References

1 Gordon, R.M. (2005) Simulation and systematic errors in prediction. *Trends Cogn. Sci.* 9. doi: 10.1016/j.tics.2005.06.003
2 Saxe, R. (2005) Against simulation: the argument from error. *Trends Cogn. Sci.* 9, 174–179
3 Rimbert, G. and Thommen, E. (2003) L'ignorance et la croyance dans la théorie de l'esprit. *Arch. Psychol.* 70, 255–271
4 Perner, J. and Horn, R. (2003) Ignorance or false negatives: do children of 4 or 5 years simulate belief with 'Not knowing = Getting it wrong?' *J. Cogn. Develop.* 4, 263–273
5 Ruffman, T. and Keenan, T.R. (1996) The belief-based emotion of surprise: the case for a lag in understanding relative to false belief. *Dev. Psychol.* 32, 40–49
6 Wilson, T.D. and Gilbert, D.T. (2003) Affective forecasting. In *Advances in Experimental Social Psychology* (Vol. 35) (Zanna, M., ed.), pp. 345–411, Elsevier
7 Hofer, B.K. and Pintrich, P.R., eds (2002) *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*, Erlbaum
8 Levin, D., ed. (2004) *Thinking and Seeing: Visual Metacognition in Adults and Children*, MIT Press

1364-6613/$ - see front matter
doi:10.1016/j.tics.2005.06.002

Letters

The false dichotomy between simulation and theory-theory: the argument's error

Jason P. Mitchell

Department of Psychology, Harvard University, William James Hall, 33 Kirkland Street, Cambridge, MA 02138, USA

Two classes of theories have emerged to explain the human ability to mentalize about the unobservable psychological states (feelings, beliefs, goals, etc.) of others: 'simulation theory' and 'theory-theory'. Recently, Saxe [1] has catalogued a number of experimental situations in which the nature of observers' mentalizing performance suggests that they cannot be using simulation to infer another person's mental states.

Curiously, in making her 'argument from error', Saxe follows many earlier commentators in insisting that, because observers do not *always* simulate, they must *never* do so. After all, evidence in support of theory-theory only argues 'against simulation' to the extent that one assumes that either simulation or theory-theory – but not some combination of the two – must fully account for mentalizing performance in all situations, for all people, and throughout all stages of cognitive development.

Is such 'either/or' logic useful for considering how people actually solve the problem of understanding other minds? Saxe suggests that compromise models (simulation/theory-theory hybrids) should be rejected for reasons of parsimony: better to explain behavior by postulating one cognitive process than some admixture of two or more. However, like all biological systems, the brain has been cobbled together through natural selection, a process notorious for tinkering with existing mechanisms without much regard for Occam's razor. And indeed, much of the progress made by cognitive neuroscience over the past three decades has been of a decidedly non-parsimonious nature, in particular the repeated observation that complex cognitive processes – such as memory [2], cognitive control [3], and semantic knowledge [4] – do not reflect the operation of unitary mechanisms but rather of multiple processes with distinct neuroanatomical correlates.

Moreover, Saxe neglects a number of empirical observations that pose significant challenges to non-simulationist accounts of mentalizing. For example, Niedenthal and colleagues have demonstrated that observers make judgments of another person's emotional state in relation to their own feelings (e.g. sad observers more readily perceive sadness in ambiguous facial displays than happy observers [5]) and that this effect is eliminated when observers are prevented from spontaneously mimicking the target's facial expression [6]. Likewise, simulation theory suggests that observers should be expected to reason differently about targets perceived to be similar vs. dissimilar to oneself (see [7] for a detailed account of this aspect of simulation theory). Support for this prediction comes from demonstrations that observers impart different emotional experiences to members of their own social groups [8], and more readily project their own goals and predilections onto similar targets (e.g. people sharing the same hobbies) than dissimilar ones [9]. Moreover, recent neuroimaging work has suggested that distinct regions of the medial prefrontal cortex subserve mentalizing about similar and dissimilar others [7]. Whereas simulation readily explains these data, theory-theory must appeal to an unspecified set of ancillary (i.e. non-parsimonious) mechanisms to do the same.

At this point, the study of mentalizing requires less debate about what may be a false dichotomy between simulation and theory-theory and increased focus on the different circumstances in which observers flexibly deploy one or the other process to understand the minds of others.

Corresponding author: Mitchell, J.P. (jmitchel@wjh.harvard.edu).
Available online 11 July 2005

References

1 Saxe, R. (2005) Against simulation: the argument from error. *Trends Cogn. Sci.* 9, 174–179

2 Schacter, D.L. and Buckner, R.L. (1998) Priming and the brain. *Neuron* 20, 185–195
3 Miller, E.K. and Cohen, J.D. (2001) An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202
4 Caramazza, A. and Shelton, J.R. (1998) Domain-specific knowledge systems in the brain: the animate–inanimate distinction. *J. Cogn. Neurosci.* 10, 1–34
5 Niedenthal, P.M. *et al*. (2000) Emotional state and the detection of change in facial expression of emotion. *Eur. J. Soc. Psychol.* 30, 211–222
6 Niedenthal, P.M. *et al*. (2001) When did her smile drop: facial mimicry and the influences of emotional state on the detection of change in emotional expression. *Cogn. Emotion* 15, 853–864
7 Mitchell, J.P. *et al*. The link between social cognition and self-referential thought in the medial prefrontal cortex. *J. Cogn. Neurosci.* (in press)
8 Demoulin, S. *et al*. (2004) Emotional prejudice can lead to infra-humanisation. In *European Review of Social Psychology* (Stroebe, W. and Hewstone, M., eds), pp. 259–296, Psychology Press/Taylor & Francis
9 Ames, D.R. (2004) Inside the mind reader's tool kit: projection and stereotyping in mental state inference. *J. Pers. Soc. Psychol.* 87, 340–353

1364-6613/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tics.2005.06.010

Letters Response

Hybrid vigour: Reply to Mitchell

Rebecca Saxe

Psychology Department, Harvard University, Cambridge, MA 02138, USA

Mitchell [1] lays serious charges. I plead innocent on all counts. My article [2] does not claim that 'because observers do not *always* simulate, they must *never* do so.' Far from ignoring cases that are well described by simulation, like the attribution of basic emotions, I explicitly grant them; and rather than promoting an 'either/or' dichotomy between Simulation Theory (ST) and Theory Theory (TT), I dedicate the main figure to a discussion of possible ST–TT hybrids.

Hybrid theories come in multiple flavours, though. Mitchell calls for 'increased focus on the different circumstances in which observers flexibly deploy one or the other process.' Previous authors [3,4] do take this tack, proposing that observers simulate in some contexts and use a naïve theory in others. But the existing dual-system models are unsatisfying. Proposals for how to distinguish the contexts requiring simulation or theorizing seem unnatural, for example dividing brief (simulation) from longer-term mental states [3], or accurate (simulation) from inaccurate attributions [4] (see Box 3 in [2]). 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.

So although I agree with Mitchell's call for hybrid theories, I disagree with the flavour he chooses. Rather than focus on the circumstances in which observers either simulate *or* theorize, I prefer to ask how the separate intuitions that motivate ST and TT can be integrated into a single more general model. For example, how could a naïve theory of mind be informed by the observer's own experiences? (see also Box 5 in [2]). A more integrated hybrid would respect the differences between ST and TT, but undermine the dichotomy. If my theory of Mitchell's motivation is right, this is what he should want too.

References

1 Mitchell, J.P. (2005) The false dichotomy between simulation and theory-theory: the argument's error. *Trends Cogn Sci.* 9, this issue doi: 10.1016/j.tics.2005.06.010
2 Saxe, R. (2005) Against simulation: the argument from error. *Trends Cogn. Sci.* 9, 174–179
3 Perner J. and Kuhlberger, A. (in press). Mental simulation: royal road to other minds? In *Other Minds* (Malle, B.F. and Hodges, S.D., eds), Guilford Press
4 Nichols, S. and Stich, S. (2003) *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding of Other Minds*, Oxford University Press

1364-6613/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tics.2005.06.017

Corresponding author: Saxe, R. (Saxe@mit.edu).
Available online 11 July 2005