

# Cloning Two Systems

<>

According to the two systems hypothesis, mindreading in humans involves at least two mindreading systems which make complementary trade-offs between flexibility and efficiency. The more flexible system is thought to be responsible, among other things, for verbal replies to communicative prompts, while the more efficient system underlies spontaneous anticipatory looking, which can be automatic (Low & Watts 2013; Low et al. 2014; Wang et al. 2015; Schneider et al. 2014). How could a mindreading system achieve a degree of cognitive efficiency sufficient for automaticity? Apperly & Butterfill (2009) argue that cognitive efficiency in mindreading is incompatible with relying on a canonical model of minds in which mental states are characterised as propositional attitudes, with arbitrarily nestable contents and functional roles that mirror evidential and reason-giving relations. They suggest that, special cases aside, relying on such a model requires reasoning as complex as any humans are capable of (see also Harris 1994; Heal 1998). But there is arguably a further, more general requirement on cognitively efficient systems: information encapsulation. To say that a system is *informationally encapsulated* is to say that the flow of information into the system is interestingly limited: there can be information in the larger unit of which it is part which, although relevant to the operation of this system, is nevertheless persistently unavailable to the system (Fodor 1983, pp. 64ff). In general, cognitive efficiency in tasks such as mindreading where there is no in principle limit to what information could be relevant plausibly requires information encapsulation. This architectural claim generates predictions which enable the two systems hypothesis to be distinguished from the view that there is just one mindreading system. Here is a recipe for generating such predictions.

Take a fact, *P*, which would not typically have any bearing on what anyone believes, and is unlikely to do so. Then identify a further fact, *Q* which is highly improbable and which is also such that, given *Q*'s truth, the truth of *P* does bear on what a protagonist believes. The way *P* bears on the protagonist's beliefs given *Q* should be clear

to any potential experimental subjects. Then contrive a situation in which  $P$  and  $Q$  are manifestly true. Will subjects track those of the protagonist's beliefs which are consequences of (and counterfactually depend upon)  $P$  and  $Q$  being true?

The view that there is just one mindreading system predicts that the answer is yes. The two systems hypothesis predicts also that the answer is yes when responses are a consequence of a flexible mindreading system. However this hypothesis predicts that the answer is no when responses are a consequence of an efficient mindreading system. This is because an efficient mindreading system achieves efficiency in part by being informationally encapsulated: as the recipe involves making an arbitrary and improbable connection between  $P$  and the protagonist's beliefs, it is unlikely that information about  $P$  would be available to an efficient mindreading system.

As an example, consider a situation in which the protagonist has just emerged from the blue room ( $P$ ), and aliens have installed a machine in the blue room which makes anyone in it temporarily omniscient ( $Q$ ) or temporarily capable only of false beliefs ( $Q'$ ). Or consider a situation in which the protagonist has just emerged from a cubicle ( $P$ ), and is therefore a psychological clone of another individual ( $Q$ ).

## References

- Apperly, I. A. & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 2009(116), 4.
- Fodor, J. (1983). *The Modularity of Mind: an Essay on Faculty Psychology*. Bradford book. Cambridge, Mass ; London: MIT Press.
- Harris, P. (1994). Thinking by children and scientists. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the Mind: domain specificity in cognition and culture*. Cambridge: Cambridge University Press.
- Heal, J. (1998). Co-cognition and off-line simulation: Two ways of understanding the simulation approach. *Mind and Language*, 13(4), 477–498.
- 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.

- Schneider, D., Nott, Z. E., & Dux, P. E. (2014). Task instructions and implicit theory of mind. *Cognition*, 133(1), 43–47.
- 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*, n/a–n/a.