Lecture 04 : Theory of Mind (II) : Origins of Mind

Stephen A. Butterfill & Richard Moore
< >

Monday, 1st February 2021

Contents

1. Models and Processes

Lecturer: Stephen A. Butterfill

After claiming that 'chimpanzees understand … intentions … perception and knowledge,' Call & Tomasello (2008) qualify their claim:

> 'chimpanzees probably do not understand others in terms of a fully human-like belief–desire psychology' (Call & Tomasello 2008, p. 191).

This is plausible. The emergence in human development of the most sophisticated abilities to represent mental states probably depends on rich social interactions involving conversation about the mental (Slaughter & Gopnik 1996; Peterson & Slaughter 2003; Moeller & Schick 2006), on linguistic abilities,[1] and on capacities to attend to, hold in mind and inhibit things (Benson et al. 2013; Devine & Hughes 2014). These are all scarce or absent in chimpanzees and other nonhumans. So it seems unlikely that the ways humans at their most reflective represent mental states will match the ways nonhumans represent mental states. Reflecting on how adult humans talk about mental states is no way to understand how others represent them.

Heyes offers a diagnosis:

> 'the core theoretical problem in contemporary research on animal mindreading is that … the conception of mindreading that dominates the field … is too underspecified to allow effective communication among researchers, and reliable identification of evolutionary precursors of human mindreading through observation and experiment' (Heyes 2015, p. 321).

But how can we more fully specify mindreading?

2. Minimal Theory of Mind

Lecturer: Stephen A. Butterfill

Which models of minds and actions underpin which mental state tracking processes?

[1] See Moeller & Schick (2006, p. 760): 'Our results provide support for the concept that access to conversations about the mind is important for deaf children's ToM development, in that there was a significant relationship between maternal talk about mental states and deaf children's performance on verbal ToM tasks.' See also Milligan et al. (2007); Kovács (2009).

2.1. What Is a Model?

A *model* is just a way some aspects of the world could be. A *model of minds and actions* is a way mental aspects of the world could be.

A model is something that can serve different purposes. Having a model does not commit you to using it for any particular purpose. The model's usefulness does not depend only on its accuracy: the ease with which it can be used to imagine, build or navigate matters. The best model for a given set of purposes may not be the most accurate. Further, it can be advantageous to have multiple models of a single thing. For example, building a house can involve creating multiple models.

Theorists specify models in various ways including by giving a theory or by constructing something physical.

A model is distinct from a theory. A model can be used to make claims about the world, but the model itself entails nothing about how the world actually is. By contrast, a theory does (Godfrey-Smith 2005).

In saying that an individual or a process *relies* on a model, we are attempting to capture the way aspects of the world seem from the individual's or processes' point of view. There is no commitment to any claim about how the model relates to the individual or process. There is no suggestion, in saying that an individual relies on a model, that they have a physical model; nor that they know any of a theory which we, as theorists, use to specify the model.

2.2. 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.[2]

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).

[2] Not to be confused with a *goal-state*, which is an intention or other state of an agent linking an action to a particular goal to which it is directed.

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).

3. Signature Limits

Lecturer: Stephen A. Butterfill

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, are both subject to signature limits associated with minimal theory of mind (Wang et al. 2015; Low & Watts 2013; Low et al. 2014; Mozuraitis et al. 2015; Edwards & Low 2017; Fizke et al. 2017; Oktay-Gür et al. 2018; Edwards & Low 2017, 2019; contrast Scott et al. 2015).

3.1. Objections

1. Low & Watts (2013) is replicable, but the paradigm involves confounds and so the results do not provide good evidence of belief tracking (Kulke et al. 2018).[3]

[3] Kulke et al. (2018) argue that although the paradigm from Low & Watts (2013) replicates, attempts to modify it to avoid confounding factors do not produce comparable results. In full:

> 'There are two broad possibilities why only the Low and Watts (2013) paradigm was robustly replicated. One possibility is that this paradigm is particularly valid (perhaps because of lower processing demands or other relevant task factors) and therefore the most sensitive and suitable one to

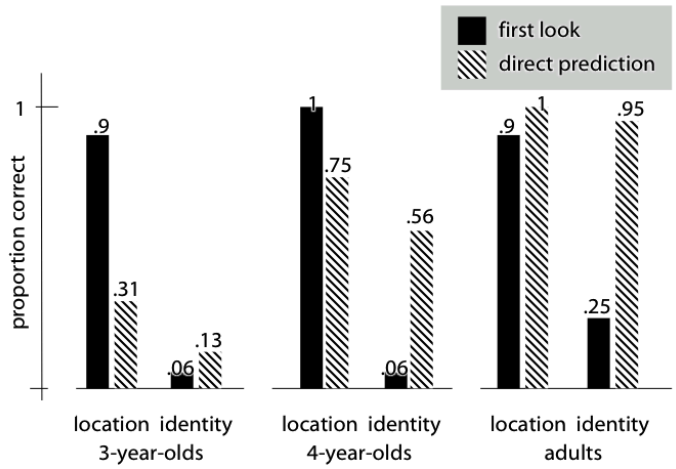

Figure 1: Signature limits illustrated. A response-by-content interaction that is robust across age-groups. *Source*:redrawn from Low & Watts (2013)

2. Infant belief-tracking is not subject to the signature limit about identity (Scott et al. 2015).
3. 'the theoretical arguments offered [...] are [...] unconvincing, and [ldots] the data can be explained in other terms' (Carruthers 2015b; see also Carruthers 2015a).

4. Automatic Mindreading in Adults

Lecturer: Stephen A. Butterfill

Is mindreading automatic? (More carefully: Does belief tracking in human adults depend only on processes which are automatic?)

A process is *automatic* to the degree that whether it occurs is independent of its relevance to the particulars of the subject's task, motives and aims.

There is evidence that some mindreading in human adults is entirely a consequence of relatively automatic processes (Kovács et al. 2010; Schneider et al.

---

tap implicit theory of mind. The contrary possibility is that this task may be particularly prone to alternative explanations because of potential confounds' (p. 8)

This motivated them to consider modified versions of the paradigm avoiding confounds, but:

'the original pattern of belief-congruent looking could be reproduced only under conditions in which the belief congruency of the locations is confounded with additional factors, and therefore, this pattern might not reflect belief-based anticipation' (p. 9)

2012; van der Wel et al. 2014; Edwards & Low 2017, 2019), and that not all mindreading in human adults is (Apperly et al. 2008, 2010; van der Wel et al. 2014).

Qureshi et al. (2010) found that automatic and nonautomatic mindreading processes are differently influenced by cognitive load, and Todd et al. (2016) provided evidence that adding time pressure affects nonautomatic but not automatic mindreading processes.

There is also limited evidence that people are unaware of automatic belief tracking processes:

> 'Participants never reported belief tracking when questioned in an open format after the experiment ("What do you think this experiment was about?"). Furthermore, this verbal debriefing about the experiment's purpose never triggered participants to indicate that they followed the actor's belief state' (Schneider et al. 2012, p. 2)

4.1. Objection

Level 1 perspective-taking in the Samson 'dot task' does not appear to be more automatic than Level 2 perspective-taking (Todd et al. 2020).[4] This finding is puzzing if we take the evidence for automatic belief-tracking at face value: why would belief-tracking but not level-1 perspective taking be automatic? Todd et al.'s finding is also incompatible with, and therefore evidence against, the conjecture that automatic belief-tracking processes rely on minimal theory of mind because minimal theory of mind involves Level-1 perspective-taking.

[4] These authors comment:

> 'not only did we consistently observe that altercentric interference was weaker when the avatar's perspective was less relevant to participants' task goal; we also consistently failed to observe any evidence of altercentric interference in L1-VPT in these conditions' (Todd et al. 2020, p. 16).

and

> 'reducing the goal-relevance of a cartoon avatar's perspective weakened both Level-1 and Level-2 visual perspective calculation. ... both Level-1 and Level-2 visual perspective calculation may be dependent on hav- ing a (remote) goal to process a target agent's perspective' (Todd et al. 2020, p. 18).

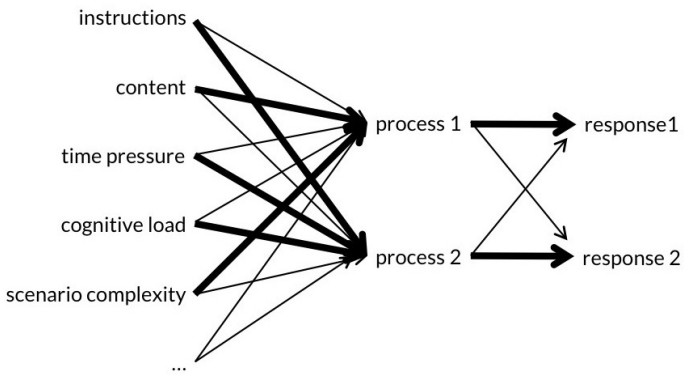

Figure 2: A modest dual-process theory claims only that two or more processes are distinct *Source*:homemade

5. A Dual Process Theory of Mindreading

Lecturer: Stephen A. Butterfill

What is a dual-process theory? In general, a modest dual-process theory claims just this:

> Two (or more) mindreading processes are distinct: the conditions which influence whether they occur, and which outputs they generate, do not completely overlap.

A key feature of this dual process theory is its *theoretical modesty*: it involves no a priori commitments concerning the particular characteristics of the processes. Identifying characteristics of the process is a matter of discovery. Further, their characteristics may vary across domains. The characteristics that distinguish processes involved in goal tracking may not entirely overlap with those that distinguish processes involved in segmenting physical objects and representing them as persisting, for example.

In the case of mindreading, we can elaborate a dual-process theory by starting with automaticity, as this is one of the most-studied features, and add a claim about signature limits which appears to be partially supported by the available evidence:

> *Dual Process Theory of Mindreading.* Automatic and nonautomatic mindreading processes are independent in this sense: different conditions influence whether they occur and which ascriptions they generate (see *Automatic Mindreading in Adults* (section §4)); and the automatic processes only rely on a minimal model of minds and actions (see *Signature Limits* (section §3)).

5.1. A Developmental Theory

1. Automatic and nonautomatic mindreading processes both occur from the first year of life onwards.
2. The model of minds and actions underpinning automatic mindreading process does not significantly change over development.
3. In the first three or four years of life, nonautomatic mindreading processes involve relatively crude models of minds and actions, models which do not enable belief tracking.
4. What changes over development is typically just that the model underpinning nonautomatic mindreading becomes gradually more sophisticated and eventually comes to enable belief tracking.

5.2. Objections

Christensen & Michael argue that the dual process theory is less well supported overall than an alternative:

> '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).

6. Conclusion: Models and Processes

Lecturer: Stephen A. Butterfill

In attempting to understand the ontogentic development of mindreading, we have been confronted with, and attempted to answer, two puzzles:

1. How do observations about tracking support conclusions about models?
2. Why are there dissociations in nonhuman apes', human infants' and human adults' performance on belief-tracking tasks?

The proposed answers are:

1. Using the method of signature limits (conjectures about models generate predictions about otherwise unexpected patterns of failure in tracking; see *Signature Limits* (section §3)).

2. Because there are multiple kinds of belief-tracking process, which (i) rely on disinct models of minds and actions; and (ii) have different influences on responses (see *A Dual Process Theory of Mindreading* (section §5)).

Glossary

automatic On this course, a process is *automatic* just if whether or not it occurs is to a significant extent independent of your current task, motivations and intentions. To say that *mindreading is automatic* is to say that it involves only automatic processes. The term 'automatic' has been used in a variety of ways by other authors: see Moors (2014, p. 22) for a one-page overview, Moors & De Houwer (2006) for a detailed theoretical review, or Bargh (1992) for a classic and very readable introduction 4–6

minimal model of minds and actions A model specified by a minimal theory of mind. 7

minimal theory of mind A theory of the mental in which: (a) mental states are assigned functional roles that can be readily codified; and, (b), the contents of mental states can be distinguished by things which, like locations, shapes and colours, can be held in mind using some kind of quality space or feature map. 6, 9

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. J., Samson, D., Humphreys, G. W., 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*, 1201–1217.

Bargh, J. A. (1992). The Ecology of Automaticity: Toward Establishing the Conditions Needed to Produce Automatic Processing Effects. *The American Journal of Psychology*, *105*(2), 181–199.

Benson, J. E., Sabbagh, M. A., Carlson, S. M., & Zelazo, P. D. (2013). Individual differences in executive functioning predict preschoolers' improvement

from theory-of-mind training. *Developmental Psychology*, *49*(9), 1615–1627.

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.

Call, J. & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187–192.

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.

Devine, R. T. & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development*, *85*(5), 1777–1794.

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

Edwards, K. & Low, J. (2019). Level 2 perspective-taking distinguishes automatic and non-automatic belief-tracking. *Cognition*, *193*, 104017.

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.

Fizke, E., Butterfill, S. A., van de Loo, L., Reindl, E., & Rakoczy, H. (2017). Signature limits in early theory of mind: Toddlers spontaneously take into account false beliefs about an objects' location but not about its identity. *Journal of Experimental Child Psychology*, *forthcoming*.

Godfrey-Smith, P. (2005). Folk psychology as a model. *Philosophers' Imprint*, *5*(6).

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

Heyes, C. M. (2015). Animal mindreading: what's the problem? *Psychonomic Bulletin & Review*, *22*(2), 313–327.

Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science*, *12*(1), 48–54.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830 –1834.

Kulke, L., von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is Implicit Theory of Mind a Real and Robust Phenomenon? Results From a Systematic Replication Study. *Psychological Science*, 0956797617747090.

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., 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*.

Low, J. & Watts, J. (2013). Attributing false-beliefs about object identity is a signature blindspot in humans' efficient mindreading system. *Psychological Science*, *24*(3), 305–311.

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.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, *78*(2), 622–646.

Moeller, M. P. & Schick, B. (2006). Relations between maternal input and theory of mind understanding in deaf children. *Child Development*, *77*(3), 751–766.

Moors, A. (2014). Examining the mapping problem in dual process models. In *Dual process theories of the social mind* (pp. 20–34). Guilford.

Moors, A. & De Houwer, J. (2006). Automaticity: A Theoretical and Conceptual Analysis. *Psychological Bulletin*, *132*(2), 297–326.

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

Oktay-Gür, N., Schulz, A., & Rakoczy, H. (2018). Children exhibit different performance patterns in explicit and implicit theory of mind tasks. *Cognition*, *173*, 60–74.

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

Peterson, C. & Slaughter, V. (2003). Opening windows into the mind: mothers' preferences for mental state explanations and children's theory of mind. *Cognitive Development*, *18*(3), 399–429.

Qureshi, A., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230–236.

Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others' mental states. *Journal of Experimental Psychology: General*, *141*(3), 433–438.

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.

Slaughter, V. & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. *Child Development*, *67*, 2967–2988.

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

Todd, A. R., Cameron, C. D., & Simpson, A. (2016). Dissociating processes underlying level-1 visual perspective taking in adults. *Cognition*, *in press*.

Todd, A. R., Cameron, C. D., & Simpson, A. J. (2020). The goal-dependence of level-1 and level-2 visual perspective calculation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, No Pagination Specified–No Pagination Specified.

van der Wel, R. P. R. D., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? evidence from a continuous measure. *Cognition*, *130*(1), 128–133.

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.