16-Month-Olds Rationally Infer Causes of Failed Actions

Hyowon Gweon and Laura Schulz

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 on 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 (Fig. 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 the yellow (between-objects condition) toy. As expected (*5*), all children pressed the button and pressed equally often between conditions [$t(26) = 1.42$, P = not significant (ns) (*6*)]. 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 (*7*). 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 nonobvious 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 covaries with the agent independent of the object. By contrast, in the between-objects condition, the statistical evidence is uninformative: The outcome covaries with both the agent and the object. Here, however, the object hypothesis is the more plausible on prior grounds: Although 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 ≤ 0.05 by Fisher's exact test: change agent versus change object, within-object, 64.3% versus 35.7%; between-objects, 21.4% versus 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 (*8*); 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.

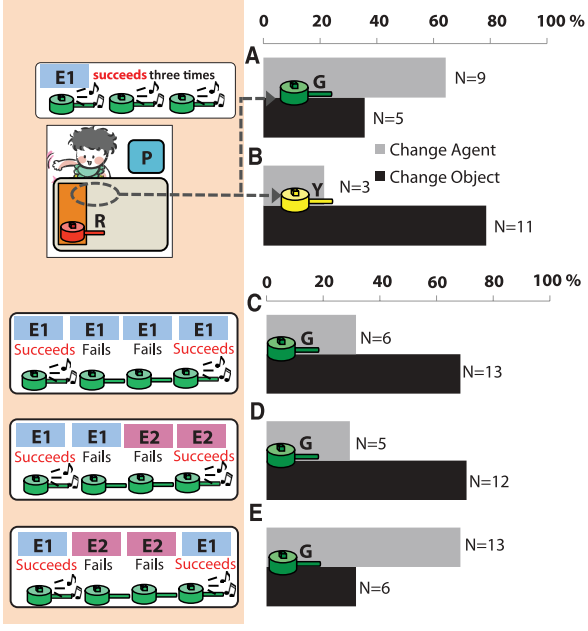

Fig. 1. Design and results. Experiments 1 [(**A**) within-object and (**B**) between-objects] and 2 [(**C**) within-agent 1, (**D**) within-agent 2, and (**E**) between-agents]. P indicates parent; E1 and E2, experimenters 1 and 2; G, Y, and R refer to toy colors: green, yellow, and red. The toy on the graph indicates the toy handed to the infant.

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 covary 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 versus change object, within-agent 1, 31.6% versus 68.4%; within-agent 2, 29.4% versus 71.6%; between-agents, 68.4% versus 31.6%).

Consistent with formal models of causal induction (*7*), 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. H. H. Kelley, *Nebr. Symp. Motiv.* **15**, 192 (1967).
2. S. J. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach* (Prentice Hall, Upper Saddle River, NJ, 2009).
3. A. Gopnik, L. E. Schulz, *Trends Cogn. Sci.* **8**, 371 (2004).
4. H. Gweon, J. B. Tenenbaum, L. E. Schulz, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9066 (2010).
5. D. A. Baldwin, E. M. Markman, R. L. Melartin, *Child Dev.* **64**, 711 (1993).
6. Materials and methods are available as supporting material on *Science* Online.
7. T. L. Griffiths, J. B. Tenenbaum, *Psychol. Rev.* **116**, 661 (2009).
8. The two within-agent conditions provided internal replication; no differences were predicted between these conditions.

Acknowledgments: We are grateful to C. Jennings, R. Saxe, and J. 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 award.

Supporting Online Material

www.sciencemag.org/cgi/content/full/332/6037/1524/DC1
Materials and Methods
Movie S1

17 February 2011; accepted 16 May 2011
10.1126/science.1204493

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. E-mail: hyora@mit.edu (H.G.); lschulz@mit.edu (L.S.)