

Meltzoff, A. N., & Gopnik, A. (2013). Learning about the mind from evidence: Children's development of intuitive theories of perception and personality. In S. Baron-Cohen, H. Tager-Flausber, & M. Lombardo (Eds.), *Understanding other minds* (3rd ed., pp. 19-34). Oxford, England: Oxford University Press.

Chapter 2

Learning about the mind from evidence: Children's development of intuitive theories of perception and personality

Andrew N. Meltzoff and Alison Gopnik

Where does our understanding of the mind come from? Different theoretical perspectives have different views on this question. Strong modularity and core knowledge theories (e.g. Leslie, 2005) propose that the essentials of our adult understanding of others are in place initially, and development involves relatively small changes in that knowledge around the edges. Strong “embodiment” and “resonance” theories (e.g. simulation and mirror-neuron based accounts of mindreading, such as Gordon, 1996 or Gallese & Goldman, 1998) also do not focus on developmental change and argue that our understanding of the mind is fundamentally not inductive. Rather than learning about the mind from evidence, both these views see our understanding as due to relatively automatic and specialized triggering or resonance processes. We “take on” the mental states of others or project our own experiences on to them—rather than inferring those states from evidence.

In contrast, “theory-theory” accounts (Gopnik & Meltzoff, 1997; Gopnik & Wellman, 1994) propose that our understanding of the mind, at least in large part, involves learning abstract causal structures from evidence—hence, the analogy to theory change in science—involving initial hypotheses, tests, and conceptual revision. In the past, these claims were largely made on the basis of the naturally occurring changes in children's understanding of the mind over time. Moreover, it was unclear just what kinds of learning mechanism would allow children to learn about a complex and invisible system like the mind so swiftly and effectively.

In the past 10 years, however, this has begun to change. First, there are results from training studies with young children, which show that providing evidence can lead to changes in children's understanding of the mind. For example, Amsterlaw and Wellman (2006) and Slaughter and Gopnik (1996) both showed that three-and-a-half-year-olds who received evidence about beliefs shifted to a new understanding of belief more quickly than those who did not. Importantly, this extended not only to their performance on the classic false-belief task, but to their understanding of related concepts like the appearance/reality distinction and the sources of beliefs. Interestingly, children showed this effect most clearly when they were asked to explain, rather than just describe the evidence. Moreover, naturally occurring variations in the availability of evidence children receive can change the timing of their belief understanding. For example, deaf children of hearing parents have markedly delayed false-belief understanding (see Gopnik & Wellman, 2012, for a review).

Even with these training effects, however, we might argue that the incoming evidence simply accelerates or delays a naturally occurring change. A more powerful demonstration of the role of evidence comes when we design experiments in which we systematically give infants or children different kinds of new evidence about a system and see what kinds of inferences they draw. This

has been the approach taken in both the statistical learning literature and the causal learning literature. When children are provided new patterns of evidence under experimental control, and the different patterns of evidence lead them to different conclusions, it seems more obvious that the evidence itself is doing the causal work. For the most part, however, this work has focused on children's learning of language (e.g. Kuhl, 2004; Saffran, Aslin, & Newport, 1996), physical properties of objects (Wu, Gopnik, Richardson, & Kirkham, 2011), or physical causal relations (e.g. Bonawitz, Lim, & Schulz, 2007; Gopnik, Glymour, Sobel, Schulz, Kushnir, & Danks, 2004; Meltzoff, Waismeyer, & Gopnik, 2012; Sobel & Kirham, 2006), rather than on their psychological learning. The experiments we discuss in this chapter move beyond this to examine infants' and young children's developing understanding of other people's minds.

Probabilistic models and Bayesian learning

In parallel with the field accumulating new data, theoretical work over the last 10 years (e.g. Gopnik, 2012; Gopnik et al., 2004; Meltzoff, Kuhl, Movellan, & Sejnowski, 2009; Tenenbaum, Kemp, Griffiths, & Goodman, 2011) has shown increasingly that it is possible to specify more precisely and formally how children learn from evidence. In particular, within the framework of probabilistic models and Bayesian inference we can think of children's learning as a process of hypothesis testing and revision (Gopnik, 2012; Tenenbaum et al., 2011). Children use probabilistic models to generate structured hypotheses, then test and revise those theories in a systematic way based on evidence. Moreover, rather than simply generating a yes or no decision about whether a particular hypothesis is true, Bayesian inference considers multiple hypotheses and assigns probabilities to those hypotheses. Bayesian methods let you determine the probability of possibilities. The integration of prior knowledge and new evidence in Bayesian reasoning also gives Bayesian inference a characteristic combination of stability and flexibility—a learner will be reluctant to give up a strongly-confirmed hypothesis, but even the most entrenched idea can be rejected if enough counter-evidence accumulates.

Moreover, according to the theory-theory view, children often are not just learning particular causal relations but are also learning abstract generalizations about causal structure. In fact, empirical research has shown that children develop more abstract, framework knowledge over and above their specific causal knowledge. For example, children may know in general that actions are caused by beliefs and desires without being able to say exactly which beliefs and desires are involved in any particular case.

These broader generalizations are important in both scientific and intuitive theories. Philosophers of science refer to “over-hypotheses” (Goodman 1955), or “research programs” (Laudan 1977), or “paradigms” (Kuhn 1962) to capture these higher-order generalizations. Cognitive developmentalists have used the term “framework theories” (Carey 2009; Wellman 1990; Wellman & Gelman 1992). For example, in their framework theories, children assume there are different kinds of variables and causal structure in psychology vs. biology vs. physics. In fact, they often understand these abstract regularities before they understand specific causal relationships (e.g. Simons & Keil, 1995).

Some nativists argue that this must mean that the more abstract causal knowledge is innate. In contrast, constructivists, including Piaget and theory theorists, hold that this abstract causal knowledge could be learned. How could this be?

Griffiths and Tenenbaum (2007, 2009; Tenenbaum, et al., 2011), inspired by both philosophy of science and cognitive development, have formulated computational ways of representing and learning higher-order generalizations about causal structure. Following Gelman, Carlin, Stern,

& Rubin (2003), they call their approach hierarchical Bayesian modeling (HBM) or, sometimes, theory-based Bayesian modeling. The idea is to have meta-representations, that is, representations of the structure of particular causal hypotheses, and of the nature of the variables and relationships involved in those causal networks. These higher-level beliefs can constrain the more particular causal hypotheses. Moreover, these higher-level generalizations can themselves be learned. HBMs stack up hypotheses at different levels. The higher levels contain general principles that specify which hypotheses to entertain at the lower level.

Computational work on HBMs has shown that, at least normatively, hierarchical Bayesian learning can actually work. Higher-level framework theories can, indeed, be updated in a Bayesian way via evidence that contacts only lower level hypotheses. Griffiths and Tenenbaum (2007) provide several simple demonstrations; Kemp, Perfors, & Tenenbaum (2007) and Goodman, Ullman, & Tenenbaum (2011) provide more comprehensive and complex ones. These demonstrations show that it is possible, in principle, to learn to proceed at several levels at once—not just at the level of specific hypotheses, but also at the level of specific theories and, even more abstractly, at the framework theory level.

Probabilistic models in development

Probabilistic models were originally designed to be ideal rational accounts of how a scientist or a computer could best solve a learning problem. They also have attractions as theories of the learning mechanisms of cognitive development. One attraction is that, at least in principle, this kind of learning would allow children to gradually move from one structured hypothesis to another very different hypothesis based on patterns of evidence—children would not be restricted to making small tweaks to innate modules or to simply accumulating new data. The probabilistic nature of Bayesian inference also captures the often gradual and piecemeal way that development proceeds. At the same time, the generative power of structured models and hypotheses might help explain the abstract and general character of children’s inferences.

In addition, the probabilistic models view gives us a new way to think about the innate bases of cognition. Rather than thinking about innate perceptual-cognitive structures as firm “constraints” on the kinds of knowledge that a human can develop, an innate “prior” might weigh certain hypotheses as more likely than others, but even these hypotheses could be overturned with sufficient counter-evidence. The work on hierarchical Bayesian learning (Griffiths & Tenenbaum, 2007) suggests that “priors” may not only take the form of specific hypotheses about particular causal relationships, but may involve broader “framework principles” about general theoretical categories and causal relations. These framework principles shape many more specific hypotheses, but they may themselves be overturned with sufficient counter-evidence.

Developmental changes in understanding the mind

We suggest that in terms of our understanding of the mind, a strong prior and innate “framework principle” is that our own mental states and those of others are likely to be similar. We can think of this as a Bayesian version of the “like-me” hypothesis that we have argued for in the past (Meltzoff, 2007, 2013; Meltzoff & Gopnik, 1993). This assumption shapes the human infants’ early learning about the mind, allowing a framework for preferring some hypotheses to others. It is, however, only the beginning of our learning about the mind. Within the framework principle, we can use evidence to elaborate on our initial understanding in complex and abstract ways. Eventually, with accumulating evidence concerning differences between our own perceptions, desires, and beliefs, and those of others, we can revise or overturn that framework principle, as shown by developmental research (e.g. Gopnik & Wellman, 2012; Moll & Meltzoff, 2011; Repacholi & Gopnik, 1997).

In this chapter, we take the problem of *developmental change* in children’s understanding of the mind to be central, and illustrate the foregoing ideas with two examples. The field’s (over) concern with the shift in children’s verbal reasoning about false belief at 3–4 years of age has obscured the important fact that the human intuition of how the mind works is an extended process, including significant changes both much earlier and later than the classic preschool shift. We consider both an earlier and a later set of developments. In both cases, we show that providing children with particular patterns of evidence, whether evidence from their own experience or about the behavior of others, can lead to novel and systematic new causal models of how the mind works. Moreover, in both cases we invoke the idea of a Bayesian framework principle. The first example concerns infants’ early understanding of other people’s visual perception. We suggest that the initial framework principle adopted preverbally, and perhaps present at birth, constrains inferences, but is itself influenced by evidence. In the second example, the development of an understanding of personality traits, we suggest that this higher-order principle is actually initially inferred from data, but then acts as a constraint on further inferences.

Understanding perception

Recent studies show that infants use first-person visual experiences as evidence for a new understanding of the perceptions of others. The research is built from the finding that young infants make a puzzling error. In gaze-following studies, 12-month-olds follow a person’s line of regard to an external object even when a blindfold occludes that person’s viewpoint. They do not make this error, however, when the person closes his eyes (Brooks & Meltzoff, 2002, 2005). Why do young infants seem to have a privileged understanding of eye closing over and above other blindfolds?

One idea is that infants have extensive evidence about the causal relation between eye closure and visual experience, but initially have much less evidence about other kinds of occlusion. Eye closure is a biological motion with which infants have extensive first-person, agentive experience. Even very young infants have strong evidence about the causal relation between whether their eyes are open or closed, and their visual experience. They can easily perform informal “tests” to assess this causal link—they can control their own vision by closing and opening their eyes. When they close their eyes, the visual world goes black, and when they open them the world pops back into view. Perhaps infants use this evidence, along with their initial “like-me” causal framework principle, to make the attribution about others’ visual experiences. What applies to me, also applies to you.

This predicts that if infants are given systematic evidence that blindfolds block their own view, they should suddenly make different attributions to others. Meltzoff and Brooks (2008) tested this idea with 12-month-olds. Infants were randomly assigned to three experimental groups that differed only in the nature of the evidence provided to the infants. Infants in the key treatment group were given massive experience with blindfolds (see Figure 2.1). When the infant looked at a toy, the adult blocked the view with a blindfold. She then lowered it in a playful manner, only to repeat the cycle for the next toy the child fixed. Infants experienced that their own view was blocked, but they were given no training about the adult’s viewpoint. A control group involved a cloth made from the same material as the blindfold, but with a small window cut out of the center. Infants in this control received the same protocol (controlling for cloth raising/lowering); however, they could peer through the windowed cloth. In a second control group, infants were familiarized with the opaque cloth while it was laying flat on the table.

At the end of training, all three groups were given a standard gaze-following test. Infants were confronted with a blindfolded adult who turned toward the distal objects. Infants who had received first-person training on the opaque blindfold responded in a completely different manner to the

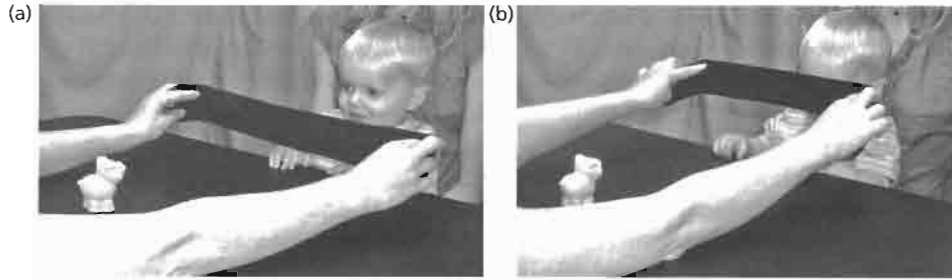


Figure 2.1 A 12-month-old boy in Meltzoff and Brooks' (2008) training procedure. Infants randomly assigned to a treatment group were given self-experience that a blindfold occluded their own perception. Infants looked at an interesting object (a). The blindfold then blocked their view of the toy (b). This was repeated over an 8-min training session. This first-person experience changed infants' interpretation of how opaque visual occluders influence other people's vision. See text for details.

Reproduced from Meltzoff, A.N. & Brooks, R. Self-experience as a mechanism for learning about others: A training study in social cognition. *Developmental Psychology*, 44, 1257–65. ©2008, American Psychological Association.

controls. Infants in this treatment group did not turn when the adult wore the blindfold, but infants in the controls still mistakenly followed the blindfolded adult's line of regard to the distal object, just like untreated infants in previous studies (e.g. Brooks & Meltzoff, 2002). It is as if infants in the treatment group had learned that the blindfold could not be seen through by *them*, and assumed that the same would be true for another person. They assumed the adult could not see. Therefore, there was no reason to follow his "gaze" when he turned to face the object; whatever the head turn was about, it was not a turn in order to see.

Making attributions about novel relations

One might argue that these experiments simply hastened a natural development—an understanding that you cannot see through opaque occluders. Could we use evidence similarly to teach infants a perceptual principle that they would not encounter naturally? Could that evidence even override a principle that they had learned earlier? In the natural course of development, by 18 months of age, infants no longer make the error of thinking that adults can see through opaque barriers (Brooks & Meltzoff, 2002; Butler, Caron, & Brooks, 2000; Dunphy-Lelii & Wellman, 2004). Meltzoff and Brooks (2008) capitalized on this to provide 18-month-olds with a completely novel self-experience—one they would not have encountered outside the laboratory. A trick blindfold was constructed that looked opaque from the outside, but was made of special material that could be seen through when held close to the eyes. Infants were randomly assigned to one of three groups: (a) experience with the trick blindfold, (b) experience with the opaque blindfold, and (c) baseline experience in which they played with the trick blindfold while it lay flat on a black table. After receiving the differential evidence, infants in all three groups saw the adult wear the blindfold in the standard gaze-following test.

As expected, infants in the baseline group and the opaque-blindfold groups refrained from following the adult's head turns when the adult wore the blindfold. The new finding is that infants who obtained evidence about the trick see-through blindfold now followed the adult's line of regard to the distal object—they treated the adult as if she could see despite wearing the opaque-looking occluder that covered her eyes.

This underscores the power of infant self-experience in making social attributions to others. Infants had learned that they could make perceptual contact with the external world through the blindfold. By employing the “like-me” framework principle, they immediately transferred this experience to others, despite the fact that the adult’s eyes were covered and it looked, from the outside, like she could not see. Moreover, this new experience now allowed them to override their earlier belief that blindfolds do obscure vision.

These results allow two inferences about development. The first is that infants are projecting their own inner experience to others, suggesting that by 12 months of age infants can attribute mental states (perceptual experience) to others. Crucially, the mentalism demonstrated is of an “on-off” variety, seeing vs. not seeing—a kind of perception-ignorance distinction. The current results do not show perspective-taking about *how* something appears to the other—only *that* it can be seen (or not) in the first place. It is widely argued that infants’ understanding of the basic on-off experience of vision is a building block for more complex mental states such as false belief. Of course, there are other findings suggesting that young children attribute visual experiences to others (e.g. Lempers, Flavell, & Flavell, 1977; Moll & Meltzoff, 2011; O’Neill, 1996; Onishi & Baillargeon, 2005; Repacholi, Meltzoff, & Olsen, 2008; Tomasello & Haberl, 2003). The specific advances of the current work are that it uses young infants (12-month-olds) and a controlled intervention paradigm with random assignment to show that infants use first-person evidence to *change* their understanding of the visual experiences of others.

A second inference concerns the level of abstraction at play. We believe that infants are learning about the spatial-causal relations among three entities: viewer, barrier, and object. These form a “visual perception triangle,” with the spatial relations determining whether the object can be seen by the viewer. Infants abstract a general lesson from the evidence of their own experience: “If the blindfold is interposed between viewer and object, the viewer cannot see the object.” This abstract description applies equally well to self and other. If infants can recognize that the spatial relation is similar—“blindfold over eyes”—they could generalize that the causal effect is similar.

The “like-me” causal framework principle allows infants to treat self and other as similar agents. What I learn about myself is immediately put to work in interpreting your behavior; reciprocally, the outcome of your actions on the world provides me with information about my own powers and the possibilities of my own future actions. This “like-me” framework principle is a human birthright (Meltzoff, 2007), underpinning unique features of human social learning and influencing the course of children’s development (Meltzoff, 2013; Meltzoff & Gopnik, 1993).

The “like-me” assumption supports learning about the world *from* watching other people. This occurs in cases of object-directed imitation and learning about cause and effect from observing social models (Meltzoff et al., 2012) as well as in learning abstract categorization rules from observing others’ sorting behavior (e.g. Williamson, Jaswal, & Meltzoff, 2010). The “like-me” assumption also supports learning *about* other people’s minds. Infants make attributions about the mental states of other “like-me” agents using their *own* first-person experience and mental states as a framework, which is a launching pad for developing an understanding of other minds.

Of course, philosophers have discussed whether an analogy between self and other plays a role in adult human affairs (e.g. Hume, 1739/1969; Husserl, 1950/1960; Smith, 1759/1966). The problem has traditionally been that the framework of equivalence was thought to be a late achievement and perhaps dependent on language, and therefore thought not to play a formative role during infancy. A quarter century of research on infancy has changed this view. In particular, the work on infant imitation indicates that young infants can represent the acts of others and their own acts in commensurate terms (Meltzoff & Moore, 1977, 1997). The generality of human imitation (face, hands, voice, object manipulation, styles of acting) establishes that human infants process a “like-me-ness”

at the level of behavior. They also recognize when their own acts are being reflected back or imitated by others, which prompts emotional and prosocial behavioral reactions by infants (e.g. Meltzoff, 2007) and special neural responses (Saby, Marshall, & Meltzoff, 2012).

What the blindfold training studies contribute is that the equivalence also is registered at the level of mental states. The infants in the blindfold gaze-following studies are not just registering equivalence in terms of behavior or visible configurations (e.g. “blindfold over eyes”), but inferring mental states. They are assuming that if a blindfold over their own eyes affects their vision, then it influences the *vision* of the blindfolded adult in the same way.

It is particularly striking how 18-month-olds react to experience with the trick see-through blindfold. Untreated children realize that opaque-looking occluders cannot be seen through, and do not follow the line of regard of a blindfolded adult. The novel intervention experience runs counter to everyday real-world experience. We arranged it so infants *can* see through this blindfold. Now when the adult dons the blindfold, infants interpret the behavior of the blindfolded adult in a new light. Now infants follow the blindfolded person’s “gaze” to distal objects. Infants attribute a psychological state (vision) to the blindfolded adult and interpret the adult’s behavior as a “turning to look.” In the absence of the novel self-experience, they do not do so.

In the cases we have described so far, evidence came from the child’s self-experience. Is self-experience the sole pathway to understanding others’ minds—a Royal Road? If we think of “like-me” as a framework principle then inferences should go in both directions—either from the self to the other or from the other to the self. We demonstrated just this in both 3-year-olds (Williamson & Meltzoff, 2011) and 18-month-olds (Meltzoff & Brooks, 2012).

In the latter study we arranged a situation in which 18-month-olds watched a blindfolded adult act in distinctive ways. The adult reached out and grabbed the toys, one by one, that were in front of her. To an adult it appeared that she was producing “visually-guided behavior.” It is as if the adult in this treatment group was demonstrating Superman’s X-ray vision. Control groups either performed the same behavior without a blindfold (controlling for “success” in grabbing the toys), or wore the blindfold and fumbled and missed the toys (controlling for “blindfold wearing”). After the exposure to the adult’s particular pattern of behavior (evidence accumulation), all infants were presented with the standard gaze-following test. Results showed that only the infants in the treatment group followed the gaze of the blindfolded adult. This suggests that self-experience is not the sole road for learning about other people’s minds. Infants can abstract information about whether the adult is (or is not) in visual contact with the world based on the cues, contingencies, and structural patterns that the other person exhibits while wearing the blindfold—that is, based on the patterning of *others’* behavior and not solely first-person experience.

To summarize, these experiments show that infants can combine an initial prior “like-me” framework principle with new evidence to infer new causal relations between objects, occluders, and experience both for themselves and others. These inferences go both ways—infants can make inferences about the behavior of others from their own experiences, but they can also make inferences about their own experiences from the behavior of others (e.g. that they will see something interesting if they follow the gaze of the “X-ray vision” adult).

We once suggested that the key thought experiment that would differentiate strong “modularity” theories from the “theory-theory” would be to place children in an alternative parallel universe with evidence that differs radically from our own (Gopnik & Meltzoff, 1997). If children developed a veridical understanding of that universe that would support the theory-theory; if they stuck to their innate understanding of this universe, that would favor the modularity predictions. However, we doubted if the granting agencies would have the funds to support the experiment. In these blindfold experiments, however, we have shown that we can do the same thing, although in a more

low-cost way. In effect, we presented children with alternative universes in which opaque-looking blindfolds are transparent, or in which some adults have the equivalent of Superman's X-ray vision. Even 18-month-olds made the correct inferences about what human behavior and experience would be like in this world.

Social attribution and the understanding of personality traits

If children are making new inferences about the mind from evidence well before they are three, they are also making new inferences well after they are five. These inferences are particularly interesting because they often straddle the unclear line between "theory of mind" and social psychology. One area of particular interest is the inferences we make about personality traits. A long tradition in social psychology (e.g. Kelley, 1967) shows that adults, at least in Western cultures, tend to explain people's actions in terms of their individual "personality traits." Our adult language is permeated with trait judgments, from brave to shy to intelligent to arrogant to introverted. Indeed, if I asked you what someone was like and you answered by giving me a description in terms of a 5-year-old's theory of mind ("well ... she believes that what she sees directly is true, and she usually tries to get what she wants ...") I would hardly be satisfied. Instead, I would expect some discussion of those personality traits that are consistent in her behavior and make her different from everyone else ("she is intelligent and charming but manipulative; he is difficult and bad-tempered but full of integrity"). This would allow me to predict her next move and explain to myself why the person acted toward me like she did.

Adults in Western societies tend to attribute behavior to such personality traits even when the evidence suggests that those actions are really the result of the situations people find themselves in. These attributions can, literally, be a matter of life and death. In the Abu Ghraib trials, for example, many observers initially attributed the atrocities to the sadistic individual personalities of those particular guards, despite the unsettling social psychology evidence suggesting that a wide range of people might behave equally badly in such circumstances.

Where do these attributions come from? It is unclear when and why children begin to explain action in terms of internal, individual, and enduring traits. Of course, even very young children tend to explain action in terms of internal mental states (Flavell, Green, & Flavell, 1990; Lillard & Flavell, 1990). However, trait explanations include two additional factors beyond mental states themselves. Traits are specific to particular, individual people, and they are constant over time and across situations.

Researchers have demonstrated that children do not spontaneously explain actions in terms of traits or endorse trait explanations for a single instance of behavior until middle childhood (Alvarez, Ruble, & Bolger, 2001; Peevers & Secord, 1973; Rholes & Ruble, 1984; Shimizu, 2000). However, other studies show that when preschoolers are given trait labels or behavioral frequency information, they can use that information to make inferences about future behavior, and that they can infer a trait label from frequent behaviors (Boseovski & Lee, 2006; Ferguson, Olthof, Luiten, & Rule, 1984; Gelman, 2003; Heyman & Gelman, 1999; Liu, Gelman, & Wellman, 2007; Matsunaga, 2002). On the other hand, these preschoolers still did not spontaneously construct trait explanations; rather they simply matched the frequency of behaviors to trait labels that were provided for them. This suggests that the failure to attribute traits is not simply a problem with word comprehension or conceptual resources, but may reflect something specific to the child's social cognition. Moreover, when children saw one behavior that could suggest a trait, e.g. one brave action, they did not predict that other behaviors would follow suit, as adults do.

We do not know the learning mechanisms that underlie the course of trait attributions in childhood. Kelley provided an early theory suggesting that person and situation covariation evidence might play an important role in attributions in adults (Kelley, 1967), and there is significant work in this area by social psychologists (e.g. Mischel & Shoda, 1995; Mischel, Shoda, & Mendoza-Denton, 2002; Plaks, Grant, & Dweck, 2005; Ross, 1977). Empirical studies confirm that adults use statistical information tracking multiple people in multiple situations to make trait attributions (e.g. Cheng & Novick, 1990; Hewstone & Jaspars, 1987; Morris & Larrick, 1995; Orvis, Cunningham, & Kelley, 1975; Sutton & McClure, 2001). However, adults already have intuitive theories of traits. They use covariation data to decide when and how to apply those theories to interpret and predict behavior, but could covariation play a role in the *development* of trait attribution itself?

Bayesian causal learning theories suggest that children systematically combine prior knowledge and current covariation evidence to arrive at the right causal hypothesis. This suggests a potential mechanism for the development of trait attribution. Children may begin by observing person and situation covariation evidence that confirms a particular type of hypothesis, particularly in conjunction with adults' linguistic accounting that internal traits cause actions. Once that theory has been strongly confirmed, it will be more difficult to overturn in the future, although it might still be overturned with sufficient evidence. Eventually, in adults raised in Western societies (Nisbett, 2003), this may result in a consistent "trait bias" that requires a very large amount of contrary evidence or concentrated effort to overcome (e.g. Blackwell, Trzesniewski, & Dweck, 2007; Dweck, 2006).

In a series of studies, Seiver, Gopnik, and Goodman (2013) examined the developmental origins of trait attribution in children raised in the USA. First, they conducted a study where 4- and 6-year-old children observed a scenario of two dolls playing on two activities (e.g. a bicycle and a trampoline). Children were either in the person condition (where the two doll characters acted consistently on the two activities, and differently from each other) or the situation condition (where both dolls played on one toy activity and did not play on the other). In some of the conditions this evidence was probabilistic—the doll would play on the toy either three out of four times or one out of four times. The children in each condition received different covariation information about the person and situation. In the person condition, covariation pattern of data indicated that some trait of the individual doll was responsible for the action; in the situation condition, the covariation pattern indicated that the situation was responsible. However, in both conditions, overall, there were the same number of examples of playing and not playing. At the end, we asked the children to explain the doll's actions (e.g. "Why did Josie play on the bicycle?") and to predict their behavior in a future situation (e.g. "What will Josie do when she sees this new diving board? Will she play on it or not?"). We also asked them to predict a new doll's response to the same situations (e.g. "What if Mary sees the trampoline, will she play on it or not?")

In the person condition, one doll always plays and the other doll never plays. This pattern of evidence suggests that something internal about the individual, rather than the situation is responsible for her behavior. In the situation condition, both characters never play with one toy and always play with the other, suggesting instead, that the situation or the toy itself is responsible for their actions. So how would children explain the dolls' behavior in these two different conditions?

Four-year-olds offered explanations that matched the pattern of evidence. In the person condition, when the evidence indicated that something about the person was responsible for the dolls' behavior, both 4- and 6-year-olds gave internal explanations for that behavior. Interestingly, and in keeping with earlier findings, these were rarely classic trait explanations, especially for the 4-year-olds. Instead, children offered explanations that highlighted "trait-like" characteristics of the person, which included both physical characteristics like age or height ("she's the big sister,"

“she’s only little”) and mental states, such as long-standing desires and beliefs (“she likes playing on bicycles,” “she thinks the water is dangerous”).

In the other condition, when the evidence suggested that the situations were driving the dolls’ actions (i.e. they both played on one activity and did not play on the other), 4-year-olds also appropriately gave more external explanations—explanations involving the specific toy activity (e.g. that the bicycle was tippy or the trampoline was safe). In contrast, 6-year-olds persisted in giving internal explanations. Like the Western adults interpreting the events of Abu Ghraib, they attributed the dolls actions to their internal states, even when the pattern of evidence went against those attributions.

In fact, this difference in attribution style between the two age groups in the situation condition suggests that the 4-year-olds were actually more sensitive to the covariation data than the 6-year-olds—they were actually better or, at least, more open-minded learners given the pattern of evidence. Seiver et al. (2013) also included a control condition where children were asked to explain why a single doll did or did not play on a single activity. In this case, the pattern of evidence provided to the children was ambiguous about the possible cause of the behavior. Six-year-olds gave internal explanations significantly more often than expected by chance; 4-year-olds were at chance. The prediction question provided additional evidence for the same developmental change.

This pattern of results suggests that American 6-year-olds have developed a specific attributional theory or person “schema”—that is a broad framework principle—that the internal qualities of a person, rather than the situation, drives behavior. This existing framework principle acted as a filter on their interpretation of the data favoring trait explanations. It did this in much the same way that infants’ “like-me” framework drove them to immediately generalize their own experience to those of others. While we argue that the “like-me” principle has an innate foundation, the trait framework seems to be something that children learn in Western society (best estimate is about 6 years of age). Six-year-olds use both the evidence at hand and their prior beliefs to arrive at a conclusion about a person-situation scenario. The 4-year-olds in contrast, seem to use a more general “bottom-up” data based strategy, and only use the most immediately available data to draw conclusions about other people’s personality.

Domains and development

An interesting characteristic of hierarchical Bayesian learning is that broad framework principles can actually constitute domains. That is, when children learn a new overarching principle that applies to a particular set of data, that principle can act as a constraint on their further inferences. Some principles, like the “like-me” principle could already divide up the world into domains very early. Indeed, there is reason to believe that young infants divide the world into “like-me” and “not-like-me” domains—at a first approximation, animate experiencing agents and inanimate unconscious objects—and treat those domains as if they follow separate rules. However, other principles like the “trait bias” could be learned from cultural-linguistic input and yet have the similar far-reaching effects overall. We can also ask how domain-specific or how general this higher-order bias actually is. Does it only apply to the case of psychological causation, or would children reason similarly about internal versus external causes of physical outcomes?

From people to magnets

To explore potential attributional bias in understanding physical causation (Seiver et al., 2013) changed the outcome of interest to a physical, rather than psychological one—“stickiness” instead of willingness to play. Without changing the task in any other way, they altered the cover story

to implicate physical instead of psychological causation. Rather than saying that the doll character was playing on the scooter, we would say that the doll was sticking to the scooter. The relevant explanatory question became, “Why did the Josie doll stick to the scooter?” For “internal” responses children talked about the properties of the doll; for “external” responses they talked about properties of the toy.

Four-year-olds in this condition behaved as they did in the psychological case. They continued to give more internal explanations in the doll condition, and external explanations in the toy condition. So 4-year-olds seemed to rely on the data, rather than on prior framework principles. However, 6-year-olds behaved differently: They lost their overall preference for internal explanations. Moreover, 6-year-olds now reliably extended the data pattern in both conditions to make future predictions. (Six-year-olds were still less likely to normatively explain the data than the 4-year-olds, however.)

Closer examination of the results suggests interesting details about the 6-year-olds’ shift from largely relying on the pattern of data provided to relying on a prior framework principle. In the physical case, the 6-year-olds gave explanations in terms of a different everyday causal theory—namely, magnetism. They appealed to the properties of magnetism, such as the relationship between magnets and metal, in their explanations and were more likely to give interactive causal explanations that implicated both the doll and the toy as causes for the outcome (e.g. “she has metal shoes and the skateboard is a magnet”). Children never produced these interactive explanations in the social case, and 4-year-olds rarely produced them in the physical case. These explanations suggest that the 6-year-old children relied on a more culturally-conferred, scientifically-based causal framework about stickiness and magnetism in particular, rather than relying on the pattern of observed data per se.

What kinds of evidence could lead to this developmental change? One interesting hypothesis is that the developments at about 6-years of age are related to the increase in peer group interaction. In many peer interactions in the USA, individual traits, rather than social roles or situations, will account for much of the variance in behavior. In a classroom of 28 otherwise similar children placed in a similar situation on the playground, some will consistently take risks and others will not. Children will see more trait-based covariation as they pay increasing attention to their peers, and acquire rich data sets across individuals and situations to draw upon.

Similarly, cross-cultural differences in covariation evidence may influence the development of attribution (Nisbett, 2003). Miller (1984) suggested that children across cultures began with similar attribution patterns and then diverged toward the more extreme adult patterns as they grew older, a claim which has been supported by further studies with children (Gonzalez, Zosuls, & Ruble, 2010; Kalish, 2002; Lockhart, Nakashima, Inagaki, & Keil, 2009).

These results suggest a mechanism by which cultural differences may influence the course of social attribution. This may either be because members of different cultures actually do behave differently, or because culture and experience influence the information children receive from adults about traits, such as adult trait language. This evidence is especially relevant to the development of person schemata. If the adults within a culture tend to linguistically describe and label behavior in terms of traits, this will lead to covariation between certain behaviors and trait labels, which might itself provide evidence for a trait-schema (see Kemp, Goodman, & Tenenbaum, 2008). If children are using covariation information about people’s behavior *and* adult trait language to make inferences about people in situations, such differences in the data could affect the development of their mature adult social cognition. An interesting test would be to explore children in a less trait-based culture, e.g. mainland China. One might predict that 4-year-olds would show a similar pattern to what we observed, but 6-year-olds would not manifest the same trait bias.

Conclusion

We have provided two examples about infants' and children's developing understanding of other minds, one substantially before the well-researched 3–4-year-old age period, and one after it. The first example concerned infants' early attribution of mental states (visual perception) to others, and the second example concerned children's changing interpretation of others' personalities. In both cases, we claim that infants and children are using evidence to develop new inferences and models of other people's minds. These inferences can be specific causal hypotheses about what will happen in a particular situation—the adult will see the distal objects through the opaque-looking occluder, or not; the doll will play on one toy, rather than another. They may also, however, involve inferences from and about general framework principles. Another person will experience the world in the same way that I do. People act based on their individual traits, rather than the situations they find themselves in. When we systematically manipulate the evidence that children receive, they draw different conclusions about the nature of the minds of those around them.

In the real world, children within a particular cultural milieu (shared language, customs, and physical world) may receive reasonably consistent, statistically discernable patterns of evidence about some aspects of the mind, such as visual perception, and so converge on the same general theories as other members in their culture. However, the example of traits and others provided in this chapter also emphasize that many aspects of mental life are likely to vary in different places and different times, and in the myriad of social, physical, and virtual environments that human beings create. Powerful theory-like inferential abilities may be particularly valuable in that sort of world.

One of the most endearing and powerful aspects about the child's social mind is that they change it based on evidence. Adult theorists are challenged to create theories explaining children's conceptual plasticity and developmental trajectory.

Acknowledgement

Work on writing this chapter was supported by NSF grants SMA-0835854 (ANM) and BCS-1023875 (AG). NSF is not responsible for the content. We thank participants in the McDonnell Causal Learning Collaborative for discussions on these topics. The authors contributed equally to this chapter.

References

- Alvarez, J.M., Ruble, D.N., & Bolger, N. (2001). Trait understanding or evaluative reasoning? An analysis of children's behavioral predictions. *Child Development, 72*, 1409–25.
- Amsterlaw, J., & Wellman, H.M. (2006). Theories of mind in transition: A microgenetic study of the development of false belief understanding. *Journal of Cognition and Development, 7*, 139–72.
- Blackwell, L.S., Trzesniewski, K.H., & Dweck, C.S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*, 246–63.
- Bonawitz, E.B., Lim, S., & Schulz, L.E. (2007). Weighing the evidence: Children's naïve theories of balance affect their exploratory play. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 113–18). New York, NY: Erlbaum.
- Boseovski, J.J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental Psychology, 42*, 500–13.
- Brooks, R., & Meltzoff, A.N. (2002). The importance of eyes: How infants interpret adult looking behavior. *Developmental Psychology, 38*, 958–66.

- Brooks, R., & Meltzoff, A.N. (2005). The development of gaze following and its relation to language. *Developmental Science*, 8, 535–43.
- Butler, S.C., Caron, A.J., & Brooks, R. (2000). Infant understanding of the referential nature of looking. *Journal of Cognition and Development*, 1, 359–77.
- Carey, S. (2009). *The Origin of Concepts*. New York: Oxford University Press.
- Cheng, P.W., & Novick, L.R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality & Social Psychology*, 58, 545–67.
- Dunphy-Lelii, S., & Wellman, H.M. (2004). Infants' understanding of occlusion of others' line-of-sight: Implications for an emerging theory of mind. *European Journal of Developmental Psychology*, 1, 49–66.
- Dweck, C. (2006). *Mindset: The New Psychology of Success*. New York: Random House.
- Ferguson, T.J., Olthof, T., Luiten, A., & Rule, B.G. (1984). Children's use of observed behavioral frequency versus behavioral covariation in ascribing dispositions to others. *Child Development*, 55, 2094–105.
- Flavell, J.H., Green, F.L., & Flavell, E.R. (1990). Developmental changes in young children's knowledge about the mind. *Cognitive Development*, 5, 1–27.
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2, 493–501.
- Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (2003). *Bayesian Data Analysis* (2nd ed.). New York: Chapman & Hall.
- Gelman, S. A. (2003). *The Essential Child: Origins of Essentialism in Everyday Thought*. Oxford: Oxford University Press.
- Gonzalez, C.M., Zosuls, K.M., & Ruble, D.N. (2010). Traits as dimensions or categories? Developmental change in the understanding of trait terms. *Developmental Psychology*, 46, 1078–88.
- Goodman, N. (1955). *Fact, Fiction, and Forecast*. Cambridge: Harvard University Press.
- Goodman, N.D., Ullman, T.D., & Tenenbaum, J.B. (2011). Learning a theory of causality. *Psychological Review*, 118, 110–19.
- Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science*, 337, 1623–7.
- 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.
- Gopnik, A., & Meltzoff, A. (1997). *Words, Thoughts, and Theories*. Cambridge: MIT Press.
- Gopnik, A., & Wellman, H. (1994). The theory theory. In L. A. Hirschfeld & S. A. Gelman (Eds), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 257–93). New York: Cambridge University Press.
- Gopnik, A., & Wellman, H.M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, 138, 1085–108.
- Gordon, R.M. (1996). "Radical" simulation. In P. Carruthers & P. K. Smith (Eds), *Theories of Theories of Mind* (pp. 11–21). Cambridge: Cambridge University Press.
- Griffiths, T. L., & Tenenbaum, J.B. (2007). Two proposals for causal grammars. In A. Gopnik & L. Schulz (Eds), *Causal Learning: Psychology, Philosophy, and Computation* (pp. 323–345). New York: Oxford University Press.
- Griffiths, T.L., & Tenenbaum, J.B. (2009). Theory-based causal induction. *Psychological Review*, 116, 661–716.
- Hewstone, M., & Jaspars, J. (1987). Covariation and causal attribution: A logical model of the intuitive analysis of variance. *Journal of Personality and Social Psychology*, 53, 663–72.
- Heyman, G.D., & Gelman, S.A. (1999). The use of trait labels in making psychological inferences. *Child Development*, 70, 604–19.
- Hume, D. (1969). *A Treatise of Human Nature*. London: Penguin Books (original work published 1739).

- Husserl, E. (1960). *Cartesian Meditations: An Introduction to Phenomenology* (D. Cairns, trans.). The Hague: Nijhoff (original work published 1950).
- Kalish, C.W. (2002). Children's predictions of consistency in people's actions. *Cognition*, *84*, 237–65.
- 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., Goodman, N.D. & Tenenbaum, J.B. (2008). Theory acquisition and the language of thought. Paper presented at the 30th Annual Conference of the Cognitive Science Society, Washington, D.C..
- Kemp, C., Perfors, A., & Tenenbaum, J.B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*, 307–21.
- Kuhl, P. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–43.
- Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Laudan, L. (1977). *Progress and its Problems: Toward a Theory of Scientific Growth*. Berkeley: University of California Press.
- Lempers, J.D., Flavell, E.R., & Flavell, J.H. (1977). The development in very young children of tacit knowledge concerning visual perception. *Genetic Psychology Monographs*, *95*, 3–53.
- Leslie, A.M. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences*, *9*, 459–62.
- Lillard, A.S., & Flavell, J.H. (1990). Young children's preference for mental state versus behavioral descriptions of human action. *Child Development*, *61*, 731–41.
- Liu, D., Gelman, S.A., & Wellman, H.M. (2007). Components of young children's trait understanding: Behavior-to-trait inferences and trait-to-behavior predictions. *Child Development*, *78*, 1543–58.
- Lockhart, K.L., Nakashima, N., Inagaki, K., & Keil, F.C. (2009). From ugly duckling to swan? Japanese and American beliefs about the stability and origins of traits. *Cognitive Development*, *23*, 155–79.
- Matsunaga, A. (2002). Preschool children's inferences about personality traits. *Japanese Journal of Developmental Psychology*, *13*, 168–77.
- Meltzoff, A.N. (2007). 'Like me': A foundation for social social cognition. *Developmental Science*, *10*, 126–34.
- Meltzoff, A.N. (2013). Origins of social cognition: Bidirectional self-other mapping and the "Like-Me" hypothesis. In M. Banaji & S. Gelman (Eds), *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us* (pp. 139–44). New York: Oxford University Press.
- Meltzoff, A.N., & Brooks, R. (2008). Self-experience as a mechanism for learning about others: A training study in social cognition. *Developmental Psychology*, *44*, 1257–65.
- Meltzoff, A.N. & Brooks, R. (2012). *Infants inferring mental states from observing others' behavior*. Unpublished manuscript. Seattle: Institute for Learning & Brain Sciences. University of Washington.
- Meltzoff, A.N., & Gopnik, A. (1993). The role of imitation in understanding persons and developing a theory of mind. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding Other Minds: Perspectives from Autism* (pp. 335–66). New York: Oxford University Press.
- Meltzoff, A.N., Kuhl, P.K., Movellan, J., & Sejnowski, T.J. (2009). Foundations for a new science of learning. *Science*, *325*, 284–8.
- Meltzoff, A.N., & Moore, M.K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*, 75–8.
- Meltzoff, A.N., & Moore, M.K. (1997). Explaining facial imitation: A theoretical model. *Early Development and Parenting*, *6*, 179–92.
- Meltzoff, A.N., Waismeyer, A., & Gopnik, A. (2012). Learning about causes from people: Observational causal learning in 24-month-old infants. *Developmental Psychology*, *48*, 1215–28.
- Miller, J.G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology*, *46*, 961–78.

- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*, 246–68.
- Mischel, W., Shoda, Y., & Mendoza-Denton, R. (2002). Situation-behavior profiles as a locus of consistency in personality. *Current Directions in Psychological Science*, *11*, 50–4.
- Moll, H., & Meltzoff, A.N. (2011). How does it look? Level 2 perspective-taking at 36 months of age. *Child Development*, *82*, 661–73.
- Morris, M. W. & Larrick, R. (1995). When one cause casts doubt on another: A normative analysis of discounting in causal attribution. *Psychological Review*, *102*, 331–55.
- Nisbett, R.E. (2003). *The geography of thought: How Asians and Westerners think differently ... and why*. New York: Free Press.
- O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, *67*, 659–77.
- Onishi, K.H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*, 255–8.
- Orvis, B.R., Cunningham, J.D., & Kelley, H.H. (1975). A closer examination of causal inference: The roles of consensus, distinctiveness, and consistency information. *Journal of Personality and Social Psychology*, *32*, 605–16.
- Peevers, B.H., & Secord, P.F. (1973). Developmental changes in attribution of descriptive concepts to persons. *Journal of Personality and Social Psychology*, *27*, 120–8.
- Plaks, J.E., Grant, H., & Dweck, C. (2005). Violations of implicit theories and the sense of prediction and control: Implications for motivated person perception. *Journal of Personality and Social Psychology*, *88*, 245–62.
- Repacholi, B.M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month olds. *Developmental Psychology*, *33*, 12–21.
- Repacholi, B.M., Meltzoff, A.N., & Olsen, B. (2008). Infants' understanding of the link between visual perception and emotion: "If she can't see me doing it, she won't get angry." *Developmental Psychology*, *44*, 561–74.
- Rholes, W.S., & Ruble, D.N. (1984). Children's understanding of dispositional characteristics of others. *Child Development*, *55*, 550–60.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in social psychology*, Vol. 10 (pp. 173–220). New York: Academic Press.
- Saby, J.N., Marshall, P.J., & Meltzoff, A.N. (2012). Neural correlates of being imitated: An EEG study in preverbal infants. *Social Neuroscience*, *7*, 650–61.
- Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–8.
- 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*, 443–54.
- Shimizu, Y. (2000). Development of trait inference: Do young children understand the causal relation of trait, motive, and behavior? *Japanese Journal of Educational Psychology*, *48*, 255–66.
- Simons, D.J., & Keil, F.C. (1995). An abstract to concrete shift in the development of biological thought: The insides story. *Cognition*, *56*, 129–63.
- Slaughter, V., & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. *Child Development*, *67*, 2967–88.
- Smith, A. (1966). *The theory of moral sentiments*. New York: Augustus M. Kelley. (Original work published 1759).
- Sobel, D.M., & Kirkham, N.Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology*, *42*, 1103–15.

- Sutton, R.M., & McClure, J. (2001). Covariational influences on goal-based explanation: An integrative model. *Journal of Personality and Social Psychology*, *80*, 222–36.
- Tenenbaum, J.B., Kemp, C., Griffiths, T.L., & Goodman, N.D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279–85.
- Tomasello, M., & Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what is new for other persons. *Developmental Psychology*, *39*, 906–12.
- Wellman, H.M. (1990). *The Child's Theory of Mind*. Cambridge: MIT Press.
- Wellman, H., & Gelman, S. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, *43*, 337–75.
- Williamson, B.A., Jaswal, V.K., & Meltzoff, A.N. (2010). Learning the rules: Observation and imitation of a sorting strategy by 36-month-old children. *Developmental Psychology*, *46*, 57–65.
- Williamson, R.A., Meltzoff, A.N. (2011). Own and others' prior experiences influence children's imitation of causal acts. *Cognitive Development*, *26*, 260–8.
- Wu, R., Gopnik, A., Richardson, D.C., & Kirkham, N.Z. (2011). Infants learn about objects from statistics and people. *Developmental Psychology*, *47*, 1220–9.