Phil 100D: Philosophy of Mind Professor Aaron Zimmerman

Handout #7b: Lewis' Response to the Knowledge Argument

1. An Objection to Jackson's Inference from Epistemic Premises to a Metaphysical Conclusion

Definition 1: A subject S individuates facts *finely* iff S allows that the fact that x is F can be distinct from the fact that y is G even if x=y and F=G.

Definition 2: S individuates facts broadly iff S does not individuate facts finely.

Let P=the fact that Samuel Clemens is happy.

Let Q=the fact that Mark Twain is happy.

Suppose that John (as described in previous handouts) sees Samuel Clemens giddy in the bar, and reports that Samuel Clemens is happy, but when asked whether Mark Twain is happy, sincerely answers, "I have no idea. Why are you asking me?"

Claim 1: If we individuate facts finely, then we will say that P≠Q because (a) John knows P but (b) does not know Q. But (c) we cannot infer from this that the two facts involve different objects or properties as both involve one guy (i.e. Samuel Clemens, a.k.a. Mark Twain) and one property (i.e. happiness).

Claim 2: If we individuate facts broadly, then we will say that P=Q even though (a) John knows P, (b) John claims not to know Q, and (c) John feels as though he learns something when he is told "Mark Twain is Samuel Clemens," and (d) John therein seems to "realize" that Mark Twain is Samuel Clemens (and that P=Q).

Let R= *the fact that the experience of seeing something red is like this [as said while seeing something red under normal lighting with normal color vision]*, or the fact that the experience of seeing something red [when in normal lighting with normal color vision] instantiates a red* *quale* [where a red* quale is the property of those visual experiences of red things that makes them all similar in regard to what it is like to have them; e.g. the visual experience of a red chair and the experience of a reddish afterimage are supposed to be alike in that each of them is red*].

Let N= the fact that the experience of seeing something red is F1-Fn [where F1-Fn are the neurophysiological properties that distinguish experiences of red things from other kinds of experience in their biological or physical aspects]

Result: (A) If she individuates facts finely, the physicalist can admit that R is distinct from N, but insist that we cannot infer from this that *being an F1-Fn experience* is a distinct property from *being this kind of experience* [i.e. being an experience that instantiates a red* quale]. (B) If she individuates facts broadly, the physicalist will insist that we cannot infer that R≠N from the mere fact that (a) Mary knows N, and (b) Mary will not (indeed cannot) assert the sentence we've used to denote R, and (c) Mary claims to come to a new realization upon leaving her room. The physicalist can defend strategy (B) by pointing out that on this broad conception of facts we cannot infer that Samuel Clemens≠Mark Twain from the fact that (a'') John knows that Samuel Clemens is happy and (b'') John will not assert "Mark Twain is happy," and (c'') John claims to come to a new realization upon being told that Mark Twain is Samuel Clemens.

2. Taking Jackson Seriously

Lewis individuates facts broadly in "What Experiences Teaches." According to the kind of view Lewis adopts there, the fact P=the fact Q so long as there is no possible world where they differ in truth value (i.e. no possible world in which P is true and Q is not and no possible world where Q is true and P is not). So the fact that 2+2=4 is the same fact as the fact that the area of a circle is always pi times the square of its radius. Since these propositions are true in all worlds, they are true in precisely the same set of worlds—so we really have one fact here (i.e. *the* necessarily true proposition) expressed by two radically different linguistic or symbolic representations.

Less bizarrely, Lewis would say that the fact that Samuel Clemens was brilliant is not distinct from the fact that Mark Twain was brilliant even though John asserts that Mark Twain was brilliant but he's unwilling to assert that Samuel Clemens was brilliant because he doesn't realize that "Mark Twain" was Samuel Clemens' pen name.

But Lewis does not think we can easily dismiss the claim that Mary learns a new fact of this sort when she leaves the black and white room—i.e. a new broadly individuated fact—by comparing Mary's realization (the realization she would express upon leaving the room that "This is what it's like to see red!") to John's realization that Samuel Clemens is Mark Twain. In Lewis' view John knew all along that Samuel Clemens is Mark Twain as this is really nothing more nor less than knowledge of the broad fact that Samuel Clemens is Samuel Clemens. (Actually, Lewis would deny this, as he seems to be a descriptivist about proper names, but this is outside the point here.) John just comes to know this fact under a new linguistic representation of it provided by the sentence 'Samuel Clemens is Mark Twain' whereas before he knew the same fact under a different linguistic representation (i.e. the sentence) 'Samuel Clemens is Samuel Clemens'. But, according to Lewis, when Mary leaves her room she *doesn't* just come to realize that the brain states people are in when they have visual experiences of red things are properly characterized in some particular way or that there is some linguistic representation that applies to these experiences that she didn't realize applied to them when she was in her black and white room.

Two sets of questions: (1) Is John's realization that Samuel Clemens is Mark Twain merely linguistic? Is it equivalent to learning that 'Samuel Clemens' and 'Mark Twain' name the same guy? (2) Is Mary's realization upon leaving the black and white room, i.e. the realization she expresses with "So this is what it is like to see red!" clearly *more substantive* than the realization John would express as "So Mark Twain is Samuel Clemens!"?

3. The Argument from Mary's Ignorance to Epiphenomenalism

Suppose Sarah is like Mary, but instead of learning everything about color vision without ever having seen colors, Sarah has learned all about olfaction and taste perception but has never tasted anything (she's been fed intravenously). For the sake of simplicity, we can suppose with Lewis that there are just two ways it might be like for us when we taste vegemite at t: V1 and V2; and that there are just two possible resultant physical states of the universe at the next moment t+1: P1 and P2.

Let V1=the set of possible worlds in which the experience of tasting vegemite at t is characterized by quale Q1≠Q2.

Let V2= the set of possible worlds in which the experience of tasting vegemite at t is characterized by quale Q2≠Q1.

Let P1=the set of possible worlds in which the microphysical structure of the universe at t+1 is X≠Y.

Let P2= the set of possible worlds in which the microphysical structure of the universe at t+1 is Y≠X.

Lewis tells us that there are two hypotheses according to which the physical state of the world at t+1 depends on the qualitative state of our experience at t, and two hypotheses according to which the physical state of the world does not depend on the qualitative nature of our experience.

Dependence
K1: If V1 then P1; if V2 then P2.
K2: If V1 then P2, if V2 then P1.

Independence
K3: If V1 then P1; if V2 then P1.
K4: If V1 then P2; if V2 then P2.

This yields eight possibilities regarding (a) the qualitative nature of our experience of vegemite at t; (b) the state of the world at t+1; and (c) the relations of dependence holding between the two.

K1V1P1	K3V1P1	K3V2P1	K2V2P1
K2V1P2	K4V1P2	K4V2P2	K1V2P2

Lewis is trying to argue (with Jackson) that if Mary and Sarah are ignorant of some broadly individuated fact, then epiphenomenalism results. If the qualitative aspect of experience is not epiphenomenal, the difference between Q1's marking our experience at t and Q2's marking that experience must make some difference to the subsequent physical state of the universe. So we don't need to consider the possibilities that contain K3 or K4. Sarah's physics lessons in the smell and tasteless room can inform her whether P1 or P2 obtains at t+2. So, without loss of generality we can suppose that she knows that the actual world is (a member of) P1 at t+1. Thus, Sarah has ruled out all the possibilities compatible with K3, K4, and P2. So she need only decide between K1V1P1 and K2V2P1. But the difference between these two hypotheses is something about which Sarah has no evidence. She knows the universe is X at t+1 (i.e. P1), and she knows that its state is dependent on the phenomenal aspect of our experience of vegemite at t. But the only difference between these two is a swap between the qualia and the psychophysical laws linking qualia with physical events in the world—whichever one of these two combinations holds, the physical world will be exactly what it would be if the other one held. And this, says Lewis, is a form of epiphenomenalism.

"The physical effect is exactly the same whether it's part of the joint possibility K1V1P1 or part of its alternative K2V2P1. It may be caused by V1 in accordance with K1, or it may be caused by V2 in accordance with K2, but it's the same either way. So it does not occur because we have K1V1 rather than K2V2, or vice versa. The alleged difference between these two possibilities [would do] nothing to explain the alleged physical manifestations of [Sarah's] finding out which one of them is realized. It is in that way that the difference is epiphenomenal. That makes it queer and repugnant to good sense" (p. 291).

Question: If we embrace the dependence of physical states on qualia when the laws are held fixed and we only allow some independence of physical states from combinations of qualia and laws

(in that the same physical effects are compatible with distinct sets of qualia and laws) have we really embraced an objectionable form of epiphenomenalism? Isn't the physical state of the universe at any time also compatible with more than one combination of prior *physical* states and *physical* laws?

4. Framing Possibilities vs. Eliminating Possibilities

As we've seen, Lewis argues that phenomenal information would have to be epiphenomenal; irreducibly qualitative information as to what it is like to see red couldn't make a difference to our knowledge of which of several possible physical states of the world will obtain as actual were red to seem one way rather than another. But I find it difficult to apply Lewis' reasoning to the example of Mary to join him in this conclusion. And I think the main problem can be traced to Lewis' model of learning or the acquisition of knowledge.

In the argument that phenomenal information must be epiphenomenal, Lewis assumes that we are given the qualitative possibilities in advance and given the physical possibilities in advance along with possibilities about their interdependencies. Mary is supposed to know that experience of red is either V1 or V2 and wonder which. This allows her to consider hypotheses about what the subsequent physical state of the world will be and whether the physical state of the world will depend on which of the possible qualitative characters experience of red will actually have. I think Lewis is led to think this is a coherent description of the scenario because he thinks belief is a matter of excluding or ruling out various possibilities.

Lewis' model of learning: A thinker S is somehow "given" a space of possibilities or possible worlds PW: (w1, w2, w3,….). Label the worlds at which a given proposition P is true the p-worlds. Then we can say that according to Lewis, S comes to believe that P just in case S includes the p-worlds within PW and excludes the not-p-worlds from PW. On this understanding, to learn something is to narrow down or winnow the space of possibilities.

This might make sense if a space of possible worlds (PW) were given to us in virtue of our grasp of the language in which we speak and think.

Suppose I understand the words "green" and "blue" and I understand the words "grass" and "sky" and I can then entertain the various possibilities: (1) the grass and the sky are both green, (2) both are blue, (3) the sky is green and the grass is blue, or (4) the sky is blue and the grass is green. When I have a normal course of experience and make the observations I make as a child, I learn that possibility (4) is actual and that (1)-(3) are "mere" possibilities. I can now "exclude" (1)-(3) from reality. Learning amounts to narrowing down or winnowing the space of possibilities: the ways the world might be a priori.

But in the case Jackson describes, there is a sense in which Mary cannot even frame the possibility that experience of red is V1 or experience of red is V2. She lacks the kind of color experience she would need to even grasp propositions that contain names for the various qualia in question. The point here is that color experience seems to give Mary a whole new range of phenomenal or qualitative color concepts that she lacks when trapped in her black and white room. Intuitively, these are not possibilities she can frame merely by having "red" in her vocabulary and using expressions to denote the experiences in question: e.g. "the kind of experience typically caused by red things." This suggests that learning is not just a matter of winnowing the space of possibilities. Learning often requires creative thought and experience. Learning depends on the creation or construction of the concepts we use to frame the hypotheses

we can then test in the way Lewis imagines to determine which of the various possibilities we use these concepts to frame are realized in actuality.

Questions: Does introspective awareness of the qualitative character of our experiences enable us to form concepts and frame hypotheses we couldn't even consider were we limited to third-person descriptions or third-person representations of our experiences? Introspection is a source of knowledge insofar as you know what you are now thinking and feeling via introspection. But isn't introspection a source of concepts as well? Does introspection on our experiences of thinking, moving and feeling give us ways of thinking of our thoughts, movements and feelings that we could not grasp or understand without thinking those thoughts, performing those movements and experiencing those feelings (or qualitatively similar experiences)? Does a positive answer to these questions allow us to challenge the coherence of Lewis' argument that phenomenal information would have to be epiphenomenal?

5. Lewis' Ability Hypothesis

What then should we say about Mary? According to Lewis, what Mary lacks is not knowledge that one possibility obtains rather than another (i.e. that our visual experiences of red things instantiate quale red* rather than some other quale). Instead, knowing what it is like to see red things (or imagine or hallucinate red things) consists in the possession of various abilities: abilities to imagine and remember red things and abilities to categorize red things viz. their color. What Mary lacks is nothing more nor less than this.

"The ability hypothesis says that knowing what an experience is like just is the possession of these abilities to remember, imagine and recognize. It isn't the possession of any kind of information, ordinary or peculiar. It isn't knowing that certain possibilities are not actualized … Therefore it should be no surprise that lessons won't teach you what an experience is like. Lessons impart information; ability is something else" (p. 293).

Reply: "A friend of phenomenal information will agree, of course, that when we learn what an experience is like, we gain abilities to remember, imagine and recognize. But he will say that **it is because we gain phenomenal information that we gain the abilities**" (p. 293).

Questions: What is Lewis' reply to this objection? Is it adequate?