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 28

Note: these lecture notes again cover more material than we discussed in class.

I. Turing machines

Putnam's functionalism claims that minds are probabilistic automata.
He explains the notion of a probabilistic automaton in terms of a Turing machine.
To understand Putnam's functionalism, then, we should look at the idea of a Turing machine.
Turing machines are just computers; see pp 257-8.
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 just a list of instructions, a procedure.
Computer programs are algorithms; cooking recipes are algorithms.
Recipes generally just give simple, linear instructions.
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.

Turing's article is needlessly obscure, for the contemporary reader, but it contains two influential elements.
First, as a founder of computing, he proved that digital computers are universal machines.
That is, the only difference between complex digital computers and simple Turing machines is the level of complexity, the immense amount of storage and processing speed, of the computer.
Essentially, though, the computer is just a mechanism which reads input, has internal states, and computes output on the basis of those states and its instructions.
Every action the computer takes is completely determined by the algorithm governing its states and its input.

Second, 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.

II. The Turing test

The Turing test for whether a machine is thinking relies on behavioral criteria for ascriptions of thought.
One might believe that the best criteria for whether people or machines have thoughts are related to their conscious mental states, as Descartes did.
Turing neglects the Cartesian criteria, and argues that machines will be able to think.
(At the beginning of his article, he criticizes analysis of ordinary language, the Wittgensteinian approach, as reducing an important question to an opinion poll.)

Of course, the problem with the Cartesian criteria is that they are impossible to apply.
Turing does consider the Cartesian view in an objection, the argument from consciousness, p 273.
The argument from consciousness says that machines can not think, since they don't have conscious states.
Turing says that the Cartesian criterion for thought (consciousness) leads to a solipsistic point of view.
We can not successfully apply the Cartesian criteria even to our closest friends and family.
So, we can never know that other people have mental states, either.
This problem, of course, is the problem of other minds.

We saw that according to identity theory, mental states are states of brains.
Since computers don't have brains, the identity theorist says that they can not 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.

Turing's question, whether machines can think, and Putnam's assertion, that minds are simply Turing machines, are clearly related.
If it turns out that minds are essentially Turing machines, it will follow that machines can think.

One undeniable difference between computers and minds is that computers are discrete state machines.
The human nervous system may not be accurately describable in terms of discrete states.
Turing considers the objection from continuity in the nervous system, pp 276-7, and insists that the rules of the game restrict its applicability.
All that matters to the question of whether something can think, given Turing's test, is the external behavior, not the internal structure.
Anyway, the human nervous system is likely amendable to a state-treatment, just one that must be very fine grained.

There seems to be some room between Turing's position, that anything that acts like a person thinks, and the solipsistic Cartesian position, that the only things to which we can attribute thought are ones which we can verify are actually conscious.
For example, we might be able to find some empirical criteria for establishing consciousness which could moderate the Cartesian view.
Or, we could adopt a different view about consciousness, one which is tied neither merely to behavior or to only our introspective mental states.
We will return to these issues about consciousness briefly at the end of the functionalism section, and then again at the end of the course.
We will put aside, now, Turing's test, to focus on the application of Turing machines to functionalism.

III. Putnam's functionalism

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.

(Turing considers installing a random number generator; see his discussion of free will, p 268.)
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 323:

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 a cow walking uptown on Lexington Avenue.
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.

Putnam identifies mental states with the states of the machine table.
Fodor and Block show that our mental states can not correlate with machine-table states, p 333.
There are more mental states that humans can have than there are machine-table states.
Instead, we must correlate mental states with computational states, p 334.
There are as many computational states as there are possible mental states, ex hypothesi.

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.

IV. 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.

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

Recall that both the identity theorist and the behaviorist had problems of multiple realizability.
The functionalist avoids these problems 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.

VI. 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 324.
What would these models show us?

VII. Strong and weak AI

Searle's article assesses the ability of models of human minds, like the ones Putnam suggests might support functionalism, to tell us anything about actual human minds.
Searle's article can be taken both as an argument against functionalism and an argument against artificial intelligence.
He distinguishes a strong and a weak AI thesis.
The weak AI thesis is just the unobjectionable claim that machines built to perform tasks that humans perform can give us some insight into the nature of our thought.
But, proponents of AI are committed to a stronger thesis.

Remember that one of Descartes's arguments for dualism was that the mind performs reasoning that no machine can perform.
Cheap calculators can now perform very complicated tasks much more quickly than even the smartest humans.
Machines are already able to do many tasks that once were inconceivable, including proving mathematical theorems that require more computation than humans can perform.
Better machines may approach or overtake human skill in other areas as well.
The strong AI claim is that computers with such skills actually have minds.
The strong AI claim is the same as McDermott's claim about Deep Blue.
The claim entails that we need not know about the structure of the brain in order to know about the structure of the mind.
All we need in order to have a mind is to simulate the behavior.
We can construe 'behavior' broadly, with the functionalist, and include internal states and their causal connections.

Putnam's claim that we can test functionalism by constructing models of human minds is also the same as Searle's strong AI claim.
Putnam's functionalism takes the mind to be the software of the brain.

Functionalism itself is not strong AI, since this catchphrase is just a metaphor.
But the claim that functionalism is an empirical hypothesis is equivalent to the strong AI claim.
To understand minds, according to strong AI, we just need to understand computer models and their software.

The first thing to notice about computers and their software is that they work according to purely formal, syntactic manipulation.
Computers merely follow algorithms.
Moreover, every step of the algorithm can be specified syntactically, by its inscription.
So, if strong AI and Putnam's claim that functionalism can be empirically tested are right, then human behavior must be describably algorithmically as well, and representable in purely syntactic form.

VIII. The Chinese room

Searle's Chinese room example provides an example of a person working according to purely formal, syntactic rules, pp 283-4.
The person in the Chinese room has all the same input as a speaker of Chinese, and produces the same output, without having any understanding of Chinese.

For the behaviorist, who already disowns our inner lives, the Chinese room shows nothing.
But, behaviorism was unacceptable because of its denial of our inner lives.
Functionalism is in fact motivated by the need to account for causal connections among mental states.
Since our mental states matter, the ability to follow formal rules can not suffice for understanding.
Even he or she internalizes all the formal rules, the person in the Chinese room lacks any understanding about the content of the symbols he or she is manipulating.

IX. Searle's argument

1. Programs are completely describable in terms of their formal, syntactic content.
2. Minds grasp the meanings, or semantics, as well as syntax.
3. Syntax alone can not produce semantics.
So, minds are not merely syntactic manipulators; i.e. minds are not mere programs.

Regarding AI, the importance of Searle's argument is that a mechanical model of the mind could not *be* a mind.
The model lacks intentional content, as well as qualia.

In arguing for the importance of the intentional, and sensory, against AI, one might think that Searle was defending the autonomy of the mental, and thus dualism.
But, Searle argues that the proponent of strong AI, instead, is committed to dualism.
Strong AI presumes that the mental is distinct from the brain, since minds might be instantiated by things that are not brains.
Searle insists that the brain, and its causal connections with sensory organs, and the rest of the body, is essential for understanding our minds.

Regarding functionalism, the problem which Searle's argument points to is that there seems to be more to our minds than algorithmic processing of sensory input.
The functionalist was motivated by the desire to account for multiple realizability.
Searle's argument is that there is a virtue in chauvinism.

If Searle is right that the machines are necessarily missing something, behaviorism and functionalism must both be wrong.
Searle also points a finger at operationalism.
Operationalism is just the idea that a term can be defined by the methods (operations) we use to verify it.
So, weight can be defined operationally in terms of our use of scales, temperature in terms of the rise of mercury in a tube.
Operational definitions came into vogue as a way of justifying our knowledge of objects, like electrons, which were not observable.
If we define the mind operationally, we are defining it in terms of its observable characteristics, i.e. behavior.
But, it seems that there is more to the mental than can be operationally defined.

X. Inverted and absent qualia

Searle's Chinese room example is closely related to a constellation of objections to functionalism regarding qualia.
In fact, in the last thirty years in the philosophy of mind, emphasis has shifted from determining the nature of mind, to the explication of intentionality and consciousness.
Essentially, functionalism has won the earlier debate, and now philosophers are trying to understand how the criticisms concerning intentionality and consciousness can be compatible with functionalism.

We will look at consciousness, and maybe intentionality, at the end of the term.
For now, we will examine the qualia-related criticisms of functionalism.

The first criticism appears in the full version of the Fodor and Block article, and is called the problem of inverted qualia.
Actually, it appears in Locke's *Essay*.
The general idea is that two people could be identical in their behavior, and indeed in their functioning, and yet not share the same phenomenal experience.

A variation of the problem arises from mere differences in physiology.
My eyes are perhaps a bit bigger or smaller than yours.
Perhaps you have more rods or cones, which are the physical basis for color perception.
Why should I believe that my sensation of red matches yours?
In fact, since I am color blind, we have no reason to believe that we have the same perceptions.
But, we don't share the same functions, either.
The problem arises for two people who do see all colors.
One person's experience might be more vibrant, or brighter, or slightly shifted to the left.

The more startling problem is how to understand the status of inverted qualia.
Take two normal sighted people, who agree on a whole range of color ascriptions.

What if every time one saw red, the other saw purple; every time one say blue, the other saw green?
They could still use the same terms; they would be functionally isomorphic.
But, they would be having different qualia.
The problem for functionalism is that if there are cases of inverted qualia, then people with the same functional states are in different mental states.
And, there seems to be no way to deny the possibility of inverted qualia.
So, functionalism fails to capture the nature of our mental states.

Further, the situation can be even worse.
David Chalmers has written in defense of property dualism from considerations of zombies.
Zombies are organisms which function just as we do, but which have no phenomenal experience.
This is also known as the problem of absent qualia.
While the possibility of zombies might seem outlandish, the privileged access we seem to have to our mental states eliminates any possibility of ruling zombies out.

Another absent qualia argument involves the Chinese nation.
This argument is due to Ned Block, but Searle anticipates it in his brain simulator reply, p 289.
The brain is essentially a collection of neurons, which discharge impulses from one to another.
Neurons fire, and induce other neurons around them either to fire or not to fire.
The story is more complicated, of course, but the differences appear only to be a matter of degree, not of kind.
The basic picture of neurons transmitting information like electrons passing along a circuit board is apt.
Imagine that we have mapped the brain, and it contains 1.3 billion neurons.
This is a fiction, but only by a factor of about a hundred - there are about a hundred billion neurons in the brain.
Now, we can set up the people of China to act as this billion-neuron brain.
We can give each person the instructions to act as a neuron does, transmit information in the way that our neurons do, to other people.
Essentially, we make a mock brain out of the Chinese nation.
The brain can be attached to a human sensory organs via radio signals from the receptor nerves.
That is, we would, according to this thought experiment, remove a person's brain, and attach an artificial processing system made out of China.

For both inverted and absent qualia cases, the functionalist seems to omit a key element of our mental lives.
But, the criticisms only apply to occurrent sensory states.