chemical perspective entirely; less commonly, a mad chemist might record a motion of NaCl molecules but be mentally so far removed from the perspective of everyday things and substances that she has no thought of salt.

The "overpaid" example is perspectival too, but a different kind of perspective is at work. Someone can know that WGL is overpaid without knowing that I am overpaid, if that person has only a public, (non-auto-)biographical perspective on me and is not in a position to refer to me more directly. Even if the person were to come into the room, point straight at me and exclaim "*You* are overpaid," I might insist that the knowledge she thereby expresses is still not quite the same knowledge I have when I know that *I myself* am overpaid. (Especially if I believe that she is mistaken as to who I am.) As Hector Castañeda emphasized in the 1960s, if I am amnesic I may know many facts about WGL, including that he is overpaid, without knowing that I myself am overpaid; so it may seem that what I know when I do know that I myself am overpaid is a different fact from any

of those I could know while amnesic, and an intrinsically perspectival fact. In order to designate the person it is supposed to designate, a mental pronoun can be tokened only from a certain point of view; only I, WGL, can think "I" and thereby designate WGL. But in the ordinary chunky way of individuating facts, the relevant one is just that the person in question, however represented, has the property of being overpaid.

Clearly, *being known to Mary* is a representation-dependent property; whether Mary knows a given fact depends on how she represents that fact. Facts can be differently represented from differing perspectives, and that is why 4 is false. Without 4, seemingly, the Knowledge Argument collapses.

Jacksonian rejoinder: All right, so 5 doesn't strictly follow from 3. But the case of knowing w.i.l. is not like the examples that refute 4. In those examples, we have multiple representations of the same fact, based on distinct epistemic routes to or takes on that fact. But knowing w.i.l. to experience visual redness is not a matter of representation; we don't have to *represent* w.i.l.; we know it just by having the experience itself.

Also, isn't 5 obviously true anyway? When I know w.i.l. to experience visual redness, it seems I'm knowing an *entirely* different fact, indeed a different kind of fact, from any fact about neurological activity or even functions being performed.