Philosophy 1320: Theories of the Mind, Stern College - Yeshiva University, Spring 2007
Russell Marcus, Instructor
email: philosophy@thatmarcusfamily.org
website: http://www.thatmarcusfamily.org/philosophy/Mind/MindHome.htm

Lecture Notes, March 26

I. Consciousness, identity theory, and behaviorism

In the lecture notes for last class, I mentioned that Armstrong defends scientific methodology since science solves problems and makes progress.
Armstrong's methodological scientism, which supports the identity theory, is unobjectionable.
But, the identity theorist shares a problem with the behaviorist.
Both the materialist and the behaviorist seem to lack an account of consciousness, of the way our mental states feel to us.
Hobbes was vulnerable to this criticism, which we can trace back, at least for sensations, to the primary/secondary distinction of Galileo and Locke.

Armstrong provides an account of the missing elements, in a way that Hobbes did not.
He considers actions that we perform absent-mindedly, like driving while not thinking about it, or walking.
Armstrong says that the missing elements of the materialist's theory are just like the experiences we neglect when not paying attention to walking or driving.
Our conscious awareness, in those cases, is missing.
But all that conscious awareness comes to is that we are not thinking about what we are doing.
If Armstrong is right about both the analogy and his account of the absent-minded activity (that it lacks only a thought) then consciousness is just thinking about thinking.
And the missing elements of the materialist's theory are just some kinds of thoughts.
We will return to the problem of consciousness and qualia at the end of the term.
Now, we proceed to examine functionalism.

II. Eliminating the semantic argument against materialism.

The beginning of Putnam's article contains some issues in the philosophy of language.
Twentieth-century philosophy, from 1879 (Frege's *Begriffsschrift*) through, say, the Chomskyan revolution, was focused on our uses of language.
One thread of this focus, started by Frege, and traceable through Russell, early Wittgenstein (as represented by the *Tractatus*), Carnap, and Quine, was a project to formalize all natural language for the purposes of precision and clarification.
Another thread, from later Wittgenstein (as represented by the *Philosophical Investigations*) through Ryle and Austin and Grice, was to examine ordinary language for its obscured meanings.
Both movements are now mostly dead, though their latent effects remain with us.

The point of Putnam's discussion of properties and concepts is just to block an argument against materialism, and for dualism.
The argument against materialism is that since the meanings of the phrases 'is in brain state B' and 'is in mental state M' are clearly different, mental states must not be brain states.
Notice that the argument derives a metaphysical conclusion from a semantic premise.
Putnam uses 'property' as what predicates stand for and 'concepts' as synonymy classes; but don't worry

about those definitions.
A clearer way to understand the difference is to consider that properties belong to objects, and concepts belong to our thoughts about objects.
Or, properties are metaphysical, and concepts are semantic.
As Putnam shows, the property of temperature is the same as the property of mean molecular energy.
That is, scientists have discovered that the temperature of an object can be reduced to, explained by, eliminated in favor of, the motion of its particles.
But, the concept of temperature, what we think of when we think of the temperature of an object, is distinct from the concept of kinetic energy.
We can see that by merely noting that people had the concept of temperature long before they knew anything about molecular theory.
Distinctions among concepts are more fine-grained than distinctions among properties.
Another example: being water and being $H_2O$ are the same property, but different concepts.

The point of the temperature example, of course, is to show that mental state properties can be identical to brain state properties, even if the correlated concepts diverge.
The property of being in pain, for example, can be identical to the property of being in a certain brain state.
That is, mental states can be identical to brain states, even if we do not know that they are.

III. Multiple realizability and identity theory

Still, Putnam does not argue that pains are brain states.
Mainly, his argument against identity theory is that it is chauvinistic: it attributes mental states only to creatures with human brains.
Some psychological states seem shared by animals.
We can imagine organisms that are made of completely different stuff, but which have the same kinds of mental states that we do.
Aliens made of silicon, instead of carbon, may have pains, color sensations, beliefs and desires, etc., and yet not have brains which are in specific ways identical to ours.
If pain is just a specific state of a brain, we must deny that the aliens have mental states.
Putnam's argument is generally known as the multiple realizability argument.
Functionalism is the result of considering both behaviorism and identity theory in the light of multiple realizability.
So, it will be worth our time to explore the problems of multiple realizability in detail.

In order to focus the range of the problems arising from multiple realizability, Fodor distinguishes token physicalism from type physicalism.
Token physicalism says that every instance, or token, of a mental state is identical with a token of a physical state.
Token physicalism is naturally taken as a denial of dualism.
For, if token physicalism is right, then there are no mental states that can not be explained by physical facts.

Type physicalism is a further claim, that every type of mental state is identical with a type of physical state.

According to type physicalism, we will be able to find specific physical states that correspond to any mental state, like pain, or the sensation of seeing red, or the belief that aliens live on Mars.
Identity theory is most naturally taken as type physicalism.
The theory will be made up of a series of clauses, like:

x has a toothache iff x is in brain state $S_{412}$
x is seeing blue iff x is in brain state $S_{7583}$
etc.

Recall that the identity theorist relies on the precedents set by other, paradigmatic theoretical reductions.
Mental states are brain states in the same way that heat is energy.
For any theoretical reduction, we specify essential properties of the reduced phenomena, e.g. pain, heat.
So, the heat of a gas is always the average kinetic energy of the molecules of that gas.
Similarly, according to the identity theorist/type physicalist, we should be able to find the specific brain states that always correspond to pain, and all other mental states.
A thing has a toothache iff it is in brain state $S_{412}$; a thing is seeing blue iff it is in brain state $S_{7583}$ ....

Type physicalism seems most plausible for mental states that correspond to occurrent sensations.
Fodor and Putnam agree that identity theory suffers from problems arising from multiple realizability.
In fact, all type physicalism suffers from multiple realizability problems; see Fodor p 332.
For, if there are different brain states which can correspond to the same mental state, the type physicalist seems to be in trouble.
It seems wildly implausible that the belief that the Mets will win the World Series this year corresponds to the same exact brain state in every one who believes it.
There will be no single $S_n$ to correspond to the same belief in different people, in the way that heat always corresponds to kinetic energy.
For one thing, we might want to attribute this belief to aliens or to machines, eventually, who clearly do not share our brain structures.

So, the first problem that multiple realizability raises for identity theory is its chauvinism.

1. Identity theory is chauvinistic.
2. Multiple realizability shows that chauvinism is wrong.
So, identity theory is false.

A second and related problem for identity theory arising from multiple realizability is that even human brains do not all work the same way.
My brain state, when I see blue, will be different from your brain state, when you see blue.
So, instead of the clauses above, identity theory will have the following sorts of clauses;

$x_1$ has a toothache iff $x_1$ is in brain state $S_{412}$
$x_2$ has a toothache iff $x_2$ is in brain state $S_{6224}$
$x_3$ has a toothache iff $x_3$ is in brain state $S_{91}$
...
So, x has a toothache iff $x=x_1$ and is in $S_{412}$ or $x=x_2$ and is in $S_{6224}$ or $x=x_3$ and is in $S_{91}$ or ...

We call a theory like the one made up of the last clause disjunctive, since it says that one mental state is

identified with any of a variety of physical states.
In the full version of their article, Fodor and Block attribute this claim, called neurological equipotentiality, to the early-20th century psychologist Karl Lashley.
In support of the Lashleyan claim, consider that language is normally processed in the left hemisphere for righties, but people with damage in the left hemisphere may process language in their right hemisphere.

At the end of his article, Fodor refers to a third multiple realizability problem closely related to equipotentiality.
He says that the identity theorist lacks a relational construal of mental states, p 332.
A relational construal sorts mental states according to the relations among stimuli and responses.
For the identity theorist, we sort, or type, mental states according to their physical properties.
Consider how we would respond to the discovery that two disparate mental states, say a leg cramp and the belief that chocolate pudding is tasty, had the same physical instantiations, i.e. were correlated with the same brain states.
If we sort mental states according to their content, how they seem to us, it does not matter that the pudding belief and the cramp sensation are instantiated by the same brain state; they are two strikingly different mental states.
But, for the identity theorist, we would have to say that they are the same state.
For, mental states are just brain states.
The following inference would be thus inevitable, for the identity theorist.

> x has a leg cramp iff x is in brain state $S_{3313}$
> x believes that chocolate pudding is tasty iff x is in brain state $S_{3313}$
> So, x has a leg cramp iff x believes that chocolate pudding is tasty.

Fodor urges that any theory of the mind should yield mental states that constitute natural kinds, for psychological purposes, p 331-2.
Another way he puts the point is that mental state terms should be projectible.
The issue of natural kinds and projectible predicates (or properties) could take us on an interesting tangent.

## IV. Natural kinds and mental states

The notions of natural kinds and projectible predicates come from a solution to an ingenious version of the old, Humean problem of induction.
The problem of induction is the problem of justifying inferences concerning the future on the basis of past experience.
We make inferences about the future all the time, any time we appeal to physical laws for prediction.
But, these predictions rely on the belief that the laws will extend from the past into the future.
Past experience only provides justification for believing that the laws have held in the past.
No experience can tell us that the laws will continue to hold.

The new riddle of induction, owed to Nelson Goodman, in *Fact, Fiction, and Forecast*, applies the problem with laws to simple terms, even the most common predicates.
You know what it means for an object to be green.
Consider the property called ‘grue’.

An object is grue if it is green until 1/1/2010, when it suddenly turns blue.
How can you tell if a plant is green or grue?
All evidence for its being green is also evidence for its being grue.
Green things and grue things are exactly alike until 2010.
One could construct other artificial properties, like the property of being a papod.
A papod is a piece of paper which, on 1/1/2010, turns into an Ipod.
All papods look exactly like pieces of paper right now.
There is, in principle, no way to tell them apart.

Still, my predictions that the plant will be green tomorrow, and that thing I think is a piece of paper will not turn out to be a papod, both will turn out true.
One way to describe the success of predictions is to say that predicates like 'green' and 'paper' are projectile, will remain constant through time, whereas predicates like 'grue' and 'papod' are not projectible.
One account of the projectibility of certain predicates is that they refer to natural kinds.
'Green' is a natural kind; 'grue' is an unnatural, deviant philosophical construct.
Scientific theories should refer to natural kinds, but not to deviant, gerrymandered kinds.

Fodor's claim, then, is that mental states are natural kinds, and that ordinary terms which refer to our mental states should be taken as projectible predicates.
Since mental states are natural kinds, theories of the mind must refer to them, not eliminate them.

V. Multiple realizability, behaviorism, and disjunctive theories

We have looked at three problems for identity theory which come under the heading of multiple realizability.
1. Chauvinism
2. Neurological equipotentiality
3. Non-relational construal of mental states
These problems all apply to the type physicalist, and identity theory is naturally taken as a version of type physicalism.
The problems of multiple realizability apply to both the identity theorist and the behaviorist.

Token physicalism can survive problems of multiple realizability, since even if we all have different brain states corresponding to relevantly similar mental states, they are all still physical states.
Recall that to accommodate multiple realizability, the identity theorist would have to adopt a disjunctive theory of mental states:

> $x_1$ has a toothache iff $x_1$ is in brain state $S_{412}$
> $x_2$ has a toothache iff $x_2$ is in brain state $S_{6224}$
> $x_3$ has a toothache iff $x_3$ is in brain state $S_{91}$
> ...
> So, x has a toothache iff $x=x_1$ and is in $S_{412}$ or $x=x_2$ and is in $S_{6224}$ or $x=x_3$ and is in $S_{91}$ or ...

Token physicalism says, at heart, that such disjunctive theories are perfectly acceptable.

Compare the disjunctive account with the behaviorist's correlations between mental states and behaviors.
There are no unique behaviors that correspond to particular mental states.
Some people react to the same painful stimulus by screaming, others by wincing, others by stomping about.
So, the behaviorist's identity sentences will look like: a thing is in pain iff it exhibits behaviors $B_1$, or $B_2$, or $B_3$, or...
That is, the behaviorist is already committed to a disjunctive theory.
If multiple realizability was not a problem for the behaviorist, maybe the identity theorist can also try a disjunctive theory, and hold on to token physicalism.
The identity theorist would then correlate pain with any of a variety of brain states, so that we can have pain-in-a-robot, and pain-in-a-Martian, etc.

One problem with the disjunctive approach is that it is incompatible with the idea that a given behavioral state or brain state may realize different psychological properties at different times.
The same behavior, say squinting, might be evidence of pain, or concentration, or blurry vision, or...
Similarly, we saw that the identity theorist's non-relational construal of mental states entailed that the same brain state may be correlated with different mental states.
Thus, on either disjunctive approach (the behaviorist's or the identity theorist's), we have lots of disjuncts on both sides of the equations.
If we have long disjunctions on both sides, we do not seem to be getting anywhere.
The fact that the behaviorist was liable to multiple realizability criticisms does not show that these criticisms are superable.
Rather, it shows that multiple realizability was a problem for the behaviorist as well.

VI. Putnam and Fodor against behaviorism (or, beating a dead horse)

Putnam and Fodor agree that the behaviorist program is fruitless, for reasons independent of multiple realizability.
Putnam's best argument is to consider two people whose motor nerves are cut, but only one of whom has cut pain fibers.
Now, kick them both, hard.
One feels pain, and the other does not.
But, they have the same behavior, even potential behavior.

Fodor's worry about the behaviorist is couched in terms of mental causation.
Often, our overt behavior is the result of long causal chains of thoughts.
Consider the chess-player example, p 330.
Every thought in the chess-player's sequence of thoughts would have to be explained in terms of dispositions to behave.
But, there is no overt behavior to distinguish among the distinct thoughts, to guide the train of thought.

More subtly, how can the behaviorist explain, 'John was disposed to produce headache behaviors because he had a headache'?
If 'he had a headache' is explained as 'John was disposed to produce headache behaviors', then the resulting behaviorist analysis becomes tautological, and bereft of any explanatory power, while the original sentence is not.

## VII. Functionalism, finally

We have considered problems for behaviorism and identity theory in order to find a better approach.
Functionalism was designed to avoid the problems we have seen, but maintain the useful insights of both theories.
Putnam's functionalism says that mental states are functional states of probabilistic automata.
Being in pain, or seeing blue, or believing that the moon is made of cheese, are functional states of an organism.
Functionalism describes the mind by appealing to an analogy with computer software: the mind is the software of the brain.
Just as the same software can be run on different hardware, the same mental states can be instantiated by distinct physical (or, even, non-physical) systems.
The mind is not identified with any particular hardware.
Two things are in the same mental states if, and only if, they have the same state of their programs.