A priori physicalism: Dennett and Lewis

Jackson and Chalmers on Physicalism

Recall: Jackson and Chalmers believes that their certain epistemic claims about consciousness (the claim that Mary learns something new when she leaves her black and white room; the claim that zombies are conceivable) supports an important epistemic premise:

> The Non-Deducibility Claim (NDC): There are truths about phenomenal consciousness that cannot be a priori deduced from the totality of physical truths.

Whereas someone with the right concepts, physical information, and deductive capacities could a priori deduce (and thus know) all the facts about water on the basis of facts about H2O, anti-physicalist arguments purport to show that the phenomenal facts *cannot* be a priori deduced from complete physical information.

A priori vs. a posteriori physicalism

One physicalist response to the anti-physicalist arguments is to reject NDC – i.e., to assert that the phenomenal facts can be deduced a priori from the physical facts. This is the strategy of:

> A priori physicalism (a.k.a. 'Type-A materialism'): Supervenience physicalism (as we have understood it) plus the claim that all truths can in principle be a priori deduced from the physical truths.

- agrees with anti-physicalists like Jackson and Chalmers about how to *interpret* physicalism, but:
- disagrees with anti-physicalists about the truth of NDC. Hence, the physical-phenomenal conditional, P → Q, is not only necessarily true but knowable a priori for anyone with the requisite concepts.
- entails that all concepts (including phenomenal ones) are analyzable in terms of structure and function. [Often goes with a neo-Frege-Russell view.]
- Proponents: Daniel Dennett; David Lewis; Keith Frankish (?)

A priori vs. a posteriori physicalism

The main alternative:

> A posteriori physicalism (a.k.a., 'Type-B materialism'): Minimal physicalism plus the denial of the claim that all truths can (in principle) be a priori deduced from the physical truths.

- disagrees with anti-physicalists like Jackson and Chalmers about how to *interpret* physicalism, but:
- agrees with anti-physicalists about the truth value of NDC. The physical-phenomenal conditional, $P \rightarrow Q$, is necessarily true but not knowable a priori for anyone with the requisite concepts. (That is, the physical-phenomenal conditional $P \rightarrow Q$ is necessarily true but only a posteriori.)
- typically goes along with the denial that all concepts are analyzable in terms of structure and function. (At the heart of the so-called "Phenomenal Concepts Strategy." [Usually goes with a rejection of neo-Frege-Russell views.])
- Proponents: Brian Loar, Ned Block, David Papineau, Janet Levin (and many others)

A priori physicalism

Most contemporary physicalists are a posteriori physicalists: As noted, they are happy to concede that Mary learns something new and that the NDC is true. This is how they rebut Jackson's argument: they argue that physicalism is perfectly consistent with NDC. But there are some a priori physicalists, and they *do* deny the NDC.

- Daniel Dennett is a case in point. He denies not only the NDC but also the intuition that Mary learns something new. Specifically, he thinks we have the intuition that Mary really would learn something only because we haven't thought hard enough about the Mary thought experiment. Thus, he rebuts Jackson by denying not only the NDC but even that Mary learns anything new when she leaves the room.

- Dennett believes that if Mary *really did* know everything Jackson says she does – and he has doubts about whether that supposition is coherent or conceivable – then we can't rule out that she might work out a way of already knowing what it's like to see red.

A priori physicalism

- Dennett's article "What RoboMary Knows" is an extended defense of this idea. In the first main section, he tries to neutralize reasons people often give for thinking that Mary really will learn something new based on the alleged "richness" of her new experience of (say) red. In the second main section, he tries to neutralize such reasons based on the idea that you can know what it's like to have a type of experience only if you've had that type of experience.

Richness and new knowledge of what it's like

In the section called "She'll Be Surprised, Dammit!" Dennett has a lot of fun at the expense of the claim that the alleged "richness" of experience shows somehow that Mary will learn something new. The argument Dennett attacks seems to run like this

(1) Mary could already know what it's like to see red prior to seeing it only if the content of that knowledge is linguistically expressible (ideally as "a string of demonstrative-free sentences" (Lycan)); but

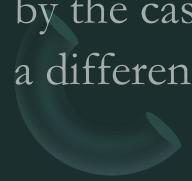
(2) due to the richness of color experience that is not possible.

Richness and new knowledge of what it's like

Dennett's response is to deny that richness entails inexpressibility:

"[A]re we really so sure that what it is like to see red or blue can't be conveyed to one who has never seen colors in a few million or billion words? […] Remember, Mary knows everything about color that can be learned by physical science, and she presumably has the attention span and powers of comprehension required to handle 10 billion words on what it is like to see red as easily as she does twenty-five words or less on triangles."

Dennett thinks Jackson's Mary thought experiment would have commanded much less agreement if, instead of an omniscient colour vision scientist who has never experienced red, it was about an omniscient geometer who has never had an experience of triangle. (Given all the knowledge and resources available to this omniscient geometer, surely they would be able to work out what experiences of triangles are like. It just doesn't seem that hard. So why should we be any more moved by the case of something (admittedly) much richer and more complex: namely, an experience of redness? That just looks like a difference of degree of informational complexity, not of kind.



Richness and new knowledge of what it's like

A worry about Dennett's argument:

- Richness isn't obviously what's at issue. If anything, it's the *simplicity* of the experience that is to the point. Imagine your visual field flooded by a single wash of red. This strikes us as a mere simple phenomenal quality.

Experience and new knowledge of what it's like

In the section called "You Had to Be There" and the following sections, Dennett takes on another reason for thinking that Mary must learn something new.

Consider the following claim:

> The Experience Requirement (ER): One can know what it is like to have a color experience only if one has had or is having color experience.

Dennett thinks that the NDC is true only if the ER is true. Thus, he reasons that if the ER is false, then the NDC is false too and the knowledge argument is neutralized. As the philosopher Torin Alter has summarized Dennett's reasoning:

Experience and new knowledge of what it's like

"[I]f the experience requirement fails – if it is possible to know what it's like to see in color without having color experiences – then why couldn't Mary put herself in a state that allows her to figure out what it's like to see in color? If there is no logical bar to obtaining this phenomenal knowledge without seeing colors, then there is no reason why Mary could not obtain that knowledge by exploiting her comprehensive physical knowledge" (from Alter's "Phenomenal Knowledge without Experience").

Dennett is going to try to show that there is no logical barrier to someone acquiring knowledge of what it is like to experience red without ever having had an experience of red.

Warmup: Swamp Mary

Dennett warms us up for his main thought experiment with "Swamp Mary and her Cosmic Accident."

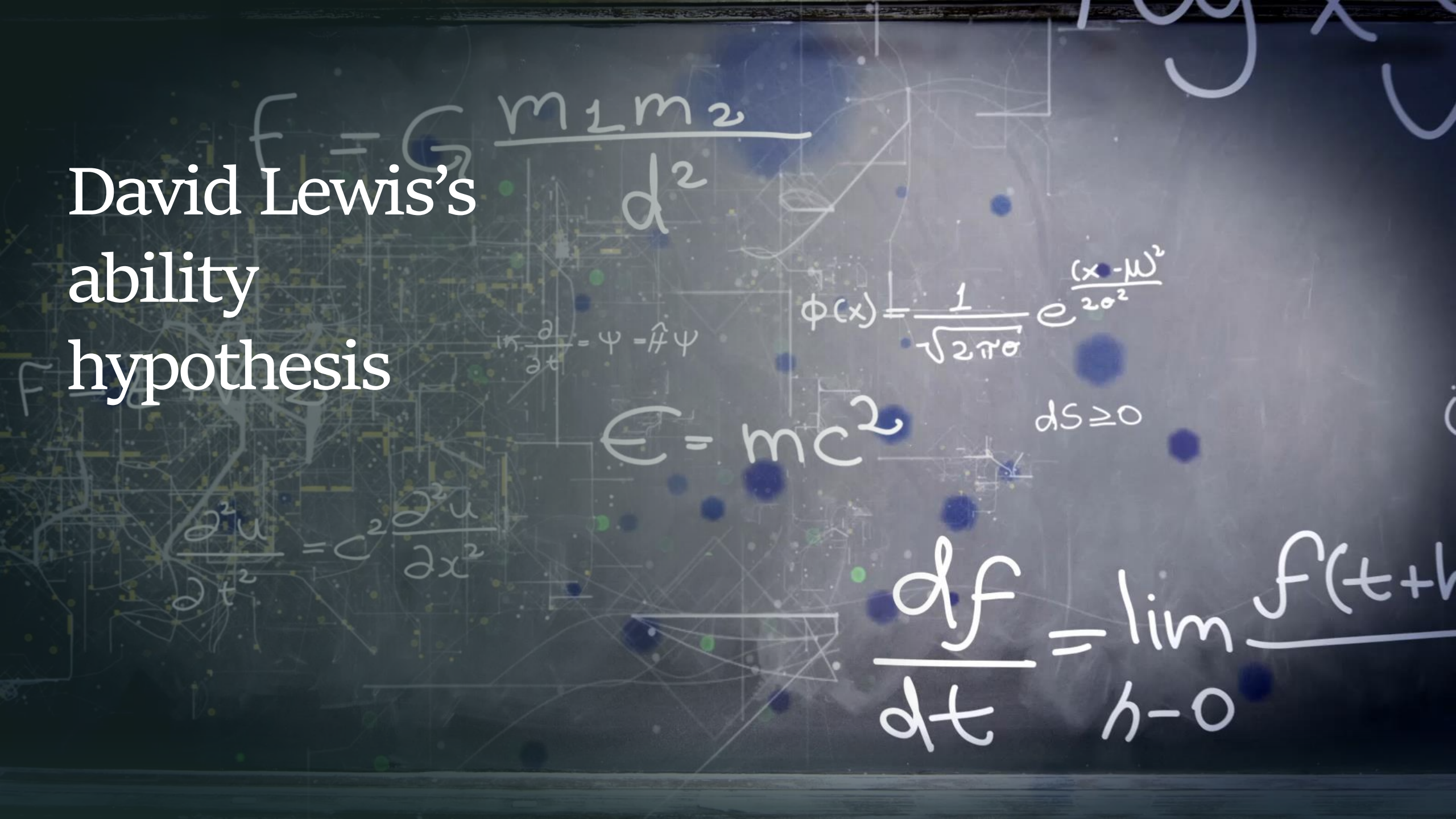
- By a 'Cosmic Coincidence', Swamp Mary pops into existence as a result of a bolt of lightning. The psychological state she is in when she pops into existence is the very psychology state Mary would go into *after* having her first experience of red.
- Note: what is induced in Swamp Mary isn't a hallucination as of red but, as Dennett insists, "the dispositional state, the competence state, that an experience of a red rose *would have* put her brain into *had* such an experience (hallucinatory or not) occurred."

RoboMary

- First consider "Unlocked RoboMary", who uses her vast physical knowledge to induce imaginative color experiences (and thereby know what it's like to see red).
    - Dennett recognizes that this will be regarded as cheating.
- Next, consider "Locked RoboMary." Locked RoboMary uses her vast physical knowledge to induce in herself the dispositional state that the imaginative color experiences cause in Unlocked RoboMary (or that the cosmic accident causes in Swamp Mary). Dennett thinks that this shows that she has *genuine knowledge* of what it's like to see red even without ever once having had anything like an experience (or imagining) of red.

Objection to Locked RoboMary

- Objection: True, RoboMary knows what it's like to see red. But her knowledge was not a priori deduced from her physical knowledge. But it is the latter – a priori deduction – that, according to Jackson, Mary would have to be able to do if physicalism is true.
- o Dennett has a response to this objection at the very end of his paper.


David Lewis's
ability
hypothesis

Lewis's ability hypothesis

Lewis is an a priori physicalist. He thinks that our concepts of phenomenally conscious states can be analyzed *entirely* in terms of structure and function. Thus, he believes, if Mary knows everything physical, then she ought to be able to a priori deduce from that vast knowledge all the truths about structure and function and, hence, all of the truths about phenomenal consciousness.

Unlike Dennett, Lewis feels the force of the intuition that Mary gains knowledge when she is released from her black-and-white prison. He claims that Mary does gain knowledge: knowledge of what it is like to see (e.g.) red.

- But! Lewis denies that this new knowledge of what it's like is *factual* knowledge, knowledge that something is the case. He analyzes it instead as the acquisition of new *abilities*. From his "Postscript to 'Mad Pain and Martian Pain'": "knowing what it's like is the possession of abilities: abilities to recognize, abilities to imagine, abilities to predict one's behavior by means of imaginative experiments" (131, cf. Nemirow).

Lewis's ability hypothesis

This "ability hypothesis" Lewis contrasts with Jackson's favored property dualistic "hypothesis of phenomenal information." Lewis favors his hypothesis over Jackson's because:

1. he thinks that the hypothesis of phenomenal information is, in an important and objectionable sense, inconceivable, and

2. the hypothesis of phenomenal information and the property dualism it supports leads ultimately to either epiphenomenalism or dualist interactionism (and he thinks both of those are trouble).

(See "What Experience Teaches" for more on the hypothesis of phenomenal information.)

Lewis's ability hypothesis

Even if you agree with Lewis that epiphenomenalism and dualist interactionism are problematic, are you so sure that the hypothesis is as hard to fathom as Lewis suspects? (True, words arguably can't convey the alleged phenomenal information that dualists say Mary gains access to when she first sees red. But does inexpressibility entail no information at all?

Lewis's ability hypothesis

It has become common to equate Lewis's claim with the claim that Mary gains "know-how," not "know-that," when she leaves her black-and-white room.

- Gilbert Ryle had argued in his book *The Concept of Mind* that "knowledge-how" isn't analyzable in terms of "knowledge-that." For example, knowing how to hit a fastball isn't just a matter of knowing various propositions (e.g. that choking up on the bat handle increases swing speed).
- The relations between knowledge-how and knowledge-that are complex and subtle. Some deny Ryle's claims.

Some objections …

Here are some (but only some!) of the objections that have been raised against Lewis's idea:

Objection 1: "The ability hypothesis is just flat-out implausible!": Could it be true that all Mary gains is "know-how"? It certainly seems as if she gains propositional knowledge. She might say after leaving the room: "Wow! I didn't know that seeing red is like this, but I now I do."

And one might insist that Mary gains new know-how precisely in virtue of acquiring new propositional knowledge.

Some objections …

Objection 2: The Argument from Meaning and Syntax (due to William Lycan): Much work on indirect questions in theoretical linguistics suggests that "S knows wh-..." clauses are systematically related to "S knows that..." clauses. Example:

- "S knows where X Vs" (e.g. "Sally knows where John lives") is true in virtue of S's knowing that X Vs at l, where "l" names a location.

So likewise:

- "S knows what it's like to see blue" is true in virtue of S's knowing that it is like Q to see blue, where "Q" names a phenomenal property of visual experience.

If this is correct, then attributions of knowledge of what it's like are plausibly taken to be attributions of factual knowledge, not (mere) attributions of certain abilities.

Some objections …

Objection 3: The Argument from Embedding (due to Brian Loar):

Some have argued that considerations to do with how "what it's like" constructions logically embed argue against the ability hypothesis.

- Some background (see the Loar reading next week for more):
- Emotivism (from metaethics)—Moral sentences are "non-factual" in that they lack truth values. (Assertions of moral sentences serve merely to express affective attitudes.)

Some objections …

Famous objection (due to Peter Geach): Moral sentences embed in ways that make emotivism doubtful. Consider:

Argument M

1) If lying is wrong, then getting little brother to lie is also wrong.

2) Lying is wrong.

3) Therefore, getting little brother to lie is also wrong.

Geach pointed out that Argument M looks like a perfectly valid piece of Modus Ponens-respecting reasoning. But it isn't if emotivism is true. (Two points: (i) truth values and (ii) availability of truth evaluable content for placement in non-assertoric positions.)

Some objections …

Now consider:

- Argument C

1) If pain feels like Q, then pain is more similar to anxiety than to elation.

2) Pain feels like Q.

3) Therefore, pain is more similar to anxiety than to elation.

This looks like a perfectly valid piece of Modus Ponens-respecting reasoning. But it isn't if the ability hypothesis is true.

Some objections …

Objection 4: Counterexamples (there are many in the literature):

If Lewis is right, then the following is true: S knows what it's like to see red if and only if S has the suite of abilities described by Lewis, in particular, the ability to remember and imagine what it's like to see red.

Counterexamples:

- Knowing what it's like without the abilities: Surely someone could know what it's like to see red, while, for example, eyeing a ripe tomato, without having these abilities because of some sort of brain damage.
- Having the abilities without knowing what it's like: Consider a subject who is adept at imagining shades she hasn't experienced that lie between shades she has experienced. This subject has the ability to imagine vermilion, a shade she hasn't experienced and which lies between scarlet and crimson, shades she has experienced. But unless she either *experiences* vermilion or *exercises* her ability to imagine it, she won't know what it's like to see vermilion.