Seminar two
CUHK
10 March 2006

Frank Jackson

What we covered in seminar one

- Kripke's attack on the DTR for proper names, his three main arguments and some of the standard replies
- Putnam's case for terms like 'water' and 'gold' naming the kind, not whatever has the folk markers and how this via the remote place version of Twin Earth, leads in his view to the conclusion that meanings aren't in the head, and how this led many to the further conclusion that mental content as well as linguistic content or meaning is broad – lead that is to externalism.
- We did not discuss ways one might avoid Putnam's conclusion; we'll do that today or more likely next time.
- We made a start on the representationalist approach to language, the way it illuminates the role of language as a putative source of information about how things are, the treatment of content as a set of possible worlds, and the need to find the right representation relation for language.

What we'll do today

- Spell out how to think about the representation relation a bit more and which one is the representation relation for language
- Note some lessons in the philosophy of mind from this
- Distinguish three questions about representation and reference as a preliminary to identifying the issue about the reference of names
- Address the question of finding the right possible worlds to capture the representational content of a given sentence

Representation and correctness relative to a mapping

- Any state that represents has correctness conditions: *S* is correct iff things are as *S* represents them to be.
- In the possible worlds scheme, this comes to being correct iff the content contains the actual world.
- However, any state that represents can be thought of as having indefinitely many different contents: a petrol gauge represents all of: the level of gas in the tank, the distance before needing to refuel, the state of the wire connecting the gauge and the tank, etc.
- We need, therefore, to think in terms of how *S* represents relative to one or another mapping.

How it looks for belief, desire and perception

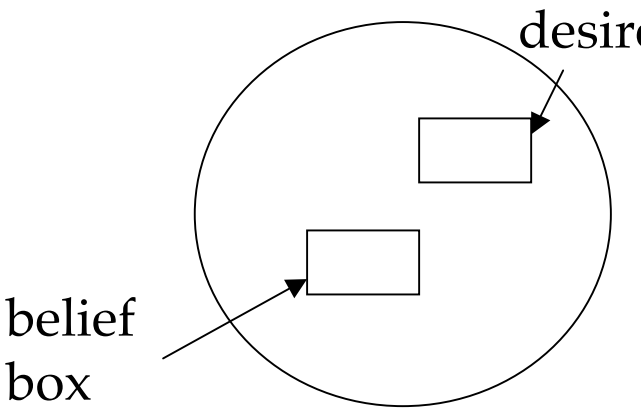
- Belief, desire and perception are all representational states: they represent how a subject takes, wants, perceives things to be, respectively.
- What it takes for a belief to be correct (true)
  - The (neural) state belief-maps onto a state of affairs that obtains
- What it takes for a desire to be correct (satisfied)
  - The (neural) state desire-maps onto a state of affairs that obtains
- What it takes for a perception to be correct (veridical)
  - The (neural) state perception-maps onto a state of affairs that obtains

Some lessons I

- Don't say there is one representation relation and the difference between belief, desire and perception is what stands in that relation – a belief, a desire or a perception. Causal co-variance is a representation relation and so any belief token will represent its ambient temperature, but very few beliefs are about that. Rather, distinguish the belief-representation relation from the desire-representation one etc.
- This means the famous belief and desire box approach to the mind may turn out to be wrong: it may be that there is one box that belief-maps onto one way things might be and desire maps into another way things might be.

Some lessons II

- Belief-box, desire-box picture



body stops moving when the information in belief box matches that in desire box

- Alternative picture: one box linked via different mappings to what's believed and desired, and the box is so connected to the body that movement stops when its state gets to the 'target' state but without there being a separate target state beforehand.
- Choice between these pictures one for neuroscience.

A bit of flesh on the bones

- What might the belief-representation relation look like?
- One way into this question is via belief-desire psychology, the idea that the core fact about belief and desire is that we move our bodies in such a way as to satisfy our desires if our beliefs are true.
- *N* belief-represents *S* iff i) *N* is a state designed to fit the facts, ii) *N* moves its host in such a way that the host's desires are satisfied if *S* obtains
- Better: take degrees of belief and desire into account. making sense of behaviour in terms of belief and desire requires attention to strength of belief and desire – otherwise could not explain buying a lottery ticket or the phenomenon of playing very safe when a lot is at risk.

The relation we need for language I

- What is the relation of language-representing?
- Language represents via its connection with the content of thought
- Locke, Grice, Bennett, Lewis view: language is a learnt, convention-governed way of communicating what we believe.
- This goes against views that *contrast* use theories of sentence meaning with representational ones: use is a key part of how a sentence gets to have its representational content.

The relation we need for language II

- Roughly, declarative *S* represents *T* in *L* iff the convention in *L* is to use *S* iff the user has a belief that represents *T*. (Very roughly, Lewis, *Convention)*
- The connection with understanding and communication:
  - you understand a language to the extent that you know the putative beliefs corresponding to the sentences of the language
  - what is offered for communication, the putative information up for transfer, are the grasped belief contents

Three questions about representational content need to be distinguished

- How a brain state gets to be a belief that represents that *p* – a matter of how the state relates to other states and the world.
- How a sentence gets to have the content it does – a matter of entering into usage conventions to express such and such a belief using so and so a sentence.
- What content some given sentence in fact has.
- The last question is our main concern; we want to know e.g. the role proper names and kind names play in representing how things are. Are Kripke and Putnam right or are they wrong?

The importance of the connection between belief and what some bit of language represents about how things are

- The connection tells us that what we say about how something is using the predicate 'is red' is how we believe it to be when we use 'is red' to describe it.
- What we say about how something is when we say it is red is that it is red.
- What we believe about how something is when we say it is red $\neq$ its having reflectance *R*.
- Therefore, being red $\neq$ having reflectance *R*
- We have a refutation of the identity theory of colour.

Second example of the same general kind

- pain is the property we ascribe with the predicate 'is in pain'
- What we believe about how someone is when we use 'is in pain' to describe them ≠ that they are in brain state *B*
- Therefore, pain ≠ brain state *B*
- We've obtained the conclusion often reached by appeal to multiple realisability without fussing about 'fools' pain.

What's the fuss about fools pain?

- Most materialists are sympathetic to the idea that mental states might be differently realised in different subjects. Perhaps neuroscience will discover that when humans are in pain, brain state *H* plays the key role of mediating between injury and protective response but in dogs it is brain state *D*. In the jargon this is the discovery that pain is multiply realisable in the way that amplification is (different amplifiers work different ways).
- These materialists conclude that pain ≠ a kind of brain state: pain is the state in common between dogs and humans in pain, there is no brain state in common; therefore, pain ≠ a brain state.
- Some materialists now resist this argument by holding that our pain is the real pain, the rest is fools pain – that is, in the imagined example, dogs aren't really in pain. Our argument finesses this debate.

Let's look at the colour and pain examples in more detail

- There are two compelling intuitions about red:
  - i) that it is the property we ascribe with 'is red' (in English) = the property we take something to have when we say it looks red = the property looking red represents it to have,
  - ii) that it is the property that we respond to when things look red.
- These two intuitions lead to trouble as on the next slide.

The colour (color) conundrum

- 1. Red = the property looking red represents something to have (obvious truth)
  2. the property looking red represents something to have ≠ reflectance *R* (or any other physical-optical property) (obvious truth)
  ∴ red ≠ any physical-optical property (valid conclusion)
- Red = the property we respond to when something looks red (obvious truth)
  the property we respond to when something looks red = a physical-optical property (a reflectance property say many colour scientists) (discovery of science)
  ∴ red = a physical-optical property (valid conclusion)
- Way out – homework!

The mental state conundrum

- 1. Mental state M = the property we ascribe in mental language (obvious truth – alternative is to say that we don't what we are saying about how things are; why bother saying anything in that case)
  2. the property we ascribe in mental language $\neq$ brain state B (for any B) (obvious truth)
  $\therefore M \neq B$ (valid conclusion)
- 1. M = the property that plays some distinctive causal role (obvious truth about mental states captured in functionalist theories)
  2. the property that plays some distinctive causal role = brain state B (discovery of science)
  $\therefore M = B$ (valid conclusion)
- Way out – next slide

Two properties instead of one I

- Often we are in the following situation:
  - there's a property Q which has such and such a property M but we don't know what Q itself is. (Maybe we'll learn what it is later.)
  - We use some word W to say that x has the property that typically M-presents.
  - The word refers to Q in some good sense.
- Some examples: words for diseases, genes, poisons, chemical properties like being acid and water.
- We respect 'we know what we are saying' by holding that in using 'W', we ascribe having the property that has M.
- This allows that we don't know what Q is; it is 'hidden'.

Two properties instead of one II

- Can think of mental state terms as like this. To be in pain is to have the property that does thus and so. To believe that so and so is to be in a state that seeks to fit the facts and that moves subjects in ways that tend to realise their wants if so and so is the case. And so on. There are two properties: the one we know about and ascribe with mental state terms is having a property that does so and so; the one we don't know about is the property that the property that fills the role.
- being in pain = having the property that plays the pain-role. This is the property we know about and ascribe with the word 'pain' and is the property that isn't $\neq$ any brain state and is the property all creatures in pain share.
- pain = brain *B* – or whatever science tells us plays the pain-role. It is the property we folk don't know about, that does the causing, and may differ from one creature in pain to another.

Trouble for teleonomy

- We've just seen how to defend the view that mental states are brain states, the famous identity theory of Smart, Armstrong and Lewis, by the two property strategy. Can we do the same for teleological theories of content? No – let's see why.
- David Papineau's early account: belief that *P* is being selected to co-vary with *P*, and desire for P is being selected to bring about *P*.
- The transparency problem for teleological theories of content – when we use '*x* believes that *P*' to describe *x* we are not expressing our belief that *x* is in a state selected to co-vary with *P*; we are not ascribing being in a certain selectional state. We *qua* the folk have never heard of selectional theories and may well lack the very concept of what it is to be selected for.

Why the two property strategy fails for teleonomy

- The two property strategy would be to offer
  - a) believing that *P* = having the property that is so and so
  - b) belief that *P* = the property that is so and so
- The claim is then that a) is something we plausibly do ascribe with the language of belief (*mutatis mutandis* for desire) – for suitable so and so, whereas b) is the 'hidden' selectional property.
- But this is only a teleological theory if the property that is so and so is a selectional property, and if it is, a) is not transparent.It is plausible that we folk know that subjects have inside them properties that play certain roles; it is not plausible that we folk know that subjects have inside them selectional properties.

Break!

Finding the right worlds to capture representational content

- We said that how a sentence represents things to be is given by a set of possible worlds – worlds where things are as they are being represented to be. How do we chose the right worlds for a given sentence?
- Our earlier discussion about the right representation relation for language tells us the set should be those that give the content of the belief about how things are that the sentence expresses.
- But how does this connect with the worlds at which the sentence is true?

Finding the right worlds

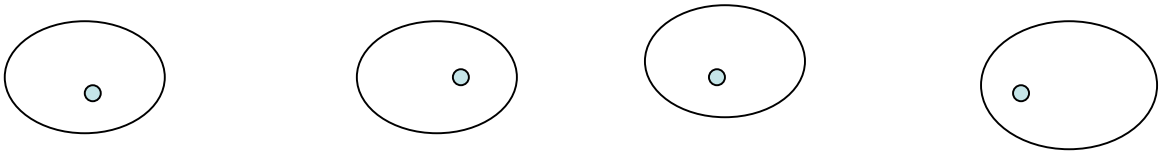
- Is the way *S* represents things to be the worlds at which *S* is true? Sometimes but not always.
- The examples that make trouble are ones where rigidification enters the picture. The effect of rigidification is to induce an illusory richness into the content of belief.
- Suppose I experience a number of robberies over time. I'm sure the same person is responsible but I have no idea who it is. I name the person 'Fred'. I say 'It's Fred again' on returning from holidays.
- What I'm claiming about how things are is that it is the same person again, and that's all. Words don't make beliefs. (Remember the tie between language-represents and what's believed.)
- But the worlds at which 'It's Fred again' is true are those where it is the person actually responsible.

More on how truth at worlds sometimes gets the content wrong

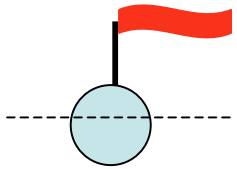
- Suppose Frank Jackson produces the sentence 'I am bearded' at 0300 to express how he takes things to be. The worlds where the sentence is true are those where Frank Jackson is bearded at 0300. But I may have no idea of who I am – I may be an amnesiac – or when it is – my watch is broken, I'm in a state of total confusion or whatever.
- Moral: going for the worlds at which the sentence is true makes the belief expressed by the sentence far too rich.
- But that's only part of the trouble: we need centred worlds, not worlds.

Why we need centred worlds to model the content

- When I produce 'I have a beard', I make a claim about the kind of world I am in, but in addition I claim that I am one of the bearded.
- What is more this claim cannot be reduced to any claim about the kind of world I am in. Knowing who you are is like the information as to where you re given by the 'you are here' dot on shopping centre maps.
- To capture this extra, we need centred worlds. The extra is that I am not only in a world where some are bearded, I am at a 'beardedness point' in one of those worlds.



How to find the right set of worlds

- Think of 'I am bearded (now)' as like a diver-below flag; part of the content is the relation to the token.
- 
- Note that a token of 'I am bearded now' said by *x* at *t* is true iff *x* is bearded at *t*, the time of producing the token.
- Content is the set of centred worlds whose centres are bearded producers of tokens of 'I am bearded now' at the time of production.
- What about the belief that I am bearded now? Replace the token sentence with the token thinking.

Saying it in terms of the distinction between *A* and *C* intensions

- Easy introduction to this distinction. Take the sentence 'Actually there are electrons', where the role of 'actually' is that 'Actually *p*' is true at world *w* iff '*p* is true at the actual world.
- As there are electrons, this means that 'Actually there are electrons' is true at every world. If we call the set of worlds where a sentence is true its *C* intension (*'C'* as all but one are counterfactual), this is to say that the *C* intension of this sentence is the universal set.
- Now consider the set of worlds *w* such that 'Actually there are electrons' is true under the supposition that *w* is actual. This set will be the set of worlds where there are electrons. If we call the set of worlds where a sentence is true its *A* intension (*'A'* for actual), this is to say that the *A* intension is that set.
- So, for some sentences their *C* and *A* intensions differ. More examples next time along with cases where the two intensions are identical.