Philosophy 355: Contemporary Philosophy — Hamilton College
Fall 2008 — Russell Marcus
Tuesdays and Thursdays, 9am - 10:15am — rmarcus1@hamilton.edu

Class 4 - Functionalism

0. Metaphysics, Epistemology, and Semantics

Much of Place's article on identity theory, and three of the questions I asked about it, focus on the question of the nature of the identity between mental states and physical states.
Place claims that the identity is contingent, the result of empirical investigation, rather than necessary.
The point of his distinction between 'is'es of identity and 'is'es of composition is to substantiate the claim that the identity of brain states and mental states is, like other identities, a mere contingent identity.
And, the claim that mental states and brain states are logically independent is to the same purpose.
All of this work is intended to block an objection, which we see (below, §1) in Putnam's article, too.

The objection is that there seems to be a basic confusion in the identity theory, since I can know a lot about my mental states without having any knowledge of my brain states.
If brain states were identical to mental states, then the terms for brain states should be interchangeable with the terms for mental states.
And, I should be able to substitute these terms for each other without changing the truth values of sentences which contain them; I should be able to make inferences, like the following, on the basis of these identities.

1. Professor Marcus is the teacher of Philosophy 355
2. Professor Marcus is identical to Russell.
So, Russell is the teacher of Philosophy 355.

But, when it comes to mental states, we have a problem, it seems.
For, the following argument, even given the identity in the second premise, does not hold.

1. Aristotle believed that he had a toothache.
2. Toothaches are stimulations of C-fibers in the brain.
So, Aristotle believed that his C-fibers were stimulated.

Now, it looked to Place, and to Putnam's interlocutor in §1 of his article, as if the problem with the above argument was in the identity at Premise 2.
Place's claims that the identities are contingent were intended to block inferences of this sort.
That is, he is arguing that we should not expect substitutivity salva veritate from contingent identities.

The problem with the Aristotle argument has nothing to do with the status of the identity in Premise 2.
The problem is clearly with the opacity of the 'believes that' context.
It seems to me that Place's interlocutor, Place, and Putnam's interlocutor, are all confusing semantic properties with metaphysical ones.
Clearly, the terms 'toothache' and 'c-fiber stimulation' have different meanings.
They have different meanings, or senses, even if they have the same referent.
Reference is a metaphysical question; meaning is semantic.
The argument from last class also fails, even though we don't have any question that lightning is electrical discharge.

1. Aristotle believed that there was lightning.
2. Lightning is electrical discharge.
So, Aristotle believed that there was electrical discharge.

Aristotle could not have believed that there was electrical discharge, even though he knew quite well that there was lightning.
Again, the problem is the opaque context, and not the identity.

I. Eliminating the semantic argument against materialism.

The beginning of Putnam's article contains some issues in the philosophy of language, but the argument is essentially the same one we saw in Place's article.

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 the argument against materialism, which may also be taken as an argument 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 argues, in the spirit of Place's argument, 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 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 can be identical to the property of being in a certain brain state.
That is, mental states could be identical to brain states, even if we do not know that they are.

II. 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, it will be convenient to distinguish, as Fodor does, between token physicalism and 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.

[Aside on dualism: Fodor, in the second paragraph of his article, says that the stumbling block for the dualist is the problem of interaction. As I mentioned in class, I think the problem of interaction is over-emphasized: once you have magic, more magic isn't really a problem. In the next paragraph, Fodor says that the problem is ontological parsimony, and I think this point is exactly right.]

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, having a toothache, seeing blue, etc.
Type physicalism seems most plausible for mental states that correspond to occurrent sensations.

Fodor and Putnam agree that 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 is in trouble.

[Aside: I feel that I may have mangled, in class, the issues of token and type physicalism and the question of whether we can have two distinct mental states at the same time. Here's what I want to say: the dualist allows that it is possible for two different mental states to be correlated with the same physical states. For, the mental state is explicable, strictly speaking, only in terms of its relations with other mental states. The brain is not the foundation, or the supervenience base, of the mind. The type physicalist/identity theorist of course denies the possibility that the same physical state could be two distinct mental states. Since the mind is the brain, a different brain state necessarily entails a different mental state and a different mental state necessarily entails a different brain state. Even the token physicalist, though, will have to deny that the two different mental states can be correlated with the same brain state. For, if the same brain state were correlated with different mental states, there would be no way for the token physicalist to explain the difference.]

It seems wildly implausible that the belief that the Mets will win the World Series this year corresponds to the same 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 reason, 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.
Elsewhere, Fodor attributes 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.

Fodor also notes that the identity theorist lacks a relational construal of mental states, p 332, which is a third multiple realizability problem, closely related to equipotentiality.
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.

III. 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.

IV. 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.

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 other mental states.
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 seem to be saying that any of an indefinite number of mental states can be correlated with any of an indefinite number of physical states, which is a lot like saying nothing at all.
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.

V. 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.

VI. 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.

Putnam's functionalism claims that minds are probabilistic automata.
He explains the notion of a probabilistic automaton in terms of a Turing machine.
Turing machines are just computers.
A Turing machine contains, in its machine table, a complete list of possible states of the system, and possible inputs, and the output.
The actions of a Turing machine (what it writes, where it goes, what state it is in) are completely determined by its algorithm, or set of rules.
An algorithm is a list of instructions, a procedure.
Computer programs are algorithms; cooking recipes are algorithms.
An algorithm can also do different things depending on the state of the system executing the algorithm.
Thus, some algorithms, like the one we generally use for long division, contain conditional clauses: if...then... statements.

Essentially, the computer is just a mechanism which reads input, has internal states, and computes output on the basis of those states and its instructions.

[Aside on Turing: Turing wanted to replace the question of whether machines can think with the question of whether computers could fool people into believing they were human. This strategy for determining whether a machine is thinking is called a Turing test. Some people argue that computers already can think: "Saying Deep Blue doesn't really think about chess is like saying an airplane doesn't really fly because it doesn't flap its wings." - Drew McDermott;
ftp://ftp.cs.yale.edu/pub/mcdermott/papers/deepblue.txt
McDermott's claim is just the chauvinism problem of multiple realizability. If we want to allow that silicon-based aliens can think then we might also allow that computers can think. If it turns out that minds are essentially Turing machines, it will follow that machines can think.]

Putnam probabilistic automaton has sensory inputs and motor outputs.
Though, the effect of an input can be merely to change the state of the system, with no motor output.
A probabilistic automaton has the same structure as a completely deterministic Turing machine, but with probabilistic responses.
A description of a system just says that there is an object with a particular machine table, which Putnam calls the functional organization of the object.
The total state of the object (or system) will be the state of the whole system at a particular time.
For a computer, the total state will include what processes are running, what output is going to the screen or the speakers, and which switches are open and closed on the circuit board.

So, Putnam's claim has four clauses, p 434:

1. Anything that has mental states is a probabilistic automaton. It must have appropriate receptors (e.g. light receptors for colors) and states.
2. Each mental state is a description. There are lots of descriptions of pain, fewer, presumably, for seeing fuchsia; more for being happy, fewer for believing that there is an elephant walking up College Hill Road.
3. An ad hoc clause to prevent us from thinking of nations, say, or ant colonies, as individual persons.
4. Any thing that has a mental state must have the appropriate receptors in the appropriate states.

The essence of Putnam's functionalism is that mental states are computational.
Or, the mind is the software of the brain.

Functionalism has proven to be a fruitful research area.
There are different forms of functionalism.
Some claim to be non-computational, even, though I do not really understand this claim.

VII. Functionalism, identity theory, and behaviorism

While Putnam rejects behaviorism, functionalism is its intellectual heir, in its reliance on the relations among sensory input and behavior/mental states.
According to both the behaviorist and the functionalist, we type mental states according to behavior, not according to the qualities available by introspection.
Actually, functionalism takes the good from both behaviorism and identity theory; see Fodor, p 332.
The functionalist takes behaviorism's attributions of mental states based on behaviors, and removes its disavowal of internal states, and its reductionist, eliminativist, program.
Behaviorism tried to reduce mental state language to behavior language, with the goal of eliminating any apparent references to immaterial substance.
Functionalism, in contrast, is compatible with substance dualism, since it makes no claim about where and how mental properties are instantiated.

In parallel fashion, the functionalist adopts from identity theory the legitimacy of mental states and an acceptance of the causal connections among them.
Fodor calls this the, "Ontological autonomy of mental particulars (p 332)".
The functionalist dispenses which identity theory's unacceptable chauvinism.

VIII. Ramsey sentences and causal-role definitions of mental states

The functionalist avoids problems of multiple realizability by identifying each mental state with the relevant properties of that state, like its interactions with other mental states, and the behaviors of people in that mental state, while eliminating reference to irrelevant particulars, like brain states.
A thing is in pain iff it has been affected in certain relevant ways, and if it has other concomitant mental and behavioral states (wincing, crying), which are causally related to it.
There is a logical trick on which functionalists rely in order to eliminate irrelevant vocabulary from the theoretical identity sentences of a formal theory of mental states, to achieve the desired level of abstraction.
The functionalist constructs what are called Ramsey sentences.

The Ramsey sentence, essentially, removes specific references to the particular causal structures (say, brain states) at work in our mental life, and replaces them with claims that something has this causal role.
Imagine a scientific description of your whole life: your experiences, your various mental states and how they are connected, the (presumably causal) relationship between your body, including your brain, and those mental states, the resulting behavior.
Replace references to the specifically mental parts of this theory, references to pains, and color terms and beliefs, with variables.

The resulting theory provides a functional, causal-role definition of your mental states.
Pain is whatever has the place that pain has in your life.
It is preceded by physical or emotional blows, and succeeded by characteristic behavior: sometimes avoidance, and sometimes valiant confrontation.
It engenders certain other mental states, fear or anger or resignation, all of which have their own causal-role definitions.

The resulting Ramsified, functionalist theory defines mental states in terms of their functional roles.

> x is in pain iff x has been affected by the kinds of things that cause pain, has other mental states that generally accompany pain, and exhibits the kind of behavior that are associated with pain.

In sorting mental states according to behavior and causal connections with other mental states, functionalism makes identity conditions on mental states very fine-grained.
Unless the machine tables of two organisms match up completely, they can never match up at all.
For, if even one state differs, it throws the whole isomorphism off.
It is unlikely that the causal-role definition of pain in any particular case will look exactly like the causal-role definition in any other case.
But, it will be similar, in many ways, especially if we take the entirety of our lives into account.
The functionalist can appeal to similarity relations among such definitions for a definition of mental states, themselves.

IX. Is functionalism an empirical claim?

Consider the identity theorist's claim that mental-state types are brain-state types.
This claim has two parts.
First, there is a correspondence between every mental state and a brain state.
This first claim is just token physicalism.
Second, there is nothing more to mental states than their corresponding brain states.
The first claim is overwhelmingly likely, independent of the identity theorist's claim.
We can see that the identity theorist's claim goes farther, since the substance dualist will also accept the first claim, but not the second.
Empirical evidence can support the first claim, but no empirical evidence could support the second.
Identity theory is thus not just an empirical claim.
Putnam says that functionalism is an empirical claim.
He argues that we can test functionalism by constructing models, robots, essentially, p 435.
What would these models show us?