Developing a Theory of Mind: A Connectionist Investigation

Philip J. Rudling*, J. Richard Eiser* and Denis Mareschal**

*Department Of Psychology, University Of Exeter, Perry Rd, Exeter EX4 5QR, UK.
{p.j.rudling, j.r.eiser}@exeter.ac.uk

**Centre For Brain And Cognitive Development, Department Of Psychology, Birkbeck College, University Of London, London, WC1E 7HX, UK.
d.mareschal@bkk.ac.uk

Introduction

It is generally accepted that by 4 or 5 years of age, children have normally developed a theory of mind (Mitchell, 1996). One of the key signs that a child is acknowledging another's mental states as a causal factor in their behavior is to pass the false belief test. The aim of this abstract is to present a connectionist model that accounts for this change in behavior without necessarily positing an explicit theory of mind.

The particular task modeled here is the deceptive box or "Smarties" test, in which a child is shown a Smarties tube (well known British confectionery) containing some pencils and then asked what another child might think is inside (Perner, Leekam and Wimmer, 1987). The younger child answers "Pencils", supplying their own knowledge, whereas the older says "Smarties", as another would not know the unusual contents of the box.

The Network

By using a connectionist approach, one can show a smooth and seamless change in performance that is accounted for simply by the training set presented to the network. The model is similar in some respects to the Cohen, Dunbar and McClelland (1991) model of the conflicting Stroop task: in this case the two conflicting channels are what is known (reality) and what is seen (appearance).

The network consists of three layers; an input bank, a partially connected hidden layer and an output. In this simplified case, the network is trained to answer the question "what is in the box?". The model is given three types of information; the agent in question (self or other), what is seen (e.g. a Smarties box), and what is known to be inside (e.g. pencils). The network is trained with situations such as "I see Smarties box and I know it contains Smarties". The network computes an output which is compared to target (c.f. opening a box to check) and the connection weights are updated according to the backward propagation rule. One crucial point in the training is that, as in the real world, the network/child does not have direct access to others' knowledge and has to use their own knowledge to predict the behavior of others. The fact that one's own knowledge is a poor predictor of other's behavior is reflected in the training set by adding noise in these situations.

Conclusions

The results of the network match the developmental progression of children, passing from an early "Pencils" stage to a "Smarties" stage, with a smooth developmental progression. The network is able to replicate other aspects of this development, for example the fact that a "noisier" environment (i.e. lots of siblings) causes change in performance to occur earlier (Perner, Ruffman and Leekam, 1994). Such an approach has qualities of both "Theory" Theory and Simulation Theory (Stich and Nichols, 1995) and also suggests that (given an initial innate endowment), a simple constructivist model can arrange its representations by a process of development though experience. The model also gives useful predictions about children's performance under certain conditions. These are currently being followed up.

References

Cohen, J. D., Dunbar, K. & McClelland, J. L. (1991) "On the control of automatic processes: a parallel distributed processing account of the Stroop effect ", Psychological Review, 97, 332-361

Mitchell, P. (1996) "Acquiring a conception of mind", Hove: Psychology Press

Perner, J., Leekam, S., and Wimmer, H. (1987) "Three year olds' difficulty with false belief: the case for conceptual deficit", British Journal of Developmental Psychology. 5. 125-37

Perner, J., Ruffman, T. and Leekam, S. R. (1994) "Theory of mind is contagious: you catch it from your sibs", Child Development, 65, 1228-1238

Stich, S. P., and Nichols, S., (1992) "Folk Psychology: simulation or tacit theory?", Mind and Language, 7, 35-71