

1 Introduction

For over two decades there have been two main competing theories of folk psychology, but recent research seems to indicate a possible resolution to this issue.

In general terms, a substantial proportion of social interaction consists of, or relies upon, the attempted prediction by one person of another's future actions, or an explanation of past actions. It is, perhaps, a necessary component of your belief that you "know someone" that you think you can successfully predict how they would act in some situations. And, of course, there are some situations in which we think all normal people would behave similarly: finding themselves in danger, for instance, they would seek to escape from or mitigate that danger. So people generally have some understanding of each other, or at least think that they do, and this supposed understanding has commonly been called "folk psychology."

The main alternative theories of folk psychology are the "theory-theory" and the "simulation theory." The former takes folk psychology to be a quasi-scientific activity, involving entities such as beliefs and desires and hypotheses about law-like regularities governing their interactions. The alternative theory is that we "simulate" other people's minds, in order to predict and explain their behaviour, imaginatively putting ourselves in their place. Hybrids of theory-theory and simulation theory have also been suggested. In this paper I look at some of the main issues regarding each of these theories, then go on to consider hybrids of them in the context of empathy.

2 The theory-theory

The term "theory of mind" was first used in psychology by Premack and Woodruff (1978). Their work with a chimpanzee named Sarah seemed to demonstrate that she attributed desires to a human pictured in certain dilemmas, which she "solved" for him by selecting pictures depicting the appropriate action on his part. The authors suggested that the system of inferences implied by Sarah's choices "...is properly viewed as a theory." (*ibid.*, p. 515) Premack and Woodruff's theory accordingly became known as the "theory-theory."

In a simultaneously published peer commentary, Dennett (1978) suggested that theory of mind is very difficult or impossible to demonstrate unambiguously where language cannot be used to communicate with the subject. Investigations into the development of folk psychology in children accordingly followed. Wimmer and Perner (1983) (the title of whose article echoed that of Dennett's commentary) reported that the three-year-olds they tested appeared not to understand that a person might hold false beliefs. Baron-Cohen et al. (1985) designed a different version of that experiment,

which has since been repeated with various modifications, being generally referred to as “the Sally-Anne test,” after the characters in the scenario, or “the false-belief test.”

Sally and Anne are two dolls. Sally has a basket, and Anne, a box. The child is shown Sally putting a marble in her basket and then leaving the room. While Sally is out, Anne takes the marble from Sally’s basket and puts it in her own box. Sally then returns, and the child is asked where Sally will look for her marble. The correct answer is “in the basket,” but very young children appear to have no concept of false belief, and say “in the box,” presumably because that is where they know the marble is. Older children, however, from around four years of age, generally “pass” the test (*ibid.*): they correctly attribute a false belief to Sally, having, according to the theory-theory, become able to recognise that other people have minds of their own. This is the concept of the child as “little scientist.” (Gopnik and Meltzoff, 1997)

Perhaps the two most obvious questions concern the relatively early age at which children start to think about other people’s minds as such, and the ability, or lack thereof, of most people to discuss (folk) psychological theories. However, Davies and Stone (2001), following Goldman (1989), make a comparison with language acquisition, pointing out that

... almost nobody now thinks that there are good objections to the whole enterprise of Chomskyan linguistics starting from the fact that ordinary folk are not very good at articulating grammatical principles. Nor is linguistics threatened by a problem about early acquisition. The linguist can respond to the two putative objections by saying, first, that knowledge of language is partly tacit and, second, that it is partly innate. (*op. cit.*, p. 21)

And the theory-theorist can respond similarly.

3 The simulation theory

A rival explanation emerged in the mid-eighties. Heal (1986) and Gordon (1986) independently suggested that, rather than theorising about Sally’s mental state, the child would imagine herself in Sally’s position, empathise with her, “put herself in Sally’s shoes,” enabling her to *simulate* Sally’s thinking and behaviour. “The basic idea is that if the resources our own brain uses to guide our own behavior can be modified to work as representations of other people, then we have no need to store general information about what makes people tick: We just do the ticking for them.” (Gordon, 2009)

Heal (1998) later argued that simulation can be understood in two different ways. It was generally assumed that the theory-theory versus simulation

theory question is an empirical one (Boden, 2006), but simulation theory is necessarily true, albeit in a weak sense, if, in wondering whether someone likes coffee, I necessarily think about coffee. Strictly speaking, a person using pure theory could think only about preferences, beliefs, etc., and not about coffee. If both the person whose tastes I am considering, and I myself, are thinking about coffee, we are “co-cognizing,” in Heal’s terminology, which amounts to simulation on my part. If we take the view that thinking about thinking about coffee necessarily implies thinking about coffee, the theory-theory would appear to be a priori false, and simulation theory a priori true. (*op. cit.*)

This might lead you to conclude, with Boden, that “[t]he empirical question is what ‘sub-personal cognitive machinery’ is involved in implementing such co-cognition.” (Boden, 2006, p. 489) However, Heal argues that such thinking leads to a “threat of collapse:” if, in simulating another’s thinking, we use mechanisms that are (substantially) the same as their’s, then surely tacit knowledge of others’ minds is embedded in one’s own mind, and simulation reduces to theorising. (*op. cit.*)

Davies and Stone (2001) argue that this threat is illusory, saying that knowledge implies representation, and, where one mechanism is used to simulate a similar mechanism, no representation is directly involved. The assumption that the mechanisms are sufficiently similar—that I am so like you, say, that I have a good chance of guessing correctly what you might do in a given situation by imagining myself in your place—does seem to require representation, but that is part of the minimal theory that any simulationist is bound to accept anyway. (*ibid.*)

If we allow that there is, after all, no real threat of empirical, sub-personal simulation collapsing into theory-theory, is there any actual evidence for such a mechanism? Gallese and Goldman (1998) use findings concerning “mirror neurons” in macaque monkeys to support an argument for simulation theory. Mirror neurons are so named because they fire both when the animal makes a certain action and when it observes another performing the same action. Individual mirror neurons have not been found in humans (there are problems with recording single cell activation in humans), but studies using various neuroimaging techniques have revealed “mirror systems” that include the areas in which mirror neurons have been found in monkeys, but also others including the somatosensory cortex, which, it has been suggested, allow people to know what it feels like to perform the observed action. (Gazzola and Keysers, 2009)

Saxe (2005a) adapts the “argument from error” of Nichols and Stich (1995) “to show that the errors that human observers make are not consistent with the ‘resonance’ Simulation Theory embraced by mirror neurone enthusiasts. Rather, observers must rely on a naïve theory of psychology.” (*op. cit.*, p. 174f) Mitchell (2005) takes the target of Saxe’s argument to be simulation theory generally but that is not the case, in fact Saxe explicitly

promotes simulation/theory hybridism, as we shall see in the next section. Nor does she reject altogether the mirror arguments: “The mirror system does offer powerful insights into the neural representation of simple actions and some basic emotions...” (*op. cit.*, p. 174)

4 Empathy and hybrid theories

Theodor Lipps, writing in German on aesthetic appreciation, coined the term *Einfühlung* (Lipps, 1903), literally “feeling-into,” which came to be translated as “empathy.” (Wispé, 1990) He believed that it is through empathy that we come to know others: we do not perceive such emotions as pride, shame, anger, sorrow, and joy in others directly, but experience them vicariously. We “feel for” the other person. (*ibid.*)

Despite its relatively recent coinage, there are now many different definitions of empathy; Batson (2009), for instance, describes eight that he has found in the literature. However, he views these as “related but distinct phenomena,” and sees that range of definitions as resulting from attempts to answer two different questions (*ibid.*, p. 3):

1. “How can one know what another person is thinking and feeling?” and
2. “What leads one person to respond with sensitivity and care to the suffering of another?”

The main concepts of empathy as described by Batson are (*ibid.*, p. 4ff):

1. Knowing another person’s inner state, including his or her thoughts and feelings;
2. Adopting the posture or matching the neural responses of an observed other;
3. Coming to feel as another person feels;
4. Intuiting or projecting oneself into another’s situation;
5. Imagining how another is thinking and feeling;
6. Imagining how one would think and feel in the other’s place;
7. Feeling distress at witnessing another’s suffering; and
8. Feeling for another person who is suffering.

Concept 1 corresponds most closely to the first of Batson’s two questions, and he found that, despite all eight concepts having individually been classified as empathy, each of the five concepts 2–6 has been invoked to explain how the situation expressed by concept 1 can come about. Batson suggests,

seemingly somewhat arbitrarily, that the theory-theory might be used to explain how concept 1 is realised via concept 5, and simulation theory used for concepts 3, 4 and 6. He exempts concepts 7 and 8 from being considered in the context of theory- and simulation theories because they "...are not sources of knowledge (or belief) about another's state; they are reactions to this knowledge. Thus, they are not likely to be invoked to explain how one knows what another is thinking and feeling." (*ibid.*, p. 9) Instead, they are relevant to question 2.

Batson's account aligns with a view common among theory/simulation hybrid theorists, that we use theory in some situations and simulation in others. (Saxe, 2005b; Gordon, 2009) Saxe, however, considers these accounts unsatisfactory. In reply to Mitchell (*op. cit.*) she writes:

Proposals for how to distinguish the contexts requiring simulation or theorizing seem unnatural, for example dividing brief (simulation) from longer-term mental states (Perner and Kuhlberger, 2005), or accurate (simulation) from inaccurate attributions (Nichols and Stich, 2003)... More importantly, in these models simulation and theorizing exist side-by-side but independently, and the observer uses them one at a time. If anything, the dichotomy between the two processes is enhanced... Rather than focus on the circumstances in which observers either simulate or theorize, I prefer to ask how the separate intuitions that motivate [simulation theory] and [theory-theory] can be integrated into a single more general model. (Saxe, 2005b, p. 364)

Shamay-Tsoori (2009) reviews the literature on empathy and, like Batson, finds a number of different concepts, but she divides them into two categories: "The critical difference between cognitive empathy and affective or emotional empathy is that the former involves cognitive understanding of the other person's point of view whereas the latter also includes sharing of those feelings..." (*ibid.*, p. 215) The investigators who focus on affective empathy "typically study aspects such as helping behavior" (*ibid.*, p. 215), so Batson's concepts 7 and 8 fall into this category.

Following brief descriptions of theory-theory and simulation theory, in connection with which she mentions mirror neurons, Shamay-Tsoori goes on to suggest that the former

...views empathy as a thoroughly "detached" theoretical analysis that involves areas of the cortex that are usually activated during mental state attribution, whereas simulation depicts empathy as incorporating an attempt to replicate the other's affective mental state via neural networks related to emotion processing... it may be suggested that cognitive empathy involves more [theory] processing, whereas affective empathy involves more simulation processing. (*ibid.*, p. 216)

Watson and Greenberg (2009) explicate what Shamay-Tsoori implies there: mirror systems reflect emotions as well as actions. “Regions in the brain associated with feeling a specific emotion are activated by seeing that emotion in another or witnessing the other in a situation that might elicit the emotion.” (*ibid.*, p. 126) More specifically, Pfeifer and Dapretto (2009) report, in children asked to imitate or just observe various emotional facial expressions, “significant correlations between activity in mirror neuron and limbic regions and each of the first three subscales of the Interpersonal Reactivity Index” (*ibid.*, p. 188), a widely used test of empathy, of which these subscales measure the more affective components.

All of these neurological accounts involve a dichotomy between neural regions that process affect and cognition, but Pessoa (2008) argues against such a view, saying that “. . . complex cognitive-emotional behaviours have their basis in dynamic coalitions of networks of brain areas, none of which should be conceptualized as specifically affective or cognitive.” (*ibid.*, p. 148) The acknowledgement that some behaviours are complex, involving both cognitive and affective processing, is surely correct, and I would suggest that such an integrated way of considering cognition and affect fits Batson’s concepts of empathy, for instance, very well. Unfortunately, I am not qualified to assess the neurological arguments, but it seems to me that this sort of division of labour, theory of mind for cognitive and simulation for affective empathy, would be an extremely neat resolution of this debate, and I hope to see it confirmed and generally accepted in the not-too-distant future.

References

- S. Baron-Cohen, A.M. Leslie, U. Frith, et al. Does the autistic child have a “theory of mind”? *Cognition*, 21(1):37–46, 1985.
- C.D. Batson. These things called empathy: Eight related but distinct phenomena. In Decety and Ickes (2009).
- M.A. Boden. *Mind as machine: a history of cognitive science*. Oxford University Press, USA, 2006.
- M. Davies and T. Stone. Mental simulation, tacit theory, and the threat of collapse. *Philosophical Topics*, 29(1&2):127–73, 2001.
- Jean Decety and William Ickes, editors. *The Social Neuroscience of Empathy*. M.I.T. Press, Cambridge, Massachusetts, 2009.
- D.C. Dennett. Beliefs about beliefs. *Behavioural and Brain Sciences*, 4: 568–70, 1978.

- N. Eisenberg and J. Strayer, editors. *Empathy and its Development*. Cambridge University Press, Cambridge, 1990.
- V. Gallese and A. Goldman. Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, 2(12):493–501, 1998.
- V. Gazzola and C. Keysers. The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. *Cerebral Cortex*, 19(6):1239–1255, 2009.
- A. Goldman. Interpretation psychologized. *Mind and Language*, 4(3):161–185, 1989.
- A. Gopnik and A.N. Meltzoff. *Words, Thoughts, and Theories*. M.I.T. Press, 1997.
- R. Gordon. Folk psychology as simulation. *Mind and language*, 1(2):158–171, 1986.
- R.M. Gordon. Folk psychology as mental simulation. *Stanford Encyclopedia of Philosophy*, 2009. URL <http://plato.stanford.edu/entries/folkpsych-simulation/>. Retrieved 8 December 2009.
- J. Heal. Replication and functionalism. In J. Butterfield, editor, *Language, mind and logic*, pages 135–150. Cambridge University Press, 1986.
- J. Heal. Co-cognition and off-line simulation: Two ways of understanding the simulation approach. *Mind and Language*, 13(4):477–498, 1998.
- T. Lipps. Einfühlung, inner nachahmung, und organ-empfindungen. *Archiv für die gesamte Psychologie*, 2:185–204, 1903.
- J.P. Mitchell. The false dichotomy between simulation and theory-theory: the argument’s error. *Trends in Cognitive Sciences*, 9(8):363–364, 2005.
- S. Nichols and S. Stich. Second thoughts on simulation. In T. Stone and M. Davies, editors, *Mental Simulation: Evaluations and Applications*, pages 87–108. Blackwell, 1995.
- S. Nichols and S.P. Stich. *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Oxford University Press, USA, 2003.
- J. Perner and A. Kuhlberger. Mental simulation: Royal road to other minds. In B.F. Malle and S.D. Hodges, editors, *Other minds: How humans bridge the divide between self and others*, pages 174–189. The Guilford Press, 2005.

- L. Pessoa. On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, 9(2):148–158, 2008.
- J.H. Pfeifer and M. Dapretto. “Mirror, mirror, in my mind:” empathy, interpersonal competence, and the mirror neuron system. In Decety and Ickes (2009).
- D. Premack and G. Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and Brain sciences*, 1(4):515–526, 1978.
- R. Saxe. Against simulation: the argument from error. *Trends in Cognitive Sciences*, 9(4):174–179, 2005a.
- R. Saxe. Hybrid vigour: Reply to Mitchell. *Trends in Cognitive Sciences*, 9(8):364, 2005b.
- Simone G. Shamay-Tsoori. Empathic processing: its cognitive and affective dimensions and neuroanatomical basis. In Decety and Ickes (2009).
- J.C. Watson and L.S. Greenberg. Empathic resonance: A neuroscience perspective. In Decety and Ickes (2009).
- H. Wimmer and J. Perner. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children’s understanding of deception. *Cognition*, 13(1):103–128, 1983.
- L. Wispé. History of the concept of empathy. In Eisenberg and Strayer (1990).