

A summary of the second half of the course

December 11, 2000

1 Theme of the second half of the course

The first half of the course was devoted to laying out some general views about the nature of mental properties of all kinds. The second half addressed questions at a more specific level about the natures of two important sorts of mental properties: **contentful** properties or **propositional attitudes**, like the property of believing that snow is white or the property of desiring to make money; and **phenomenal** properties, like the property of experiencing red and the property of being in pain.

2 Broad and narrow content

The first question about propositional attitudes is whether they are **broad** (a.k.a. ‘wide’) or **narrow** properties. Narrow properties are those that are necessarily shared by **doppelgangers**: things that are exactly alike in how they are internally and governed by the same laws of nature. (In practice we can treat a brain in a vat that is just like my brain as one of my doppelgangers, since it’s clear that the brain is the part of the body that really matters as far as mental properties go.) Broad properties are just those that are not narrow.

There’s a strong intuitive case for the view that contentful properties are narrow: it just seems obvious that a brain in a vat that was just like your brain would have all the same beliefs and desires as you have. However, it’s not hard to see that for many values of ‘*p*’, the property of believing (or desiring) that *p* is broad. For example, I believe that New York is a large city, whereas doppelgangers of mine who live on other planets don’t have any beliefs about New York. Again, I believe that water is wet but my doppelgangers who live on worlds where there is no water don’t believe this. **Deflationists about broad content** agree, but claim that there is still something right about the intuition that propositional attitudes are narrow. (**Scepticism about narrow content** is just the denial of deflationism about broad content.) There are various different ways of working out the deflation-

ist view: the one I presented makes use of an idea introduced by **Russell**. The idea is that when I have beliefs and desires about things and properties out in the world, I think about them *by description*. I count, for example, as believing *that Bismarck was famous in virtue of the fact that I believe that the most famous nineteenth-century chancellor of Germany was famous*, and the fact that Bismarck was the famous nineteenth-century chancellor of Germany. I count as believing that, in turn, because I believe something even more complicated in which descriptions are substituted for the names ‘nineteenth century’ and ‘Germany’. According to Russell, when we keep doing this we can eventually get back to a most basic set of beliefs, my **explicit** beliefs or **judgments**: these are beliefs I don’t have in virtue of believing anything else. (Similarly for desires.) Deflationism about narrow content, as I presented it, says that the property of *explicitly* believing some proposition is always a narrow property. I believe that New York is a large city and my doppelgangers don’t, but we all make the same *judgment*, e.g. that the city which we live in and which is called ‘New York’ is a large city. Because New York is in fact the city I live in which is called ‘New York’, I get to count because of this as believing that New York is a large city; my doppelgangers don’t.

The **shortest spy** problem is a problem for the simplest Russell-style model of what it takes to have beliefs about external things. If Vladimir is the shortest spy, intuitively you don’t believe that he is a spy even if you believe that the shortest spy is a spy. This problem can be blocked either by restricting the class of descriptions which grant thoughts about external things, or by the object of thought not only to satisfy the relevant description but also to stand in certain other relations to the thinker.

Burge argues that for a very wide range of values of ‘*p*’, the property of believing that *p* is broad, on the grounds that a doppelganger of someone who believed that *p* who lived in a linguistic community in which some of the words that occur in the sentence ‘*p*’ are used differently. He doesn’t explicitly address deflationism about

broad content, but it's pretty clear that he thinks that the phenomenon of broadness is deeper than a deflationist could allow. The standard deflationist response to Burge is to claim that the examples only show that we often believe what they believe—e.g. that arthritis is a disease—in virtue of having some descriptive belief which picks out the thing that it's about by means of a linguistic description, e.g. the description 'the disease that people in my community call 'arthritis'.

3 What determines content?

The next part of the course was devoted to a different sort of question about propositional attitudes. According to functionalism, we get to have beliefs and desires because our brains have physical properties that play certain roles. The question is: which of the physical aspects of and relations among the states of human beings' brains matter in determining what they believe and desire? This question is partly empirical but may also be partly conceptual. (If analytic functionalism is true, it is a conceptual question whether certain physical facts would play a role in determining content *if they obtained*, even though it is an empirical question whether facts of that sort actually obtain in human beings.)

The first question, discussed by Fodor, is: is mental representation (in human beings) **structured** or **unstructured**? Do structural relations among brain states—e.g. the fact that brain state A is like brain state B as far as one bit of the brain is concerned, and like brain state C as far as another bit of the brain is concerned—matter for determining content? Or is it only the **causal** relations among the states that matter? Fodor defends the view that structural relations do matter. BMJ discuss a further question that arises once we have decided that mental representation is structured: does the brain represent in a **map-like** or a **sentence-like** way? Unlike a collection of sentences, a map can represent many facts as obtaining without having minimal parts that represent one fact each. If mental representation is sentence-like, it might make sense to suppose that someone could change their beliefs about just one thing while leaving all their other beliefs unchanged; there's no clear sense to be made of this possibility on the map-like view.

No-one thinks that structural relations among one's brain states are *sufficient* to determine what one believes and desires: it also matters how a brain state is linked up causally to other brain states, sensory inputs and behav-

ioral outputs. But there is disagreement over which of these causal relations are important, and how the causal relations in question fix the content of thought. BMJ adopt a view in which connections to *behavior* are paramount, with content being fixed by the principles of belief-desire psychology. By contrast, many philosophers adopt some form of **information semantics**, in which the typical *causes* of a brain state in the external world determine its content. A brain state that is typically caused by cats will mean 'cat'. A difficulty for this view is the **disjunction problem**: since we typically are misled by things that look like cats but aren't, why not interpret that brain state as meaning 'cat, or thing that looks like a cat'? Proponents of the view answer this challenge in many different ways, some of which are surveyed by BMJ.

One variant of information semantics is **teleological semantics**, according which what determines the meaning of a brain state is not what it actually tends to be caused by, but what it *ought* to be caused by (if it is functioning properly); this is normally understood in evolutionary terms. Teleological semantics gives very counterintuitive results when applied to creatures who didn't evolve in the usual way—for example, **Swampman**, who is like an ordinary person but was formed by a completely random process. Intuitively Swampman can have lots of beliefs and desires: teleological semantics must, it seems, disagree.

These views about how content is determined by no means exhaust all the possibilities. In class, I defended a mixed view according to which sensory inputs and behavioral outputs both play a significant role in determining mental content.

4 The knowledge argument

This influential argument purports to show that *physicalism* is false, and to establish that in fact *property dualism* is the true theory of the natures of *phenomenal* mental properties. (See the previous handout for definitions of these terms.) You know the story of Mary who is confined to a black-and white room but gets to learn all about physics, neuroscience, etc. There are different things you might want to say about Mary before leaving the room. In roughly increasing order of strength, you might claim that...

1. There is some knowledge Mary doesn't have
2. There is some information Mary doesn't have

3. Mary doesn't know that p , for some p : e.g. that THIS is the experience people normally have when they look at red things (or that THIS is what the experience of red is like, etc.)
4. There are some properties such that Mary doesn't know which things have them
5. Mary doesn't eliminate all possible worlds different from the actual world as regards phenomenal properties

The step from each step to the next one seems slight. But 5, together with the premise that Mary *does* eliminate every possible world different from the actual world in *physical* respects when she is in the room, implies that there is a possible world different from the actual world in mental but not physical respects, i.e. that physicalism is false.

Lewis and Loar both try to defend physicalism by blocking the knowledge argument. **Lewis** denies 2–5, but accepts 1. According to Lewis, what Mary gains is an **ability**, i.e. *knowledge how*—the ability to imagine and remember experiences of a certain sort, and to recognise them when they recur.

Loar denies 4 and 5, but accepts 1–3. According to Loar, before leaving the room Mary didn't know that people had THIS sort of experience when looking at red things, and hence lacked some information. The proponent of the knowledge argument points out if physicalism is true, 'THIS sort of experience' refers to some physical property P , and that Mary already knew that people looking at red things have P . It seems to follow that if physicalism is true then Mary *did* know that people had THIS sort of experience when looking at red things, contradicting Loar's claim. Loar blocks the inference by claiming that the phrases ' P ' and 'THIS sort of experience' refer to the same property but are associated with different **concepts** of that property—i.e. different ways of thinking about it, different words for it in the language of thought. While she is in the room, Mary can think about the property, but only using a 'physical' concept of it. So it's true to say that she knows that people looking at red things have P , but false to say that she knows that people looking at red things have THIS sort of experience. For the latter claim to be true she'd have to have a **phenomenal** concept of the property, and that's something she won't acquire until she experiences red for herself. Physical and phenomenal concepts are **conceptually independent**: no amount of a priori reflection will settle what you should

believe using phenomenal concepts, given the facts about what you believe using physical concepts.

5 Zombies

A **zombie** is something exactly like an ordinary person in all physical respects, but lacking consciousness, i.e. lacking phenomenal properties. A **zombie world** is a whole possible world physically just like the actual world but where nothing has any phenomenal properties. If physicalism is true, there are no zombie worlds. **Chalmers** argues against physicalism by arguing for the claim that a zombie world is possible. The premises of his argument are (1) that a zombie world is **conceivable**, and (2) that if a zombie world is conceivable, it is possible. (A scenario is conceivable if it cannot be shown to be contradictory by a priori reflection alone.) Chalmers' argument for the first premise is just that there doesn't *seem* to be anything implicitly contradictory about the supposition of a zombie world. His argument for the second premise turns on a theory about the cases in which the inference from 'it is conceivable that p ' to 'it is possible that p ' fails: e.g. 'it is conceivable that water isn't H_2O ' doesn't entail 'it is possible that water isn't H_2O '. The best way to understand Chalmers' theory is in terms of the notion of explicit thought that was introduced in section 2. Chalmers claims, in effect, that when I conceive of water not being H_2O this isn't one of my *explicit thoughts*. What I'm explicitly conceiving of is something like the proposition that the transparent liquid that flows in rivers and falls from the sky as rain isn't H_2O . It's in virtue of my conceiving this, and the fact that water is in fact the liquid that plays that role, that I count as conceiving of the proposition that water isn't H_2O . Chalmers concludes that while such examples do show that conceivability doesn't entail possibility in general, they don't refute the independently plausible principle that when a proposition is *explicitly* conceivable it must be possible. But if I conceived of a zombie world in full physical detail, I would be conceiving explicitly of a world physically like ours in all respects, but mentally unlike it—in fact, I would be explicitly conceiving of the zombie world. So, if Chalmers' principle is true, a zombie world must be possible.

Physicalists respond to Chalmers' argument by denying one of his premises. Analytic functionalists (and behaviorists) deny the first premise: according to them there is an implicit contradiction in the claim that a creature just like me in physical and hence in functional respects could

be unlike me mentally. Others deny the second premise, and hence must deny Chalmers' claim that explicit conceivability entails possibility. Many do this on the basis of a view somewhat like Loar's: our explicit thoughts can involve concepts which refer to the same property even though they are conceptually independent. Thus it can happen that we explicitly think a thought that is impossible, attributing a property to something using one concept while denying it using another concept, in such a way that the impossibility can't be discerned by a priori reflection.

6 Intentionalism

Intentionalism is the view that phenomenal properties just are propositional attitudes of a certain kind. In particular, the phenomenal character of someone's visual experience is just a matter of what propositions **visually appear** to them to be the case; hence, two people who are alike in what visually appears to them to be the case must be alike as far as the phenomenal properties associated with visual experience are concerned. **Harman** argues for intentionalism by means of the argument from **transparency**: if there was something more to the phenomenal character of one's visual experience than what visually appeared to be the case, one could attend to that; but in fact, all one can attend to in one's visual experience are the features external objects are represented as having; so intentionalism is true.

One sort of objection to intentionalism is the argument from **aberrant experiences**: double vision, unfocused vision, afterimaging, etc. An intentionalist must account for the difference between, e.g., blurred vision and normal vision as a difference in what visually appears to be the case. But certain sorts of experience are very difficult to characterise in this way.

The most influential argument against intentionalism is the **inverted spectrum** argument. The idea is to come up with an intuitively convincing case in which two people have inverted visual phenomenal properties—the phenomenal state one of them is in when looking at a ripe tomato is the same as the one the other is in when looking at an unripe tomato, etc.—but the same visual contents. The difference in phenomenal properties is established by a difference between their visual systems, such as the presence of 'inverting lenses' in the eyes of one of them, which clearly makes a difference to phenomenal properties when we apply it to any one person. The sameness of what visually appears to be the case is typi-

cally argued for on the grounds that in the usual examples (i) both people function as normal members of the same **linguistic community**, sincerely saying the same things in the same circumstances; (ii) both perform with equal **success** in the world, making it implausible that one has systematically false visual contents; (iii) the difference between them is symmetric, so that it would be **arbitrary** to ascribe true contents to one and false contents to the other, and (iv) both are assigned the same contents by some plausible **functionalist theory of content** in which the typical causes and/or effects of a given person's brain states determine that person's propositional attitudes. The intentionalist can respond to any given purported counterexample either by denying that the phenomenal properties are really different, or by denying that the visual contents are the same.

Another argument against intentionalism relies on the fact that phenomenal properties seem to be *narrow*, whereas contentful properties—at least, the ones which are claimed by most intentionalists to matter as far as phenomenal properties are concerned—are broad. Hence the former properties can't be the same as the latter ones. If you vary the world outside of a person but leave them unchanged, you can change what visually appears to be the case to them, but not their phenomenal properties. **Block's Inverted Earth** argument can be regarded as an example of this: whisking me off to Inverted Earth while putting inverting lenses in my eyes won't affect my phenomenal properties (since they are narrow), but will, in the long run at least, affect what visually appears to be the case to me. The intentionalist can respond to this argument either by denying that phenomenal properties are narrow, or by denying that the propositional attitudes that constitute the phenomenal aspect of experience are broad.