Current Directions in Psychological Science
2015, Vol. 24(2) 87–92

DOI: 10.1177/0963721414556653
cdps.sagepub.com

When Younger Learners Can Be Better (or at Least More Open-Minded) Than Older Ones

Alison Gopnik[1], Thomas L. Griffiths[1], and Christopher G. Lucas[2]
[1]Department of Psychology, University of California, Berkeley, and [2]School of Informatics, University of Edinburgh, United Kingdom

Abstract
We describe a surprising developmental pattern we found in studies involving three different kinds of problems and age ranges. Younger learners are better than older ones at learning unusual abstract causal principles from evidence. We explore two factors that might contribute to this counterintuitive result. The first is that as our knowledge grows, we become less open to new ideas. The second is that younger minds and brains are intrinsically more flexible and exploratory, although they are also less efficient as a result.

Keywords
cognitive development, causal learning, Bayesian models, simulated annealing

There is a tension in the field of cognitive development. Children perform worse than adults on many measures. As they grow older, children become more focused, they plan better, and, of course, they know more. Yet very young children are prodigious learners, and they are especially good at learning about causes. Preschoolers, toddlers, and even infants construct everyday causal theories about objects, living things, and minds (e.g., Wellman & Gelman, 1992; Gopnik & Meltzoff, 1997). How can the youngest children learn so much so quickly and accurately when their knowledge and cognitive abilities seem so limited?

We suggest that the apparent limitations in children's knowledge and cognitive abilities may actually sometimes make them better learners. Empirically, we have recently found a similar pattern across different problems and age ranges. Younger learners are, surprisingly, better than older ones at inferring unlikely or unusual abstract causal hypotheses from evidence.

There are some other examples of this counterintuitive developmental pattern. Younger infants can learn distinctions between sounds that are not used in their native language better than older infants and adults (Kuhl, 2004; Werker, Yeung, & Yoshida, 2012), and younger children are better at generating alternative uses for a tool than older children (Defeyter & German, 2003). These findings also suggest that younger learners might sometimes be open to more possibilities than older ones.

Theoretically, we propose two possible complementary explanations for this pattern, inspired by viewing children's learning through the lens of computer science. Younger learners may do better because they are less biased by their existing knowledge, or because their brains and minds are inherently more flexible.

Empirical Studies

Many studies have shown that children as young as 15 months old can learn specific cause-effect relationships from statistical data (Gopnik et al., 2004; Gopnik & Schulz, 2007; Gopnik & Wellman, 2012; Gweon & Schulz, 2011). These studies have shown the typical developmental pattern—either younger and older children perform similarly, or older children do better. In the new studies we describe here, we investigated whether children can use patterns of data to infer more abstract, general causal principles, or *overhypotheses*—that is,

Corresponding Author:
Alison Gopnik, Department of Psychology, 3210 Tolman Hall #1650, University of California, Berkeley, Berkeley, CA 94720-1650
E-mail: gopnik@berkeley.edu

hypotheses about which kinds of more specific hypotheses are likely (Griffiths & Tenenbaum 2007; Kemp, Perfors, & Tenenbaum, 2007).

For example, suppose you observe that stomachaches are caused by eating bad food, rashes by touching weeds, and coughs by inhaling pollen. You might form the overhypothesis, or "framework theory" (Gopnik & Wellman, 2012), that illnesses have biological causes. When you then seek the cause of a new illness, such as AIDS, you might think biological causes such as viruses, bacteria, or genes are more likely than psychological causes such as anxiety.

In the kind of studies we report here, learners see a series of events and choose between two abstract hypotheses, A and B, that could explain those events. Hypothesis A initially seems less likely than B, at least from the adult perspective, but it is better supported by the evidence the learner has seen. Younger learners turn out to be more likely to infer A than older learners, who, despite the data, are more likely to stick with B.

The first study exhibiting this pattern explored how preschoolers learn high-level principles of social cognition (Seiver, Gopnik, & Goodman, 2013). Adults in Western cultures believe that actions are caused by personal traits that are stable over time but differ among individual people, such as bravery or timidity. Western adults explain what people do in terms of such traits even when the evidence shows that people are actually reacting to particular situations—in other words, they have a "trait bias" (Kelley, 1967). We gave 4- and 6-year-old children statistical evidence that supported either a trait or situation explanation (Seiver et al., 2013). In the "person" condition, a character called Sally (represented by a doll) was usually willing to play on both a skateboard and a diving board (represented by miniature toys), while a character called Josie usually avoided both toys. This data pattern supported the hypothesis that something about Sally or Josie caused them to approach or avoid the toys. In the "situation" condition, neither character approached the skateboard, though both approached the diving board, supporting the hypothesis that something about the toys caused the action pattern. In a control condition, the data supported both the "trait" and "situation" hypotheses equally. Then we asked children why each character approached or avoided the toy.

Four-year-olds accurately inferred the right kind of cause from the data (Fig. 1). When the data supported a personal-trait explanation, the children did, too, often inventing trait-like causes (e.g., "Josie's the little sister, and Sally's the big one"). But they also said that the character acted because of the situation when that fit the data—"it looks scary" or "it looks fun." Six-year-olds, in contrast, did much worse in the situation condition. Like adults, they showed a strong bias toward trait explanations even when the evidence did not support them.

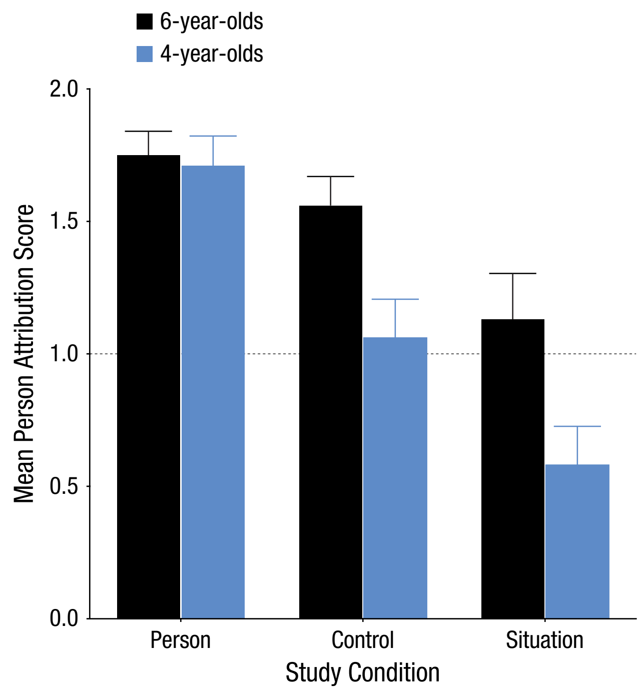

Fig. 1. Mean number of person (vs. situation) attributions (out of two) among 4- and 6-year-olds as a function of condition (Seiver, Gopnik, & Goodman, 2013). Four-year–olds correctly explained actions in terms of personal traits in the person condition and in terms of situations in the situation condition. In a control condition, which supported both types of attributions equally, they were equally likely to choose either attribution (indicated by the dashed horizontal line). Six-year-olds showed a marked bias toward personal-trait explanations in the control and situation conditions. Error bars show standard errors.

Notably, the children's inferences extended beyond these particular dolls and toys. Their explanations invoked more general principles—older sisters are better than younger ones at many skills; people are unlikely to play with anything that looks scary. We also asked them to make predictions about new actors and situations. In the person condition, all the children said that the brave character would also be brave if she faced a new situation, such as jumping on a trampoline. In the situation condition, 4-year-olds followed the data and predicted that Mary, a new character, would also be scared by the skateboard but not the diving board. However, 6-year-olds thought the character would act the same in both situations, in spite of the data, consistent with a trait bias.

In another series of studies, participants had to infer an abstract principle about a machine that played music when you put some combinations of blocks on top of it and not others (Lucas, Bridgers, Griffiths, & Gopnik, 2014). The machine could work on an "individual" principle, such that some individual blocks made the machine work and some did not—each cause did or did not lead to the effect. Adults assume that causal systems work this way (Cheng, 1997), just as they assume that actions are caused by traits. But the

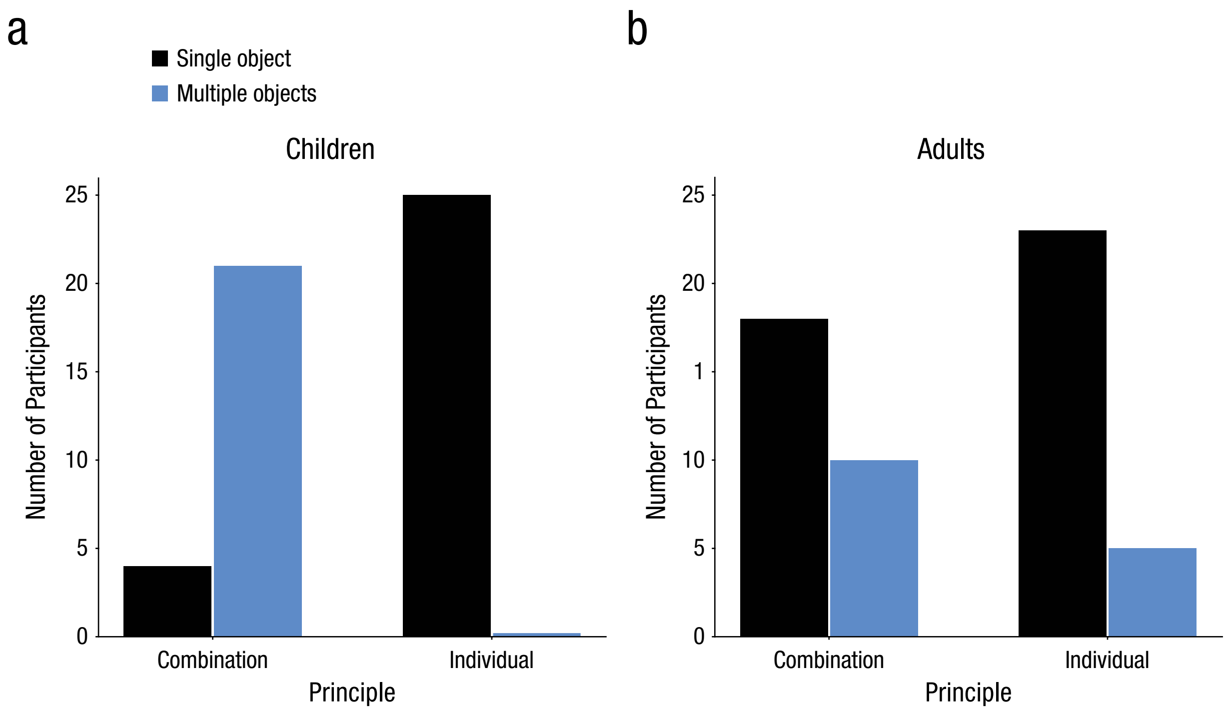

Fig. 2. Children's and adults' choices of objects to activate a machine after seeing evidence that a machine operates according to an individual or combination principle (Lucas, Bridgers, Griffiths, & Gopnik, 2014). When both age groups saw evidence for a combination principle, meaning that two or more objects were necessary to activate the machine, only children tended to choose multiple objects to activate the machine (a). In contrast, adults tended to choose only one object, despite the evidence (b). When both age groups saw evidence for an individual principle, meaning that only one object was necessary to activate the machine, both children and adults tended to choose a single object to place on the machine, consistent with the evidence.

machine could also work on a more unusual "combination" principle, such that causes had to be combined to produce an effect: Some two-block combinations made the machine go, though individual blocks did not.

We showed 4-year-old children and adults an unambiguous pattern of events that supported one principle or the other. Then they saw an ambiguous pattern with a new set of blocks, which could be consistent with either the "individual" or the "combination" principle. Then we asked them to activate the machine.

If the machine worked on the combination principle, multiple blocks would be necessary to make it go; a single block should suffice on the individual principle. Again, children had to generalize beyond particular hypotheses about which specific block combinations made the machine go and infer a general principle about how the machine worked.

Preschoolers correctly learned both the individual and combination principles from the unambiguous examples and used them to interpret the ambiguous new data and design the right action (Fig. 2a). The adults stuck with the individual principle even when the evidence weighed against it—they continued to place individual blocks on the machine even in the combination condition (Fig. 2b).

The third study looked at a different kind of abstract causal principle. Older children (and nonhuman primates) have difficulty with higher-order relational concepts such as "same" and "different" (Gentner, 2010; Penn, Holyoak, & Povinelli, 2008). Chimpanzees quickly learn that a square stimulus leads to a reward while a round one does not, but they need hundreds of trials to learn that a reward follows when two stimuli are the same rather than different.

We gave 18- to 30-month-olds a causal higher-order relation problem (Walker & Gopnik, 2013, 2014). A machine played music when an experimenter put two similar blocks on it but not when she put two different blocks on, or vice versa. Toddlers then had to choose between two novel pairs of blocks—one pair of two similar blocks and one pair of two different blocks—to activate the machine.

Surprisingly, these toddlers were adept at the task, in contrast to the older children in previous studies. Then we gave 3-year-olds exactly the same task as the toddlers. They performed at chance level. Further studies showed that this was because they assumed that the individual objects, rather than the relations between them, would activate the machine, in spite of the data.

So, the same counterintuitive pattern emerged across all three studies. But why would children perform worse as they grow older? This is still an open question, but we propose two potential explanations below.

A lot of knowledge can be a dangerous thing

First, the very fact that older learners know more may make it more difficult for them to learn something new. Once a learner has inferred a general principle (e.g., that people act because of their traits, or that individual objects, rather than combinations of objects or relations between them, have causal powers), that principle may constrain his or her interpretation of new data. Causal relationships that conflict with that principle may then be more difficult to learn.

Probabilistic-model-based approaches to cognitive development can provide a more precise version of this idea (for more on such approaches see, e.g., Gopnik & Tenenbaum, 2007; Gopnik, 2012; Gopnik & Wellman, 2012; Kushnir & Xu, 2012; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). A Bayesian learner assesses how likely various hypotheses are, given a pattern of new data. Learners do this by using Bayes's rule to combine two probabilities. One is the prior probability of any particular hypothesis—how likely the hypothesis was before the learner saw the data. The other is the likelihood—how likely it was that that hypothesis would have generated the new data.

As a result, if the prior probability distribution strongly favors one hypothesis—that is, if the learner initially thinks that hypothesis A is much more likely than B—then the learner will need more evidence to overturn A and accept B instead. If the prior is "flat"—that is, if the learner initially thinks that A and B are equally likely—then the learner will require less evidence to accept B.

In an extension of this idea, called *hierarchical Bayesian learning* (Griffiths & Tenenbaum, 2007), data at a more specific level, like the relations between stomachaches and food, can be used to learn a higher-level principle—in this case, the overhypothesis that illnesses have biological causes. This kind of learning might explain the counterintuitive pattern in our studies.

From flexibility to efficiency

Another factor may be that as children grow older, there are changes in the way they learn that make them intrinsically less flexible and less able to attend to unusual possibilities. There are complementary computational, neuroscientific, and evolutionary reasons for thinking this might be true.

A Bayesian learner, whether that learner is a child or a computer, must have some technique for searching through the vast space of possible hypotheses and trying to find the most likely option. Recent studies have explored the search methods children might use (e.g., Bonawitz, Denison, Griffiths, & Gopnik, 2014; Denison, Bonawitz, Gopnik, & Griffiths, 2013).

Using an analogy to physics, computer scientists talk about different search "temperatures." In "high-temperature" searches, the learner searches broadly but is less likely to "settle" on any one answer for long—the learner bounces widely around in the space of hypotheses like a molecule bouncing around in a hot liquid.

From a Bayesian perspective, raising the temperature of a search will have an effect equivalent to "flattening" the prior—initial differences among hypotheses will make less of a difference. In addition, however, it will have the effect of weakening the likelihoods.

High-temperature searches are wide ranging but very variable, and the learner can move away from good hypotheses as well as bad ones. Low-temperature searches are more likely to quickly lead to "good enough" hypotheses. However, the learner risks getting stuck in a "local minimum"—passing up potentially better but more unusual hypotheses that are further away from his or her initial guess.

One way to compromise between the advantages and drawbacks of high and low temperature is to start with a high-temperature search and gradually "cool off." This is called *simulated annealing* in computer science, by analogy to the heating and cooling that leads to robustness in metallurgy (Kirkpatrick, Gelatt, & Vecchi, 1983). By beginning with a high-temperature search, a learner can explore the possibilities more widely before focusing more narrowly on the likely candidates.

If children initially perform high-temperature searches and gradually "cool off" to perform low-temperature ones as they grow older, this might explain why younger learners sometimes infer unusual hypotheses better than older learners. How could we discriminate between this simulated-annealing idea and the related flat-prior idea? In Lucas et al. (2014), we included a "baseline" condition. Participants in this condition saw only the ambiguous events—they never saw the unambiguous new data that pointed to each principle. If adults initially thought that the "individual" hypothesis was more likely than the "combination" hypothesis, and children did not, that should have been reflected in this baseline condition. But, in fact, both children and adults preferred the "individual" hypothesis initially. The difference seemed to be that children were more willing to switch to the alternative hypothesis. A Bayesian model consistent with the annealing possibility matched children's judgments. However, more studies of the dynamics of learning are necessary to distinguish these possibilities.

Findings in neuroscience also mesh well with the annealing idea (e.g., Thompson-Schill, Ramscar, & Chrysikou, 2009). An early period of neural flexibility and plasticity is succeeded by a more narrow and inflexible, though more efficient, set of procedures. In particular, as children get older, frontal areas of the brain exert more control over other areas. This frontal control is associated with focused attention and better planning and executive control. However, this control has costs. Empirically, disruptions to frontal control, resulting in a more "child-like" brain, can actually lead to better performance in cognitive tasks that involve exploring a wide range of possibilities (e.g., Chrysikou et al., 2013). There may be an intrinsic trade-off between exploitation and exploration—between swift, focused, efficient adult action and wide-ranging, exploratory child-like learning.

A pattern of early cognitive exploration also makes sense from an evolutionary perspective. Across many species, flexibility, brain size, and intelligence are associated with a long, protected period of immaturity—a long childhood. Human beings have the largest brains, the most flexible intelligence, and the longest childhood of any species. One explanation for this distinctive life history is that an early protected period allows young organisms to explore possibilities in an unconstrained way. This early exploratory learning, in turn, allows learners to act more effectively when they grow older (Buchsbaum, Bridgers, Weisberg, & Gopnik, 2012). Childhood may be evolution's way of performing simulated annealing.

Adults may sometimes be better at the tried and true, while children are more likely to discover the weird and wonderful. This may be because as we get older, we both know more and explore less.

Recommended Reading

Gopnik, A. (2012). (See References). An accessible short review of recent work on probabilistic models of cognitive development.

Gopnik, A., & Wellman, H. M. (2012). (See References). An extensive, thorough, accessible review of the theoretical and empirical work on Bayesian causal models as a constructivist account of cognitive development, which includes a gentle nontechnical instructional tutorial explaining how the models actually work and discussion of the idea of search and developmental change.

Kushnir, T., & Xu, F. (Eds.). (2012). (See References). An edited volume including chapters by a wide variety of researchers applying probabilistic models to a very wide range of problems and domains.

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). (See References). The source of much of the empirical work described here, including several more experiments and an explanation and formal model relevant to the annealing ideas.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). (See References). A general and accessible review of probabilistic models and Bayesian inference in cognitive science.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

The research reported here was supported by the McDonnell Foundation Causal learning collaborative initiative and National Science Foundation Grants BCS-1023875 and BCS-331620.

References

Bonawitz, E., Denison, S., Griffiths, T., & Gopnik, A. (2014). Probabilistic models, learning algorithms, response variability: Sampling in cognitive development. *Trends in Cognitive Sciences*. doi:10.1016/j.tics.2014.06.006

Buchsbaum, D., Bridgers, S., Weisberg, D. S., & Gopnik, A. (2012). The power of possibility: Causal learning, counterfactual reasoning, and pretend play. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 2202–2212. doi:10.1098/rstb.2012.0122

Cheng, P. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*, 367–405.

Chrysikou, E. G., Hamilton, R., Coslett, H., Datta, A., Bikson, M., & Thompson-Schill, S. (2013). Noninvasive transcranial direct stimulation over left prefrontal cortex facilitates cognitive flexibility in tool use. *Cognitive Neuroscience*, *4*, 81–89.

Defeyter, M. A., & German, T. (2003). Acquiring an understanding of design: Evidence from children's insight problem solving. *Cognition*, *89*, 133–155.

Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T. L. (2013). Rational variability in children's causal inferences: The Sampling Hypothesis. *Cognition*, *126*, 285–300. doi:10.1016/j.cognition.2012.10.010

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*, 752–775. doi:10.1111/j.1551-6709.2010.01114.x

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science*, *337*, 1623–1627. doi:10.1126/science.1223416

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 3–32. doi:10.1037/0033-295X.111.1.3

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Gopnik, A., & Schulz, L. (2007). *Causal learning: Psychology, philosophy, and computation*. New York, NY: Oxford University Press.

Gopnik, A., & Tenenbaum, J. B. (2007). Bayesian networks, Bayesian learning and cognitive development. *Developmental Science*, *10*, 281–287. doi:10.1111/j.1467-7687.2007.00584.x

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*, 1085–1108. doi:10.1037/a0028044 1085-1108

Griffiths, T. L., & Tenenbaum, J. B. (2007). Two proposals for causal grammars. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation*. Oxford, England: Oxford University Press.

Gweon, H., & Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science*, *332*, 1524. doi:10.1126/science.1204493

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 15, pp. 192–238). Lincoln: University of Nebraska Press.

Kemp, C., Perfors, A., & Tenenbaum, J. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*, 307–321.

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*, 671–680.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–843.

Kushnir, T., & Xu, F. (Eds.). (2012). *Rational constructivism in cognitive development* (Vol. 43). New York, NY: Academic Press.

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, *131*, 284–299. doi:0.1016/j.cognition.2013.12.010

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral & Brain Sciences*, *31*, 109–130. doi:10.1017/S0140525X08003543

Seiver, E., Gopnik, A., & Goodman, N. D. (2013). Did she jump because she was the big sister or because the trampoline was safe? Causal inference and the development of social attribution. *Child Development*, *84*, 443–454. doi:10.1111/j.1467-8624.2012.01865.x

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279–1285. doi:10.1126/science.1192788

Thompson-Schill, S. L., Ramscar, M., & Chrysikou, E. G. (2009). Cognition without control: When a little frontal lobe goes a long way. *Current Directions in Psychological Science*, *18*, 259–263. doi:10.1111/j.1467-8721.2009.01648.x

Walker, C. M., & Gopnik, A. (2013, October). *Infants infer higher-order relational principles in causal learning*. Paper presented at the Cognitive Development Society Meeting, Memphis, TN.

Walker, C. M., & Gopnik, A. (2014). Toddlers infer higher-order relational principles in causal learning. *Psychological Science*, *25*, 161–169. doi:10.1177/0956797613502983

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, *43*, 337–375. doi:10.1146/annurev.ps.43.020192.002005

Werker, J. F., Yeung, H. H., & Yoshida, K. (2012). How do infants become experts at native speech perception? *Current Directions in Psychological Science*, *21*, 221–226.