

16-month-olds rationally infer causes of failed actions

Hyowon Gweon and Laura Schulz

Department of Brain and Cognitive Sciences,

Massachusetts Institute of Technology

Address correspondence to:

Hyowon Gweon or Laura Schulz

77 Massachusetts Avenue

Cambridge, MA 02139

hyora@mit.edu, lschulz@mit.edu

ABSTRACT

Sixteen-month-olds ($N = 83$) rationally use sparse data about the distribution of outcomes among agents and objects to solve a fundamental inference problem: deciding whether event outcomes are due to themselves or the world. When infants experienced failed outcomes, their causal attributions affected whether they sought help or explored.

To achieve our goals, we need to solve a fundamental inference problem: we need to distinguish our influence on event outcomes from the impact of the outside world. The distinction between attributions to the self and the world has been critical in disciplines ranging from social psychology (1) to artificial intelligence (2). The problem becomes urgent when our actions fail to achieve expected outcomes. If we try to turn a light and are left in the dark, did we do something wrong (e.g., flip the wrong switch), or is something wrong in the world (e.g., a bulb burned out)? These attributions have different implications for our subsequent actions. If we are the problem, we should change something about the agent (e.g., vary our actions or ask for help finding the switch); if the problem is external, we should try to change the world (or at least the light bulb). Consistent with empirical work showing that children draw accurate inductive inferences from minimal data (3, 4), we show that infants can use sparse evidence about the distribution of failed outcomes to answer the question “Is it me or the world?”

In Experiment 1, infants were seated next to a parent and shown toys that differed only in color (green, yellow, and red). The experimenter pushed a button on the green toy and the toy played music. She placed the red toy on a cloth near the infant, and handed the infant either the green (*Within-Object* condition) or yellow (*Between-Objects* condition) toy. As expected (5) all children pressed the button, and pressed equally often between conditions ($t(26) = 1.42, p = ns$). The toy never worked for the child.

To decide whether the problem lies with the agent or the object, infants should consider both the relative plausibility of the two hypotheses and the statistical evidence for each (6). In the *Within-Object* condition, neither hypothesis initially appears very probable: the infant might be doing something subtly wrong (e.g., not pressing hard enough), or something non-obvious might be wrong with the toy (e.g., it might have broken during the transfer). However, the statistical evidence favors the agent hypothesis: the outcome co-varies with the agent independent of the object. By contrast, in the *Between-Objects* condition, the statistical evidence is uninformative: the outcome co-varies with both the agent and object. Here however, the object hypothesis is the more plausible on prior grounds: while the infant’s actions are not obviously different from the experimenter’s, the toy clearly is. Moreover, there are now many ways the toy might have failed (e.g., the yellow toy might have broken at any point, or yellow toys might never work). As predicted, infants were more likely to try to change the agent (by handing the toy to their parents) than the object (by pulling the cloth or pointing to get the red toy) in the *Within* than *Between-*

Objects condition (All p-values $\leq .05$ by Fisher's Exact test: Change Agent vs. Change Object, *Within-Object*: 64.3% vs. 35.7%; *Between-Objects*: 21.4% vs. 78.6%).

These results suggest that infants rationally use sparse data to make causal attributions. However, other interpretations are possible. Infants who received the experimenter's toy might have been less likely to want a new toy than those who did not. Alternatively, infants in the *Within-Object* condition might have asked for help not because they attributed failure to themselves but because they inferred that the toy was broken and wanted the parent to fix it.

Experiment 2 addressed these possibilities. Infants were assigned to one of three conditions, identical to the *Within-Object* condition except as follows: *Within-Agent 1*: a single experimenter successfully activated the green toy twice and failed twice; *Within-Agent 2*: two experimenters each activated the green toy once and failed once (7), or *Between-Agents*: one experimenter activated the green toy twice and another experimenter failed twice. Children pressed the button equally often across conditions ($F(2,51) = 0.59, p = ns$) and the toy never activated.

These conditions differ only with respect to the statistical evidence. The outcomes in the *Within-Agent* conditions (considering also the infant's failure) vary independent of the agent, suggesting the failure is due to the object; the outcomes in the *Between-Agents* condition co-vary with the agent, independent of the object, suggesting the failure is due to the agent. As predicted, infants were more likely to first change the agent than the object in the *Between-Agents* than *Within-Agent* conditions (Change Agent vs. Change Object, *Within-Agent 1*: 31.6% vs. 68.4%; *Within-Agent 2*: 29.4% vs. 71.6%; *Between-Agents*: 68.4% vs. 31.6%).

Consistent with formal models of causal induction (6), these results suggest that infants track the statistical dependence between agents, objects, and outcomes and can use minimal data to draw inferences that support rational action. When the infants inferred that they were the source of failure, they sought help; when they believed the failure was due to their object, they explored others. Seeking instruction and engaging in exploration are both potentially effective strategies for learning; infants' differential response to failure depending on the evidence for its causes augurs well for their success.

References and Notes

1. Kelley, H. H. *Attribution theory in social psychology*. Nebraska Symposium on Motivation, **15**, 192 (1967)

2. S. J. Russell, P. Norvig, *Artificial intelligence: a modern approach*. (Prentice hall, Upper Saddle River, NJ, 2009).
3. Gopnik, A., Schulz, L.E., *Trends in Cog. Sci.*, **8**, 217 (2004).
4. Gweon, H., Tenenbaum, J., Schulz, L.E., *PNAS.*, **107**, 9066 (2010).
5. D. A. Baldwin, E. M. Markman, R. L. Melartin, *Child Dev.* **64**, 711 (1993).
6. T. Griffiths, J. Tenenbaum, *Psych Rev.* **116**, 661 (2009).
7. The two *Within-Agent* conditions provided internal replication; no differences were predicted between these conditions.

Acknowledgments

We are grateful to Charles Jennings, Rebecca Saxe, and Joshua Tenenbaum for helpful comments on the draft. The research was funded by an NSF Faculty Early Career Development award, a John Templeton Foundation Award, and a James S. McDonnell Foundation

Figure 1. Design and results, Experiments 1 (A: *Within-Object*, B: *Between-Objects*) and 2 (C: *Within-Agent* 1, D: *Within-Agent* 2, E: *Between-Agents*). P: Parent, E1 & E2: Experimenters 1 and 2. G, Y, and R refer to toy colors: Green, Yellow, and Red. The toy on the graph indicates the toy handed to the infant.