

A summary of the first half of the course

October 30, 2000

1 The questions

What is it to believe that snow is white, to want to be famous, to be in pain, to be angry, to be intelligent, to be conscious? In other words, what is the nature of the properties of believing that snow is white, wanting to be famous, being in pain, etc.?

2 Dualism

What is common to **dualists** is the view that some mental properties are properties no *entirely physical* thing could have. According to a **substance dualist**, certain mental properties that people have are properties that nothing that had any physical properties could have. Thus there are two different sorts of thing that do not overlap: things that have mental properties, and physical things. A **property dualist** holds a weaker view: even if a given mental property is *compatible* with physical properties, it is quite separate from the physical properties of thing things that have it: a thing's physical nature doesn't determine whether it has that mental property.

An influential argument for substance dualism is the argument from the possibility of disembodiment, which we considered in the second lecture. (The handout from that class is on the web, so I won't discuss it further here.)

You could deny dualism even if you believed in ghosts and all kinds of mental life having nothing to do with the physical world. But people are typically motivated to deny dualism by their belief in **materialism**, also known as **physicalism**. Materialism says that there is nothing more to the nature of the world than its physical nature—or at least, nothing that is relevant to the philosophy of mind. We can capture this claim as a claim about **supervenience**: any possible world that is a minimal-physical duplicate of the actual world is just like the actual world in all respects—or at least, in all mental respects.

There are many arguments for materialism and against dualism. One of them is the **causal argument**, discussed by BMJ (pp. 6–13). Another is the **problem of other**

minds: it is suggested that if mental properties were independent of physical properties, we could not be justified (as we in fact are) in drawing conclusions about a person's mental life on the basis of evidence about the physical state of their body. There is also a widespread view that materialism is the natural or default position, so that the burden of proof is on the opponents of materialism to show that it is false.

Behaviourism, the Identity Theory, and Functionalism are all compatible with materialism, and typically supported by materialists.

3 Behaviorism

A **behavioral disposition** is a property that says how a thing *would* behave if it were put in such-and-such circumstances. For example, a glass has the disposition to break when struck. A **behaviorist** (about a given mental property) says that the property is just a matter of one's behavioral dispositions. A very simple-minded sort of behaviourism says that the mental property just *is* some particular behavioral disposition: e.g., that pain is the disposition to scream (in more or less any circumstances). This is very implausible. A better sort of behaviourism allows that you might need a whole lot of facts about a thing's behavioural dispositions in order to determine any mental property. Even this sort of behaviourist must admit that mental properties are properties which **supervene** on behavioural dispositions: that is, any two possible things that were alike in all their behavioural dispositions are alike in their mental properties.

One argument that has been powerful in motivating behaviourism is an attempt to generalise the 'problem of other minds' argument against dualism. The key premise of this argument is that it is only if mental states are a matter of behavioral dispositions that we can explain why the method we actually use to find out about other peoples' mental properties—observing their behavior in response to various stimuli—should be justified. Dennett seems to be giving a version of this argument in support of a mod-

ified form of behaviorism about beliefs. (Dennett's view apparently entails that what someone believes supervenes on the facts about their behavioural dispositions *and their history*.) Dennett supports the key premise by giving a detailed description of the complex, intricate method we use to find out about other peoples' beliefs (and desires), the 'intentional strategy': the challenge to the non-behaviorist is to explain why this very specific method should be so effective.

The simplest arguments against behaviorism are those turn on *counterexamples* to the behaviorist's supervenience claim: (what seem to be) possible things that are alike in behavioural dispositions but different in mental properties. In class, we discussed **incurable paralytics** (who surely could have mental lives even though their behavioral dispositions are the same as those of the brain-dead), **Martian Marionettes** (BMJ p. 150), **Blockhead** (BMJ pp. 111–120) and Putnam's **X-worlders**.

We considered several other arguments against behaviorism. The **first-person objection** could be taken as asserting that behaviorism can't explain the reliability of introspection (this is false: behaviorism is compatible with the view that we can never go wrong in forming beliefs about our own mental states), or it could be the bald assertion that we can tell just by introspecting them that our experiences, sensations, etc. aren't a matter of behavioral dispositions. The **causal objection** says that behaviorism is incompatible with the allegedly obvious claims that mental properties cause behavioral dispositions, or that mental properties cause behavior. I claimed that what BMJ call the **match-up problem** is really just a special case of the **problem of multiple interpretations**: even given the truth of belief-desire psychology, many different stories about someone's beliefs and desires are compatible with the facts about their behavioral dispositions. Dennett makes an interesting move in solving this problem: the idea is that *natural* interpretations—ones that assign true beliefs and ordinary desires—are better than unnatural ones, and are automatically true unless they are defeated by the facts about behavioral dispositions. BMJ discuss different versions of this idea under the names 'the principle of charity' and 'the principle of humanity' (pp. 147–149).

4 The Identity Theory

To hold the so-called **identity theory** with regard to a mental property M is to think that M is a *brain property*:

a property which an organism has in virtue of the physiological state of its brain.

Smart and Lewis both advocate the view that *pain*, and other *sensations* and *experiences*, are brain properties. But Lewis does not hold that properties like the properties of *being in pain*, *believing that snow is white*, *having an experience as of something yellowish-orange*, etc.—the properties we are primarily concerned with in this course—are brain properties. In Smart's case, it is hard to tell whether he holds the identity theory about these sorts of properties.

One might argue for the (full version of the) identity theory by emphasising the analogy between mental predicates (like 'is in pain') and natural kind predicates (like 'is made of water'). In the case of natural kind predicates, what I called the "standard view" says that relevant parts of empirical science can tell us the nature of the properties expressed by those predicates. For example, chemistry is supposed to have told us that the property of being made of water is the property of being composed of H₂O molecules. If there are other planets, or other possible worlds, where some other kind of stuff plays the "water role", it still isn't water. The identity theorist can at least ask why we shouldn't expect the same kind of thing to happen with mental predicates.

Kripke presents an influential argument against the identity theory. The main point of the argument is just this: According to the identity theorist it is necessary that the people who are in pain are exactly those whose brains are in a certain state. But this doesn't *seem* to be necessary: we seem to be able to imagine possible situations in which people are in pain without being in that brain state, or vice versa. Those who would claim that this is a misleading appearance need to *explain it away* somehow: Kripke claims that this can't be done, and concludes that the identity theory is false. If Kripke's argument is any good, it can be generalised into an argument against *physicalism*, not just the identity theory: for according to any physicalist the **zombie world**—a world just like this one physically, but lacking any mental life—is impossible, even though we seem to be able to conceive such a situation.

The most important argument against the identity theory is the **multiple realizability argument** (Putnam, p. 201; BMJ, pp. 42–45). If *being in pain* is a brain property, however, it is necessary that all and only creatures with a brain that is in a certain physiological state are in pain. But it just seems possible that brainless creatures,

or creatures whose brains worked in very different ways from ours, could be in pain and indeed have mental lives just as rich as ours. So *being in pain* is not a brain property, and neither are any other such properties. If there are actually any creatures that are in pain without having brains, then on the plausible assumption that *pain* is a property which *in the actual world* is had by exactly those things that are in pain, *pain* cannot be a brain property either.

5 Functionalism

In general, a **functionalist** about a mental property *M* holds that to have *M* is to have the *n*th one of a sequence of properties that in a certain population have such-and-such pattern of causal relations to each other and to certain *input* and *output* properties.

For our purposes, the most important sort of functionalism is **common sense functionalism**. The common sense functionalist relies on common sense about mental matters—sometimes called **folk psychology**—to specify the relevant pattern of causal relations, the inputs, and the outputs. The **Ramsey-Lewis method** is a general method for explaining the meanings of **theoretical terms**, i.e. expressions which get their meaning from the theories in which they occur. Provided that folk psychology is entirely about the causal connections between one’s perceptions, one’s mental states, and one’s behaviour, the theory one gets by applying the Ramsey-Lewis method to folk psychology to come up with definitions of psychological predicates will be a version of functionalism.

Common sense functionalism can answer the behaviorist’s challenge to explain why our actual methods of finding out about mental properties should be good ones. If believing that snow is white is being in a state which plays a functional role partly defined in terms of the characteristic behavior that state tends to cause, it’s no mystery that we can reliably tell whether people believe that snow is white by observing their behavior. In fact the common sense functionalist can claim to improve on the behaviorist in this regard, since the common sense functionalist can also explain why various non-standard methods—e.g. looking at someone’s brain—should be good ways of finding out about people’s mental properties in abnormal cases. This advantage isn’t limited to the common sense functionalist: it can also be claimed by the (across-the-board) identity theorist, and by the empirical functionalist (see below), provided that they give folk psychology a role

in *fixing the reference* of words for mental properties.

How does a common sense functionalist specify the *population* in which the properties are supposed to play the functional roles? An incautious reading of Lewis’s ‘Psychophysical and Theoretical Identifications’ would lead one to conclude that the population must consist of all the things that exist at the same possible world as the thing whose mental properties we are discussing. This view is defeated by a version of the multiple realizability argument. In Lewis’s ‘Mad Pain and Martian Pain’, the population is specified as the “appropriate” population, where what counts as appropriate depends on certain context-sensitive, pragmatic factors. BMJ present a version of common sense functionalism on which the “population” is just the single individual to which we are attributing mental properties: a creature is in pain just in case it is in the relevant member of a sequence of properties which play the common sense functional roles *in it*. This ‘**individualist**’ approach (my phrase) is standard, and often presupposed in accounts of functionalism.

Common sense functionalism is also known as **analytic** or **a priori** functionalism, because it is supposed to be something one is in a position to know provided one grasps the *meaning* of mental predicates. The other leading sorts of functionalism are *empirical* functionalisms of one sort or another. These views are very like the identity theory, except that instead of giving the role of telling us about the nature of mental properties to the science of brain physiology, they give it to some more abstract science, like cognitive psychology or the science of the structures of neural networks. Empirical functionalism is vulnerable to a new version of the multiple realizability argument, a.k.a. the **chauvinism** objection: It seems possible for there to be beings which have rich mental lives despite being quite unlike us as regards the sorts of structure studied by cognitive psychology, etc.

We looked at two general arguments against all kinds of functionalism, Ned Block’s **Chinese Nation argument** and John Searle’s **Chinese Room argument**. Each of these purports to describe a system which is just like an ordinary person at the level of functional organisation, but which nevertheless lacks some or all of the mental properties of an ordinary person. Most functionalists have been happy just to “bite the bullet”, i.e. to accept that systems of that sort could have mental properties, provided that they were set up correctly.