

PHL342 Week 4 Notes

Prof: James John

Topic: Metaphysical Interlude: Souls, Behavior, and Brains: The Rise of Functionalism

Section Zero: Readings:

1. Jerry Fodor, "The Mind-Body Problem" (pp.114-122, stop at "An obvious objection...")
2. Janet Levin, "[Functionalism](#)" in [Stanford Encyclopedia of Philosophy](#) (secs. 1-3)
3. Michael Rescorla, "[The Computational Theory of Mind](#)" in SEP (sec. 1)

Section One: Video 2 (Turing Test 2):

1. Back to Lucas's Argument: A Gödelian Version of the Mathematical Objection
"Gödel's theorem must apply to cybernetical machines, because it is of the essence of being a machine, that it should be a concrete instantiation of a formal system. It follows that given any machine which is consistent and capable of doing simple arithmetic, there is a formula which it is incapable of producing as being true—i.e., the formula is unprovable-in-the-system—but which we can see to be true. It follows that no machine can be a complete or adequate model of the mind, that minds are essentially different from machines (1961, 113).
2. An Interpretation of Lucas's Reasoning
 - a. Given Gödel's results, any machine instantiating a formal system that is consistent and capable of doing arithmetic will be such that there are statements that the machine can neither prove nor disprove but that human mathematicians can see to be true.
 - b. If human mathematicians can see statements to be true that machines instantiating formal systems can neither prove nor disprove, then human mathematical reasoning—hence, human intelligence—is not the instantiation of a formal system.
 - c. Thus, human intelligence is not the instantiation of a formal system— and "minds are essentially different from machines."
3. But!!!
 - a. Step 2 is false: we might be able to see the truth of the relevant statements even if our intelligence does involve the instantiation of a formal system.
 - b. Perhaps we are **inconsistent** formal systems. As Turing himself argued in a different article, there is no guarantee that we are consistent in our reasoning, though we may often seem to ourselves to be. Inconsistency could arise through reliance on heuristics in our thought.
 - i. Recall that the theorem only holds for those they are both consistent and outfitting with vocabulary sufficient for expressing the truth of arithmetic.
 - ii. Maybe the formal system our minds implement is inconsistent as such the result does not apply.
 - iii. Turing claims that there's no guarantee that we are consistent in our reasoning though we may often be seen as ourselves to be.
 - c. How could inconsistency arrive from us if we are mechanistic things, like digital computers?
 - i. through reliance on heuristics
4. An Important Qualification: the Need for Heuristics

- Example 1: Newell and Simon's Logic Theory Machine
 - i. It found proofs of theorems by showing how they can be derived from axioms via permitted inference rules.
 - ii. This program did NOT work by mere brute force, by way of an algorithm that does nothing but search through all possible proof solutions.
5. The Reason Why Not
- "Suppose that the rules of inference are such that at any instant, a string of symbols could be changed in ten different ways. This means that if we want the machine to check all possible 6-step 'proof chains,' it would have to go through one million (i.e., 10^6) different possibilities. We can see that such a '**combinatorial explosion**' [...] means that exhaustive search through the space of possible symbolic transformations would have been prohibitively time-consuming"*
- (Walmsley 2012, 52-53).
6. An Important Qualification: the Need for Heuristics
- Example 2: Deep Blue and AI work on chess-playing computers.
 - Again, this program did NOT work by mere brute force, by way of an algorithm that does nothing but search through all possible moves at any given moment of play.
7. The Reason Why Not
- Though "perfect" or "exhaustive" chess is a theoretical possibility, "the numbers involved in chess make this approach practically unfeasible. There are somewhere around $(10^{120}) \times 40$ -move chess games [...]. But current estimates put the total number of atoms in the known universe at around 10^{80} , so even if you used one atom per tree branch, there wouldn't be enough matter in the universe to represent all of the possible games. Further, current calculations put the age of the universe at less than 10^{25} nanoseconds, so even if you had been evaluating one tree branch every nanosecond since the big bang, you wouldn't yet have had time to calculate your first move"*
- (Walmsley, 56-57).
8. Heuristics
- a. A heuristic is a "rule of thumb" for solving a problem that, while not guaranteeing a solution, nevertheless makes one fairly likely.
 - b. Newell et al. 1957: a heuristic is "a process that may solve a given problem, but offers no guarantee of doing so" (114).
 - The Logic Theory Machine's programmed heuristics were based on observation of how actual logicians and mathematicians construct proofs.
 - An instantiation of a formal system that includes heuristics as relevant instructions.
 - Deep Blue's programmed heuristics were worked out in consultation with several leading chess grandmasters.
9. The Reason Why Heuristics Matter in the Present Context
- If human intelligence is computational and yet involves reliance on heuristics, then the algorithms it employs may on occasion leave us with contradictory—hence, inconsistent—beliefs.

- a heuristic cannot lead to a certainly right result.
- if your thought processes unfold that computational lines. Once where heuristics are amongst the coded instructions, then you may very well end up with contradictory beliefs.

10. Objection 6: Lady Lovelace's Objection

- a. Lovelace, in her memoir, on Babbage's "Analytical Engine": "The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform" (from Turing, 450).
- b. What, exactly, is the objection here? (What does it have to do with whether computers can do new things or with whether they can surprise us?)

11. Turing's Reply

- For his part, Turing offered several responses to the objection that "a machine can 'never do anything really now'" (450). Among them:
 - i. For all we know, learning of some sort is possible in digital computers.
 - ii. How sure can we be that what we mean by "new" or "original" isn't the result of hidden mechanistic factors?

12. Parting Questions

- (1) According to the "heads in the sand" objection, "[t]he consequences of thinking machines would be too dreadful." Why think such a thing? How many different reasons can you come up with? How should someone untroubled by AI respond?
- (2) What, exactly, is the "argument from consciousness" against thinking machines? How plausible is it?
 - (a) no machines can be conscious in the way I am. Since intelligence requires consciousness there must be a major difference between machines and human intelligence.
- (3) Of all of the objections Turing considers, "Lady Lovelace's objection" has received the most attention. What do you think the objection is supposed to be? How plausible is Turing's reply?

Section Two: Town Hall& Video 3 (the Mind-Body Problem & Functionalism):

1. Today's Agenda

- a. Review from last week
- b. The metaphysics of mind: from dualism to functionalism

2. The Metaphysics of Mind

- a. Last week we discussed Turing on the nature of digital computation and on the question whether digital computers can think.
- b. As we saw, the success of early AI motivated the computational theory of mind. Another key impetus behind the development of that theory was the rise of functionalism about mental states.
- c. Functionalism is a metaphysical account of the nature of mental states.

3. The Mind-Body Problem (as represented by Campbell 1984)

Each of the following propositions is plausible. But they cannot all be true. Consider:

- (1) The human body is a material thing.
- (2) The human mind is a spiritual thing.
- (3) Mind and body interact.
- (4) Spirit and matter do not interact.

(1), (2), and (4), for example, jointly entail:

(5) Mind and body do not interact.

But that contradicts (3)!

4. Some Options

- **Idealists** reject proposition (1). They say: the human body is not material.
- **Materialists** reject proposition (2). They say: the human mind is not spiritual.
 - they do not deny or have to deny existence of things with consciousness and intentionality
 - is not about deny the existence of mind
- **Dualists** (accept the first 2 & reject one of the other)
 - **Non-interactionist dualists** (e.g., epiphenomenalists) reject proposition (3). They say: mind and body do not interact.
 - The basic concern about non-interactionist dualism is that it seems to fly in the face not only of common sense but also of standard cognitive science.
 - **Interactionist dualists** reject proposition (4). They say: spirit and matter can and do interact.

5. Interactionist Dualism vs Materialism

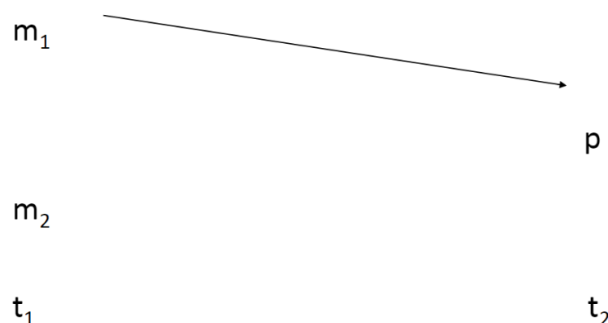
- a. The idea basically was that it is just too difficult to see how phenomenality and intentionality could be states/processes of purely material systems.

6. Dualism and the Problem of Mental Causation

- Two forms of (interactionist) dualism:
 - i. **Substance dualism**: Mental states and processes are wholly immaterial states and processes of a wholly immaterial substance.
Example: Plato Descarte
Mind is a purely immaterial thing (a soul, a spirit).
 - ii. **Property dualism**: Mental states and processes are (at least partly) immaterial states and processes of an otherwise material substance.
No wholly immaterial substances.
A brain (a psychical object). Your brain's physical material properties do not exhaust its nature.
- Both face a serious problem of explaining how what is “dualist” about them can be compatible with what is “interactionist” about them.

7. A Concern for Substance Dualism: the “Pairing Problem”

The Pairing Problem



- a. at time t1, we have 2 distinct mental events, m1 and m2, each one taking place in a different person's mind.
 - b. m1 be a mental state in Smith's mind, m2 in Jone's (Smith's twin). *make a decision
 - c. At t2, arm of Smith's body goes up causes p. m2 causes other event.
 - d. Smith are exactly a like and m1 and m2 are exactly alike. only difference is m1 is Smith's and m2 is Jone's.
 - e. In virtual of what, is p caused by m1 and not m2? That arrow is Smith's decision to make Smith's arm to get up is because of what?
 - f. We use space to distinguish them. (Smith and John)
 - g. if substance dualism is true, then m1 and m2 are not events in Smith and John's bodies, they are a event in Smith and John's imaterial lives. There is not spatial pathway to connect m1 or m2 to p.
8. A Concern for Property Dualism:
The "Exclusion Problem"
*closed physical system
- a. Consider **Causal Closure**: "every physical event with a cause at all has a sufficient physical cause." The dualist seems to face a choice:
 - i. Either deny Closure (at risk of violating the conservation of energy?)
 - injection of new energy
 - ii. Or accept Closure (and along with it massive mental-physical causal overdetermination?)
 - more than one sufficient causes.
9. A Radical Materialist Reaction
- ❖ By the early 20th century, many philosophers and psychologists were convinced that the only way to deal with dualism's difficulties with mental causation was to embrace behaviorism.
 - i. **Eliminativist behaviorism** in psychology: There are no mental states/processes; there is only observable bodily behavior.
 - ii. **Logical behaviorism** in philosophy: All claims apparently about mental states/processes are analyzable without remainder in terms of claims about nothing but observable bodily behavior or dispositions towards observable bodily behavior.
 1. Smith is hungry -> Smith is either behaving in such ways or disposed in certain ways.
 2. If Smith is offered food, Smith will eat.
10. Against Eliminativist Behaviorism
- a. Behaviorism in psychology and the other mind sciences died out as a research program. After the "cognitive turn," the most successful science posited unobservable internal "mental" states and processes.
 - b. Some examples:
 - Cognitive psychology: working memory, feature maps, selective attention
 - Theoretical linguistics: syntactic rules, logical forms, Wh-movement
11. Functionalist treats minds, mental states as real entities in the world. They do not treat them as neural, brain events but computational functional role items.
12. Against Logical Behaviorism
Where to even begin?!?!?

- a. **Apparent counterexamples:** “perfect actors” (no pain but has the behavioral dispositions) and “Super Spartans” (pain but no behavioral dispositions).
 - i. I’m in pain \Leftrightarrow i’m not being disposed to engage in a pain behavior
 - 1. a perfect actor can behave as she is in pain but she is not really.
 - ii. Train themselves. Take a hammer and slam it down. They felt not tendency to shout and “ouch”. but he is indeed in pain.
- b. **The circularity problem:** there is no plausible way of analyzing mental claims in terms of observable bodily behavior that doesn’t involve at least some mentalistic residue.

13. A Different Materialist Reaction

- a. **“Central state materialism”** (aka the “mind-brain identity theory”): mental states/processes are wholly material states/processes of the brain and central nervous system.
- b. A slight complication: the view was first advanced so as to apply only to phenomenal mental states/processes. Something like behaviorism was true, it was thought, for non-phenomenal aspects of mind.

14. But! The Problem of Multiple Realizability

“Consider what the [identity theorist] has to do to make good his claims. He has to specify a physical-chemical state such that any organism (not just a mammal) is in pain if and only if (a) it possesses a brain of a suitable physical-chemical structure; and (b) its brain is in that physical-chemical state” (Hilary Putnam, “The Nature of Mental States”(1967)).

15. A Dilemma, or “The Mind-Body Problem Reformulated”

- a. It is difficult to see how intentionality and phenomenality could be material, so dualism!
- b. But the problem of mental causation is very grave indeed for dualism!
- c. But materialism seems no better.
- d. Behaviorism: yuck! Identity theory: the problem of multiple realizability! So where do we go from here?!?!?

16. A New Approach: Functionalism

- a. Functionalism defines mental states in terms of causal (“functional”) roles. It says that mental states are internal states with certain characteristic causes and effects.
- b. The first, most basic form of functionalism was “machine functionalism.”
- c. To understand what it involves we need to recall what we learned about Turing machines and their corresponding machine tables.

17. An Example Case: a Toy Soda Machine

- a. The Rules for Our Soda Machine
 - i. If input (N) and in state (0), then enter state (5) and output (wait).
 - ii. If input (N) and in state (5), then enter state (10) and output (wait).
 - iii. If input (N) and in state (10), then enter state (0) and output (soda).
 - iv. If input (D) and in state (0), then enter state (10) and output (wait).
 - v. If input (D) and in state (5), then enter state (0) and output (soda).
 - vi. If input (D) and in state (10), then enter state (5) and output (soda).

(Note: We could have changed the rules so that the soda machine gives change.)
- b. The Machine Table for the Soda Machine

		<u>Inputs</u>	
		N	D
<u>Inner States</u>	0	5 / wait	10 / wait
	5	10 / wait	0 / soda
	10	0 / soda	5 / soda

c. Defining “Inner States” of the Soda Machine

- i. Equipped with this machine table, we can define the notion of an inner state of our soda machine. For example:
- ii. Inner state 5 = the state that is caused by inputting a nickel when the machine is in state 0 and by inputting a dime when the machine is in state 10, that causes an output of wait when given an input of a nickel and an output of a soda when given an input of a dime, and that makes the machine enter into state 10 when given a nickel and enter into state 0 when given a dime.
- iii. This definition defines “inner state 5” in terms of (i) the inputs that typically cause it, (ii) the outputs it typically causes, and (iii) the inner state changes it typically causes.
 1. defined the internal state of the machine in terms of functional roles.

18. Super Important!

- Our machine table, and the definitions it provides, makes our toy soda machine multiply realizable: it could be made out of anything so long as it is capable of having inner states like our 0, 5, and 10 with causal powers like the ones spelled out in the table. Functionalism says that mental states like believing that water is wet and feeling a sharp pain can be given functional definitions just like these.

19. From Soda Machines to Minds: A Simple Theory of Pain

		<u>Inputs</u>	
		Hammer Smashing Finger	Taking Aspirin
<u>Inner States</u>	P	P/“Ouch!!!”	R/“Ahh!!!”
	R	P/“Ouch!!!”	R/“Ahh!!!”

20. Defining “Inner” Mental States

- a. This machine table allows us to define the mental state of pain:
- b. Inner state P = the state that that is caused by a hammer smashing one’s finger, that causes one to say “Ouch!!!” when one is in P and a hammer

smashes one's finger, and that causes one to say "Ahh!!!" when one is in P and one takes an aspirin.

c. oversimplified model

21. Mental States as Multiply Realizable

- This is incomplete (as well as ridiculously oversimplified). But the basic idea is the same as the idea pertaining to our soda machine. This machine table, and the definitions it provides, makes the inner mental states of pain and relief multiply realizable. Anything could have them so long as it is capable of having inner states like our P and R with causal powers like the ones spelled out in the table.

22. A Complication

- If functionalism is true, then functionalists owe us an answer to an important question:
- "What, exactly, are the functional roles in terms of which the theory defines mental states?"
- Think back to our soda machine and pain-relief mind examples.
 - The functional roles for our soda machine were derived from the rules encoded by the machine table. **We stipulated these rules.**
 - The functional roles for our pain-relief mind were derived from the rules encoded by the machine table. We stipulated these rules.
- But we don't stipulate the rules (assuming there are such) according to which our minds operate. That's a job for God or Mother Nature or
 - So how do functionalists answer the important question?
 - In two very different ways...

23. A Complication—Resolved

- "Commonsense" or "analytic" functionalism**—The functional roles in terms of which mental states are defined are a priori platitudes entailed by our ordinary concepts of belief, desire, pain, and all of our other mental states.
 - Quick review: A proposition is a priori if the justification for believing it is independent of sense-experience; a proposition is a posteriori (not a priori) if the justification for believing it depends (at least partly) on sense experience.
- "Psycho"-functionalism**—The functional roles in terms of which mental states are defined are at least partly derived from the cognitive sciences.

24. From Machine Functionalism to Contemporary Functionalism

- Contemporary advocates of both commonsense functionalism and psycho-functionalism have moved beyond the simplistic machine table-based way of supplying mental state definitions covered here.
- Instead, they endorse a more sophisticated approach, one which defines mental states in functional terms using the method of "Ramsey sentences." —See section 4.1 of Levin's Stanford Encyclopedia of Philosophy entry on functionalism for a succinct summary.

25. An Additional Wrinkle

- An extra complication has to do with the crucial distinction between "role state" and "realizer state" functionalism.

26. First- vs. Second-Order Properties

- A second-order property is the property of having a (first-order) property that meets a certain condition.

- Example: If you are from Rome, you have the property of being from Rome. Now suppose that being from Rome is one of my favorite properties. Then, in addition to having the first-order property of being from Rome, you have the second-order property of having a property that is one of Jim's favorites.

27. Role State Functionalism

- a. Role state functionalism—Mental states are identical with second-order properties that consist in having first-order properties that realize certain functional roles.
- b. Example: pain = the property of having a property that realizes the PAIN functional role (where the PAIN functional role is given by the machine table/Ramsey sentence)

28. Realizer State Functionalism

- a. Realizer state functionalism—Mental states are identical with the first-order properties that realize certain functional roles.
- b. Example: pain = the property that realizes the PAIN functional role (where the PAIN functional role is given by the machine table/Ramsey sentence)

29. A Final Question

- a. Do any of the functionalist views we've just canvassed allow for the desired causal efficacy of the mental?
- b. It isn't obvious that the answer is "yes." How serious a problem is that for the functionalist?