

Ideal Observers in Theory of Mind

Noah D. Goodman¹

Massachusetts Institute of Technology

Dec. 5, 2006

After a great deal of research, and significant insights, many foundational questions remain in the psychology of theory of mind. Why does theory of mind have such a striking developmental trajectory? How do adults solve the “inverse problem” to recover beliefs and desires from actions? How can we attribute mental states, even in the absence of actions? Perhaps most puzzling, why do we ascribe unobservable mental states to others in the first place?

Ideal observer models, the optimal solution to a given task, have proven to be a valuable tool to understand perception and learning, often explaining counterintuitive effects and illusions. Can ideal observers, and the related technique of rational analysis, help us to understand theory of mind? As an example, we discuss our recently proposed rational analysis of children’s false belief reasoning. Our analysis realizes a continuous, evidencedriven transition between two causal Bayesian models of false belief. Both models support prediction and explanation; however, one model is less complex while the other has greater explanatory resources. Because of this explanatory asymmetry, unexpected outcomes weigh more heavily against the simpler model. We tested this account empirically by showing children the standard outcome of the false belief task and a novel “psychic” outcome. As predicted from the analysis, we found children whose explanations and predictions were consistent with each model, and an interaction between prediction and explanation. We found that unexpected outcomes only induce children to move from predictions consistent with the simpler model to those consistent with the more complex one, never the reverse.

Despite achieving one of the goals of scientific modeling—motivating novel experimental work—this model has significant limitations: it only applies to a simple task in a very simple world. We conclude by considering the formal techniques needed to extend this type of model to more general situations, and the prospects for useful ideal observer analysis of observers of other agents.

¹Much of this work is joint with: Elizabeth Bara • Bonawitz, Chris L. Baker, Vikash K. Mansinghka, Alison Gopnik, Henry Wellman, Laura Schulz, and Joshua B. Tenenbaum.