MIND-READING | Michaelmas 2018 | Maarten Steenhagen (ms2416@cam.ac.uk)
http://msteenhagen.github.io/teaching/2018mre/

Session 3: **Simulation theory**

Heal, Jane (1998), 'Co-Cognition and Off-Line Simulation: Two Ways of Understanding the Simulation Approach', *Mind and Language*, 13(4): 477–498.

Why we need to draw on more than theory alone

Recall, Jane Heal (1996) objected that the theory theory has to assume that we have a (tacit) grasp of a systematic and general theory of *relevance,* a precondition of solving the Frame Problem in Artificial Intelligence. This is implausible, because it would require assuming that humans (tacitly) work with a theory that would solve one of the biggest problems in AI: give a theoretical reduction of intelligent decisions. Many people are convinced there is no such theory, and even if there were, it would be computationally too heavy to attribute to finite creatures like us.

> The theory theorist may try to avoid this unpleasant outcome by denying that we have this systematic theoretical grip on relevance. […] Instead, he says, there is a non-theoretical background machinery—our remarkable cognitive system—which delivers to us the factors relevant to any given problem. (Heal 1996, 84)

We are individually able to make up our minds, at least typically. Our 'remarkable cognitive system' can do things that, as far as we know, no purely computational system can do. However,

> in presenting this picture of matters, the theory theorist has conceded the primary point that the simulationist is urging with respect to content. To apply the remarkable machinery to someone else's worldview so as to extract from it the thoughts relevant to answering a particular question is precisely to simulate his or her thought. (Heal 1996, 85)

Mental simulation

The core of the simulation theory is the idea that we use our ability to make up our own mind to make mental state attributions to others. ('Putting oneself in the other's shoes')

> Example: If I want to find out what someone whose trousers are on fire will do, I could set fire to my own trousers and see what I will do (first: panic; then: looking for something to cover and suffocate the flames). Assuming that others are relatively similar to myself, I can then attribute these reactions and intentions to them in predicting their behaviour.
>
> But clearly, I am also able to predict what I myself would do in the case my trousers were on fire. In that case I have to rely on *imagining* (e.g. visualising or supposing) my trousers to be on fire. (Without imagination, it wouldn't be a prediction or attribution, but actually having the states.)

This suggests I can find out about mental states either by actually making up my mind (cf. Moran 2001), or by imaginatively making up my mind (roughly, counterfactually making up my mind).

Although the idea of *simulation* (or 'replication', in Heal's terms) is crucial, that we draw on *imagination* or 'off-line processing' merely captures the typical way we make mental state attributions to others: we could just as well actually make up our mind and reason by analogy to the other. Imagination is only needed to explain our access to non-actual mental states.

Empirical evidence for simulation

Heal's argument against the theory theory does not itself seem to be a direct argument in favour of simulationism.

> *Primitivism about mindreading:* we can always take our ability to attribute mental states to others to be a primitive capacity, or simply to be part of our primitive 'remarkable cognitive system'. In that case, we may just be able to read off other people's states in the same way we are able to introspect our own. (Note, this is more plausible for non-cognitive states; arguably we do not have introspective access to cognitive states like beliefs and desires.)

Some have taken empirical findings in neuroscience to support the simulation theory.

> a particular set of neurons, activated during the execution of purposeful, goal-related hand actions, such as grasping, hold- ing or manipulating objects, discharge also when the monkey observes similar hand actions performed by another individual. We designated these neurons as 'mirror neurons' (Gallese 2001, 35)

The central finding here has been that of the 'mirror neurons'. Goldman defines these as follows:

> Mirror neurons are a class of neurons that discharge both when an individual (monkey, human, etc.) undergoes a certain mental or cognitive event endogenously and when it observes a sign that another individual undergoes or is about to undergo the same type of mental or cognitive event. (2008, 90)

This is controversial: (i) Even if some neural systems show this 'mirroring' feature, the simulationist would still have to establish that their activation in endogenous events is *primary*. (ii) Further, and more importantly, we should not conflate enabling conditions and capacities. (We might be low-level, sub-personal simulators without being high-level, personal level simulators.)

An a priori argument for simulation theory

But is the simulation claim an empirical hypothesis? Or is it an a priori truth? Heal thinks it is confused to think of the discussion as about contingent fact:

> it is commonly taken that the inquiry into … the extent of simulation in psychological understanding is empirical, and that scientific investigation is the way to tell whether ST … is correct. But this perception is confused. It is an *a priori truth* … that simulation must be given a substantial role in our personal-level account of psychological understanding. (1998, 477–478)

Heal's argument exploits the assumption that it is *a priori* that a belief about carrots is about carrots. A belief about carrots couldn't not be about carrots. This has implications for our understanding of those beliefs. She uses the analogy of thinking about a photograph of something:

> A person may have an excellent understanding of vegetables without having any sort of grasp on photographs of vegetables, indeed without so much as knowing that photographs exist. A person may also have knowledge about photographs in general (their varieties, how they are taken, how printed, how used etc.) without knowing that vegetables exist. But a person cannot be credited with rich and adequate knowledge of photographs of vegetables without knowing such things as what colour is likely to predominate in a close-up colour photograph of a well lit pile of clean carrots, as opposed to a pile of cabbages. Clearly this knowledge cannot be supplied merely by grasp on the general notion of a photograph. (1998, 481)

To represent to yourself that I have in my office a photograph of a pile of cabbages requires you to represent cabbages, besides representing the photograph itself. This is because a grasp of the idea of a photograph of *x* requires you to grasp how *x* was involved in the photograph's production.

Similarly, we cannot represent to ourselves that someone believes that there are cabbages on the kitchen top without ourselves entertaining the proposition that there are cabbages on the kitchen top. This is because a grasp of a belief about P presupposes a grasp of how P is relevant for action and further thought, given circumstances.

An objection to this as an argument for simulation theory

Does Heal's defence of simulation theory not itself presuppose that when we make an attribution, we already had access to their belief states in the first place? For otherwise, how would I know *which* propositions to entertain?

Heal's response is that when someone attributes mental states to another person, his first concern is with the other person's situation, and not with their mind.

> He is not looking at the subject to be understood but at the world around that subject. It is what the world makes the replicator think which is the basis for the beliefs he attributes to the subject. The process, of course, does not work with complete simplicity and directness. The replicator does not attribute to someone else belief in every state of affairs which he can see to obtain in the other's vicinity. A process of recentring the world in imagination is required. (Heal 1986, 139)