Philosophy 240: Symbolic Logic
Fall 2010
Mondays, Wednesdays, Fridays: 9am - 9:50am

Hamilton College
Russell Marcus
rmarcus1@hamilton.edu

Class 7 - September 10
Philosophy Friday #2: Syntax and Semantics
John Searle, "Can Computers Think?"
Gottlob Frege, Preface to *Begriffsschrift*

I. Theories of the Mind

One of the more interesting applications of logic is in artificial intelligence programming.
We have mainly been thinking of logic as a language for formally representing our beliefs and inferences.
We can also use it to program computers, and to simulate human reasoning with machines.
Such uses of logic lead us naturally to ask the question, "What is a mind?"
In particular, can machines like robots have minds?
Similarly, we could ask the question whether animals have minds.

There are four prominent theories about the nature of mind.
I will briefly mention three theories, and the major difficulties they face.
Then, we will focus on the fourth.

The dualist says that minds are non-physical substances, sometimes called souls.
My mind is somehow attached to my body, but independent of it.
Some one who thinks that the soul can live past the death of the body is a dualist.
The central problem with dualism is the problem of interaction: how does an immaterial substance interact with a physical substance?
Or, why does the mind get drunk when the body does the drinking?
That is, the dualist must explain some way for the mind to communicate with the body.
But, the communication can't be physical (since the mind is isolated from the physical world), and it can't be mental (since the physical world is isolated from the mental world).
If there are two distinct kinds of substances, it seems impossible that they would be able to communicate.

Behaviorism was the first serious materialist theory developed in the twentieth century.
The behaviorist says that minds are just behaviors.
So, to say that someone is hungry (a mental state) is just to say that she is disposed to go to Commons, or to have a snack (both behaviors).
Behaviorism was favored by psychologists.
By defining the mind as dispositions to behave, mental states became accessible to scientists, legitimate objects of serious empirical study, in contrast to the earlier speculations of Freudian psychology.

The behaviorist avoids the problem of interaction, but loses any account of our internal states.
For example, consider two chess players.
The first player makes a quick, move, without thought or hesitation.
The second player, faced with the same board, quietly and unnoticeably considers lots of different possible moves.
After, say, fifteen minutes, the second player makes the same move as the first player.
Now, both players exhibited the same resultant behavior.
But, we certainly want to claim that they did so as a result of very different mental processes.
The behaviorist has little ability to differentiate the two players.

The identity theorist identifies minds with brains: the mind is the brain.
According to identity theory, only human beings can have minds, since only human beings have brains.
Identity theory thus differentiates our two chess players on the basis of their quite different neural processes.
But, identity theory is chauvinistic.
If we met an alien made out of a different substance, with a different organization, but who behaved and interacted with us as if it were human, we would surely grant that the alien had a mind.

In response to the difficulties with these theories of the mind, many philosophers defend functionalism.
Most functionalists agree with identity theorists that there are no immaterial souls, and that there is more to our mental life than our behavior.
The functionalist identifies minds with the brain's processing, rather than the brain itself.
According to functionalism, anything that behaves like something with a mind, and that has internal processes that map onto our internal processes, has a mind.
Since computers do not have brains, the identity theorist says that they can not think.
Functionalists argue that what is important about minds is the software, rather than the hardware: the mind is the software of the brain.
They are sympathetic to the claim that machines can think, that there can be artificial intelligence.
"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

Functionalism is both consistent with a materialist view of the world and subtle enough to accommodate both internal processes (like the identity theorist) and behavioral correlates of psychological states (like the behaviorist).
Thus, it has been popular with computer scientists eager to find profound implications of their work.
Sometimes, these scientists can over-reach, as in the thermostat case attributed to John McCarthy by Searle.
McCarthy claims that a thermostat has three beliefs: that it is too cold, that it is too hot, and that it is just right.
McCarthy's use of 'belief' is clearly metaphorical, not literal, even if he did not intend it in that way.
Similarly, when my wife says that the mosquitos believe that she is tasty, we are best interpreting her words as metaphors.
On the other hand, if we say that a chimp or a dolphin has beliefs, well, we don't necessarily know whether to take it literally or metaphorically, or to what degree.
The functionalist claims that criteria for mentality are both external (and verifiable) and internal (and only susceptible to supposition).
Functionalism provides clear-enough criteria, but those criteria are, like the hypothesis that you are sentient, not amenable to external verification.
I just have no way of knowing for sure whether you are a carefully-crafted robot.
But, I know precisely what the conditions for you to be sentient are.

One of the main problems with functionalism involves the qualitative aspect of conscious experience.
I know what mangoes taste like, independently of their chemical properties.
I assume that you do too, despite my not being able to verify that you do.
It seems unlikely that a machine would experience the sweet taste of a mango in the same way that you and I experience it.
We will not explore that problem, here, though.
Searle's worry concerns a different mental property, which we can call intentionality.

II. Searle and Strong AI

Searle's article can be taken both as an argument against functionalism 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.
Weak AI is uncontroversial, except for the most enthusiastic dualists.
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, along with some plausible internal causes of that behavior.
To understand minds, according to Strong AI, we only 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.
The syntax of a program or system of formal logic concerns its form, or shape.
Our rules for wffs are syntactic.
The semantics of a system or program concerns the meanings of its terms.
When we interpret a set of propositional variables as meaning something, we are providing a semantics, as when we give a translation key for a formal argument.
It will be useful, both now and later, to make a clear distinction between syntax and semantics.

III. Syntax and Semantics

This course is centrally focused on constructing and using formal systems of logic.
Whenever we introduce a formal system of logic, we introduce two languages: an object language and a meta-language.
The object language is the language that we are studying.
The meta-language is the language we use to study the object language.
The rules for well-formed formulas are written in the meta-language, but they are about how to construct the object language.
The rules for forming wffs are syntactic.
Similarly, the rules for constructing truth tables, indeed the truth tables themselves, are written in a meta-language.

That's why we use ⊤ and ⊥, which are not symbols of our object language.
The rules for assigning truth values are semantic rules.

Whenever one constructs a formal language, one provides both a syntax and a semantics for that language.
The syntax tells how the formulas are constructed.
The semantics tells how to interpret the formulas.

Soon we will introduce inference rules for our logic.
The inference rules will be specified syntactically.
They will hold for any interpretation of the formulas, which makes them both powerful and uncontroversial.

Separating the syntax of our language from its semantics allow us to treat our formal languages as completely uninterpreted, or topic-neutral.
We can play with the symbols, according to the rules we specify, as if they were meaningless toys.
We can interpret our languages variously, comparing interpretations in order to see the properties of the language itself clearly.
Frege, indeed, was motivated specifically by the possibility of specifying a syntactic criterion for logical consequence.
As I mentioned earlier in the term, he wanted to ensure that some odd results which had arisen in mathematics in the nineteenth century were not illegitimate.
He wanted to ensure that all deductions are secure, and that we do not implicitly smuggle into our results unjustifiable interpretations.
We, like Frege, want to make sure that we do not presuppose a hidden, implicit premise.

For example, consider the first four axioms, or postulates, of Euclidean geometry.

E1. Between any two points, one can draw a straight line.
E2. Any straight line segment can be extended indefinitely, to form a straight line.
E3. Given any straight line segment, a circle can be drawn having the segment as radius and one endpoint as center.
E4. All right angles are congruent.

Euclid relied on a commonsense interpretation of these terms, especially concepts like 'straight' and 'right angle'.
Given those ordinary concepts, it seemed obvious that the parallel postulate, Euclid's fifth postulate, would also hold.

E5. If a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

E5 is equivalent to Playfair's postulate, which makes it easier to see the problem with hidden premises.

PP. Given a line, and a point not on that line, there exists a single line which passes through the point and is parallel to the given line.

In the two millennia between Euclid and the early nineteenth century, geometers tried in vain to prove E5.
They mainly did so by trying to find a contradiction by denying E5 or PP.
They supposed that there were more than one parallel line through the given point.
They supposed that there were no parallel lines through the given point.
Both suppositions led to odd kinds of spaces.
But, neither supposition led to an outright contradiction.
By the early nineteenth century, some mathematicians realized that instead of leading to contradiction, the denials of E5 and PP led to a more abstract conception of geometry, and fecund new fields of study.
Riemann, and others, explored the properties of elliptical geometries, which arise when adding the claim that there are no parallel lines through the given point to E1-E4.
Lobachevsky, Gauss, and others explored the properties of hyperbolic geometries, which arise when adding the claim that there are infinitely many parallel lines through the given point to E1-E4.
In both elliptical and hyperbolic geometries, the notions of straightness and right-angularity, among others, have to be adjusted.
Our original conceptions had been smuggled in to the study of geometry for millennia, preventing mathematicians from discovering important geometric theories.

These geometric theories eventually found important applications in physical science.
Both E5 and PP are equivalent to the claim that the sum of the angles of a triangle is 180° (pi).
Consider an interstellar triangle, formed by the light rays of three stars, whose vertices are the centers of those stars.
The sum of the angles of our interstellar triangle will be less than pi, due to the curvatures of space-time corresponding to the gravitational pull of the stars, and other large objects.
Space-time is not Euclidean, but hyperbolic.

Euclid seems to have recognized worries about the parallel postulate, since he does not invoke the fifth postulate freely; rather, he waits until he absolutely requires it.
Frege, recognizing that hidden premises had undermined mathematical progress, wanted to make sure that all branches of mathematics, indeed all of human reasoning, was not liable to similar problems.
He thus formalized the study of logical consequence, turning logic into a mathematical subject.
The preface to Frege's *Begriffsschrift*, the title of which means concept-writing, makes his motivation clear.

> So that nothing intuitive could intrude [into our concept of logical consequence] unnoticed, everything had to depend on the chain of inference being free of gaps. In striving to fulfil this requirement in the strictest way, I found an obstacle in the inadequacy of language: however cumbersome the expressions that arose, the more complicated the relations became, the less the precision was attained that my purpose demanded...The present *Begriffsschrift*...is intended to serve primarily to test in the most reliable way the validity of a chain of inference and to reveal every presupposition that tends to slip in unnoticed, so that its origin can be investigated.

By separating the syntax of logic, its formation and derivation rules, from its semantics, its interpretations and our ascriptions of truth and falsity, we are attempting to fulfil Frege's dream of a secure theory of logical consequence.

III. The Chinese room

Computers, in their most basic form, contain a complete list of possible states of the system, and possible inputs, and the output, all specifiable syntactically.
The actions of a computer 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 the machine is in such-and-such a state, and receives such-and-so input, then it does this-and-that and moves into this other state.

Computers merely follow algorithms.
Moreover, every step of the algorithm can be specified syntactically, by its inscription.
When we play a video game, we see cars and people, and hear music.
We interact with the machine on a semantic level.
But, the computer is just processing syntax, crunching 0s and 1s.
So, if strong AI and functionalism are right, then human behavior must be describable algorithmically as well, and representable in purely syntactic form, using a formal language like the one we use in logic.

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 671.
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.
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.

IV. 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.
Premise 3 might be taken as definitional.
The role of the Chinese room argument is to support premise 2.

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 in order to be a mind.
Syntax alone seems insufficient.
But, if our reasoning really proceeds according to rules of formal logic, then it would seem that we can have a purely syntactic description of our mental lives.
Who knows what 'causal powers' refers to, anyway?

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.
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.

Paper Topics

1. Is there artificial intelligence? How might the defender of strong AI respond to Searle's criticisms? See Dennett, especially.

2. What is a mind? Compare and contrast two or three theories of the mind. See Churchland's first chapter, and consider Searle's argument.

3. Is logic purely syntactic? Consider Frege's microscope analogy, from the preface to the *Begriffsschrift*, and the discussions of semantics from later in the term.

Suggested Readings

Block, Ned. "Troubles with Functionalism." In *Readings in the Philosophy of Psychology*, vol. 1, Ned Block ed., Harvard University Press, 1980. The source of the Chinese Nation thought experiment, and a sophisticated, detailed examination of the theory of mind most closely associated with artificial intelligence.

Boolos, George, John Burgess and Richard Jeffrey. *Computability and Logic*, 5th ed. Cambridge University Press, 2007. A great technical book on the logic of computation.

Churchland, Paul M. *Matter and Consciousness*, revised edition. MIT Press, 1988. The first few chapters are an excellent overview of the most important theories of the mind.

Dennett, Daniel. *Brainstorms: Philosophical Essays on Mind and Psychology*. MIT Press, 1990. See especially the first essay, "Intentional Systems," for a philosophical defense of AI.

Frege, Gottlob. *Begriffsschrift*. In van Heijenoort. The original.

Searle, John. "Can Computers Think?" In *Philosophy of Mind: Classical and Contemporary Readings*, David Chalmers ed., Oxford University Pres, 2002.

Searle, John. "Minds, Brains, and Programs." In *The Nature of Mind*, David Rosenthal ed., Oxford University Press, 1991. Includes responses and counter-responses from Jerry Fodor and Searle.