09: Signature Limits

s.butterfill@warwick.ac.uk

1. Minimal Theory of Mind

An agent's *field* is a set of objects related to the agent by proximity, orientation and other factors.

First approximation: an agent *encounters* an object just if it is in her field.

A *goal* is an outcome to which one or more actions are, or might be, directed.

Principle 1: one can't goal-directedly act on an object unless one has encountered it.

Applications: subordinate chimps retrieve food when a dominant is not informed of its location (Hare et al. 2001); when observed scrub-jays prefer to cache in shady, distant and occluded locations (Dally et al. 2004; Clayton et al. 2007).

First approximation: an agent *registers* an object at a location just if she most recently encountered the object at that location.

A registration is *correct* just if the object is at the location it is registered at.

Principle 2: correct registration is a condition of successful action.

Applications: 12-month-olds point to inform depending on their informants' goals and ignorance (Liszkowski et al. 2008); chimps retrieve food when a dominant is misinformed about its location (Hare et al. 2001); scrub-jays observed caching food by a competitor later re-cache in private (Clayton et al. 2007; Emery & Clayton 2007).

Principle 3: when an agent performs a goal-directed action and the goal specifies an object, the agent will act as if the object were actually in the location she registers it at.

Applications: some false belief tasks (Onishi & Baillargeon 2005; Southgate et al. 2007; Buttelmann et al. 2009).

2. Signature Limits

A *signature limit of a model* is a set of predictions derivable from the model which are incorrect, and which are not predictions of other models under consideration.

Automatic belief-tracking in adults (and belief-tracking in infants) is subject to signature limits associated with minimal theory of mind (Wang et al. 2015; Low 2010; Low et al. 2014; Mozuraitis et al. 2015; Edwards & Low 2017; contrast Scott et al. 2015.)

	Propositional attitude	Relational attitude
level-1 perspective taking	Y	Y
level-2 perspective taking	Y	N
false beliefs about non-existence	Y	N
false beliefs about location	Y	Y
false beliefs about identity	Y	N
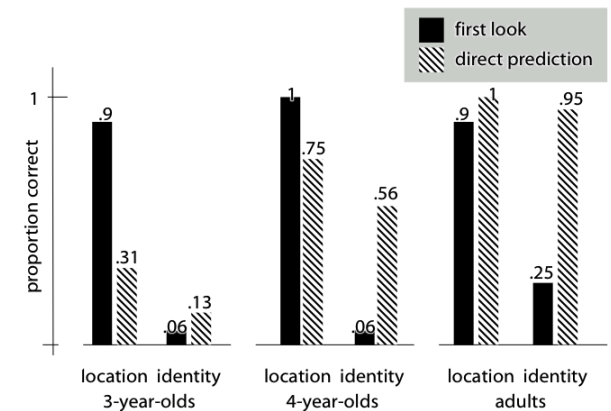
source: Low & Watts (2012)

For adults (and children who can do this), representing perceptions and beliefs as such—and even merely holding in mind what another believes, where no inference is required—involves a measurable processing cost (Apperly et al. 2008, 2010), consumes attention and working memory in fully competent adults Apperly et al. 2009; Lin et al. 2010; McKinnon & Moscovitch 2007, may require inhibition (Bull et al. 2008) and makes demands on executive function (Apperly et al. 2004; Samson et al. 2005).

3. Objections

'the theoretical arguments offered [...] are [...] unconvincing, and [...] the data can be explained in other terms' (Carruthers 2015b; see also Carruthers 2015a).

'A cooperative multi-system architecture is better able to explain infant belief representation than a parallel architecture, and causal representation, schemas and models provide a more promising basis for flexible belief representation than does a rule-based approach of the kind described by Butterfill and Apperly' (Christensen & Michael 2016; see also Michael & Christensen 2016; Michael et al. 2013).

References

Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential theory of mind task. *Cognition*, *106*(3), 1093–1108.

Apperly, I. A., Carroll, D., Samson, D., Humphreys, G., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? evidence from adults' ability to follow instructions from an ignorant speaker. *The Quarterly Journal of Experimental Psychology*, *63*(6), 1201–1217.

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: Neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience*, *16*(10), 1773–1784.

Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology*, *45*(1), 190–201.

Bull, R., Phillips, L., & Conway, C. (2008). The role of control functions in mentalizing: Dual-task studies of theory of mind and executive function. *Cognition*, *107*(2), 663–672.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*(2), 337–342.

Carruthers, P. (2015a). Mindreading in adults: evaluating two-systems views. *Synthese*, *forthcoming*, 1–16.

Carruthers, P. (2015b). Two systems for mindreading? *Review of Philosophy and Psychology*, *7*(1), 141–162.

Christensen, W. & Michael, J. (2016). From two systems to a multi-systems architecture for mindreading. *New Ideas in Psychology*, *40*, 48–64.

Clayton, N. S., Dally, J. M., & Emery, N. J. (2007). Social cognition by food-caching corvids. the western scrub-jay as a natural psychologist. *Philosophical Transactions of the Royal Society B*, *362*, 507–552.

Dally, J. M., Emery, N. J., & Clayton, N. S. (2004). Cache protection strategies by western scrub-jays (aphelocoma californica): hiding food in the shade. *Proceedings of the Royal Society B: Biological Sciences*, *271*(0), S387–S390–S387–S390.

Edwards, K. & Low, J. (2017). Reaction time profiles of adults' action prediction reveal two mindreading systems. *Cognition*, *160*, 1–16.

Emery, N. J. & Clayton, N. S. (2007). How to build a scrub-jay that reads minds. In S. Itakura & K. Fujita (Eds.), *Origins of the Social Mind: Evolutionary and Developmental Perspectives*. Tokyo: Springer.

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, *61*(1), 139–151.

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*(3), 551–556.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108*(3), 732–739.

Low, J. (2010). Preschoolers' implicit and explicit False-Belief understanding: Relations with complex syntactical mastery. *Child Development*, *81*(2), 597–615.

Low, J., Drummond, W., Walmsley, A., & Wang, B. (2014). Representing how rabbits quack and competitors act: Limits on preschoolers' efficient ability to track perspective. *Child Development*, *forthcoming*.

McKinnon, M. C. & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, *102*(2), 179–218.

Michael, J. & Christensen, W. (2016). Flexible goal attribution in early mindreading. *Psychological Review*, *123*(2), 219–227.

Michael, J., Christensen, W., & Overgaard, S. (2013). Mindreading as social expertise. *Synthese*, *191*(5), 817–840.

Mozuraitis, M., Chambers, C. G., & Daneman, M. (2015). Privileged versus shared knowledge about object identity in real-time referential processing. *Cognition*, *142*, 148–165.

Onishi, K. H. & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(8), 255–258.

Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain*, *128*(5), 1102–1111.

Scott, R. M., Richman, J. C., & Baillargeon, R. (2015). Infants understand deceptive intentions to implant false beliefs about identity: New evidence for early mentalistic reasoning. *Cognitive Psychology*, *82*, 32–56.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by two-year-olds. *Psychological Science*, *18*(7), 587–592.

Wang, B., Hadi, N. S. A., & Low, J. (2015). Limits on efficient human mindreading: Convergence across chinese adults and semai children. *British Journal of Psychology*, *106*(4), 724–740.