Philosophy 110W - 3: Introduction to Philosophy, Hamilton College, Fall 2007
Russell Marcus, Instructor
email: rmarcus1@hamilton.edu
website: http://thatmarcusfamily.org/philosophy/Intro_F07/Course_Home.htm
Office phone: 859-4056

Lecture Notes, November 8

I. Functionalism and theories of the mind

We have been looking at various answers to the question, "What is a mind?"
The dualist says that minds are non-physical substances.
The central problem with dualism is the problem of interaction: how does an immaterial substance interact with a physical substance?
The behaviorist says that minds are just behaviors.
The behaviorist avoids the problem of interaction, but loses any account of our internal states.
The identity theorist tries to identity minds with brains.
But, the identity theorist has three difficulties which I grouped as problems of multiple realizability: chauvinism, equipotentiality, and non-relational construal of mental states.
The functionalist tried to identify minds with the software of the brain, rather than the brain itself.
We saw that the functionalist ran into difficulties in cases of absent or inverted qualia.
For both inverted and absent qualia cases, the functionalist omits a key element of our mental lives.
Still, functionalism seems to have the best account going.
Perhaps the problem is with our understanding of qualia.
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 consciousness can be compatible with functionalism.
As Searle will show, there are also problems with intentionality.
To set up the Searle argument, we must look more carefully at the computational aspect of functionalism.
The computer analogy originates with one of the founders of computer science, Alan Turing.

II. Turing machines

A Turing machine is a computer, distilled to its essential form.
Some functionalists explain human minds in terms of Turing machines.
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 the inputs 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.

III. 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 467.
The argument from consciousness says that machines can not think, since they do not 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 do not 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 the functionalist'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 470-1, 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.
The functionalist adopts the latter position, allowing that machines can think.
The behaviorist has no ability to rule out thinking machines.
The identity theorist does not allow for thinking machines, but we saw that identity theorist was too chauvinist in just this regard.
So, can computers think?

IV. Strong and weak AI

Searle assesses the ability of models of human minds to tell us anything about actual human minds.
Searle's article can be taken both as an argument against functionalism (and behaviorism) and as an argument against artificial intelligence.
He mentions a strong AI thesis.
Elsewhere he characterizes a weak AI thesis, which 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.

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.
It 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.
To understand minds, we just need to understand computer models and their software.
Notice that Searle's characterization of strong AI is the same as our characterization of functionalism in terms of computers; the mind is the software of the brain.

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 functionalism are right, then human behavior must be describable algorithmically as well, and representable in purely syntactic form.

V. The Chinese room

Searle's Chinese room example is closely related to the qualia objections to functionalism.
Searle provides an example of a person working according to purely formal, syntactic rules, p 477.
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.

The behaviorist, who already disowns our inner lives, can claim that 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, and their relations to other mental states, matter, the ability to follow formal rules can not suffice for understanding.
Even if 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.
Searle extends the argument to robots.
Even if they are highly complex, essentially they are doing the same thing that they would be doing if I were controlling the robot from the Chinese room.
This is why Searle claims that his argument does not depend on the current state of technology.
His claim is not that the particular machines that we have now lack minds.
Such a claim would leave open the possibility that some machines will have minds.
Instead, Searle's claim is that any syntactic processor, completely describable in terms of formal processing, is necessarily not a mind.

V. Searle's argument

Searle presents his argument as follows:
1. Brains cause minds.
2. Syntax is not sufficient for semantics.
3. Computer programs are entirely defined by their formal, syntactic structure.
4. Minds have semantic contents.
C1: Computer programs are not sufficient for minds.
C2: The way that brains cause minds can not be by running a computer program.
C3-4: Anything that causes minds, including any artefact that we might make, must have causal powers at least equivalent to those of the brain.

Premise 1 and premise 4 are obvious, since we all have brains and minds, and we all process meanings.
In our case, anyway, brains cause minds, and minds understand.
Premise 1 does not assume that only brains cause minds, or that they are the only semantic processors.
The role of the Chinese room argument is to support premise 2.
Premise 3 might be taken as definitional.
We looked at Turing's article, and Turing machines, to justify premise 3.

Regarding AI, the importance of Searle's argument is that a mechanical model of the mind could not *be* a mind.
Searle says that any artefact would have to have the causal powers of the mind.
Now, C3 seems to be a non-sequitur.
1000 is greater than 10.
But, it does not follow that anything greater than 10 must be at least as great as 1000.

Still, Searle's point stands that syntax alone seems insufficient.
So, what is it about our brains, and perhaps our bodies, that allows us to understand, as well as process, information?
Searle thinks it has something to do with the way our bodies are connected to the world.

Regarding functionalism, the importance of the argument is that intentional states, not only qualia, are a problem.

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.
In other words, consciousness is essentially a biological phenomenon.
Perhaps the chauvinism of identity theory was right after all.