

24.119 Minds and Machines

Handout 6: Behaviorism

Motivation

The Newcomers arrive and live among us. They are humanoid in form, although their skin is tinged with green, and their inner structure is nothing like ours. In fact -- the skin aside -- as far as outward appearances go the Newcomers are astonishingly like us. In short, they behave as we behave. At any rate -- not to beg any questions -- they move their bodies and emit sounds just as we do. There are the usual social tensions, of course. A rabble-rousing human politician raises the question: what if the Newcomers have no minds at all? What if they lack, not just conscious experience, but beliefs, desires, hopes, expectations, intentions, and the rest? If we do think they have minds, aren't we just guessing? The Leader of the Newcomers speaks:

I am a Newcomer. Hath not a Newcomer eyes? Hath not a Newcomer hands, limbs, dimensions? Fed with the same food, hit with the same weapons, subject to the same irritation of our sensory surfaces, do we not behave just as humans? If you prick us, do we not cry out? If you tickle us, do we not laugh? If you ask us, do we not answer? And if you have minds, do we not also? If we are like you in the rest, we will resemble you in that.

How could the politician's suggestion possibly be correct? Note that the crucial behaviorist thought isn't simply that we have *excellent evidence* that the Newcomers have minds. Rather, it is anyone who behaves just as the Newcomers behave is *absolutely guaranteed* to have a mind (cf. Block, "[The Mind as...](#)", section 1.1, on the Turing Test).

As discussed in class, a similar motivation for behaviorism can be extracted from the Star Trek episode "[The measure of a man](#)".

Two kinds of behavior

Any instance of *physical behavior* is a physical change to a creature's body (perhaps in relation to his environment), such as the rising of the creature's arm, or the emission of certain sounds from the creature's mouth.

Any instance of *agential behavior* is something a creature does, such as raising his arm, or saying that it's time for lunch.

We can similarly define *physical behavioral dispositions* and *agential behavioral dispositions*. Dispositions are properties like fragility, solubility, and elasticity: something is fragile just in case (to a first approximation) it *would* break if struck or dropped. Note that something might be fragile even though it is never in fact struck or dropped. Similarly for a soluble sugar cube, or an elastic band.

Kinds of behaviorism

The kinds described here all take "behavior" to be *physical* behavior. When Ryle spoke of behavior, he meant *agential behavior*. So Ryle's version of behaviorism is left out (for a bit more on Ryle, see Byrne, "[Behaviourism](#)" and the REP entry [Gilbert Ryle](#)).

Eliminative (or revisionary) behaviorism

Our commonsense or folk theory of the mind is to a large extent false (much as, say, ordinary talk about witches in the seventeenth century was to a large extent false). So we should replace our folk theory of the mind by a more scientific theory couched in the language of physical behavior. The Harvard psychologist [B. F. Skinner](#) showed considerable sympathy for this view. (Note that the eliminative behaviorist will *not* find the Newcomer example persuasive: he thinks that the Newcomers do *not* have minds -- and neither do we!)

Analytical (or logical) behaviourism

Statements containing mental vocabulary can be analysed into statements containing just the vocabulary of physical behavior.

There is an important difficulty for any behaviorist analysis of the propositional attitudes (e.g. beliefs, desires, intentions). What someone does, or is disposed to do, depends not only on the fact that he holds a particular belief, but also on his desires (and, further, on his other beliefs as well). Therefore there can be no question of a simple atomistic behavioral analysis: one which matches each belief with a different kind of behavior (whether specified in the language of agential or physical behavior). A given belief may issue in practically any sort of behaviour, depending on the agent's other attitudes. In fact, the same problem seems to afflict behaviorist accounts of sensations like pain.

Although analytical behaviorism is not very plausible, there are three other behaviorist theses that have a bit of life in them yet:

Behaviour-as-necessary

Necessarily, anything that has no physical behavioral dispositions of a certain kind and complexity does not have a mental life.

Behaviour-as-sufficient

Necessarily, anything that has physical behavioral dispositions of a certain kind and complexity has a mental life.

Supervenient behaviorism

In the jargon, psychological facts "supervene" on physical behavioral dispositions: If any two possible creatures x and y differ with respect to types of mental states, then they differ with respect to types of behavioral dispositions. (See the REP entry [Supervenience](#).)

(Note that analytic behaviorism entails supervenient behaviorism, but not conversely.) Of course these theses are somewhat vague as stated: to make them more precise we need to know exactly what is supposed to count as a behavioral disposition. My fingers would twitch if a probe were inserted into a certain region of my brain. In one sense this is a behavioral disposition of mine (the disposition to have twitching fingers when my brain is probed). But surely the behaviorist doesn't mean to include *these* dispositions in the formulation of her behaviorism. That wouldn't fit with the intuitive motivation for behaviorism, that whether someone has a specific kind of inner organization (e.g. a brain) isn't relevant to whether she has a mind.

To refute the view that a certain level of behavioral dispositions is necessary for a mental life, we need convincing cases of thinking stones, or utterly incurable paralytics, or disembodied minds.

To refute the view that a certain level of behavioral dispositions is sufficient for a mental life, we need convincing cases of rich behaviour with no accompanying mental states. One possible example is Block's Aunt Bubbles machine, adapted so that it responds behaviorally as well as verbally.

Such an example would work equally well against supervenient behaviorism (see Block's [note 1](#)). But supervenient behaviorism could be refuted by something less ambitious. Creatures who are in intense pain but do not betray this in their verbal or non-verbal behavior, behaving just as pain-free human beings, would be the right sort of case. But even if this works for pain, an analogous story to refute supervenient behaviorism restricted to the propositional attitudes will be less intuitively convincing. (Exercise: why?)
