



## Constructing a New Theory From Old Ideas and New Evidence

Marjorie Rhodes,<sup>a</sup> Henry Wellman<sup>b</sup>

<sup>a</sup>*Department of Psychology, New York University*

<sup>b</sup>*Department of Psychology, University of Michigan*

Received 4 April 2012; received in revised form 20 August 2012; accepted 20 August 2012

---

### Abstract

A central tenet of constructivist models of conceptual development is that children's initial conceptual level constrains how they make sense of new evidence and thus whether exposure to evidence will prompt conceptual change. Yet little experimental evidence directly examines this claim for the case of sustained, fundamental conceptual achievements. The present study combined scaling and experimental microgenetic methods to examine the processes underlying conceptual change in the context of an important conceptual achievement of early childhood—the development of a representational theory of mind. Results from 47 children (*M* age = 3.7 years) indicate that only children who were conceptually close to understanding false belief at the beginning of the study, and who were experimentally exposed to evidence of people acting on false beliefs, reliably developed representational theories of minds. Combined scaling and microgenetic data revealed how prior conceptual level interacts with experience, thereby providing critical experimental evidence for how conceptual change results from the interplay between conceptions and evidence.

*Keywords:* Theory of mind; Conceptual development; Social cognition; Microgenetic methods

---

In developing evolutionary theory, a critical step in Darwin's thinking came when the numerous species and subspecies of mockingbirds on the Galapagos Islands led him to realize that the idea of "fixity of species"—that species are fixed and stable over time—was wrong. Darwin's insights proceeded in a progression of ideas that unfolded over years, but recognizing within-species change laid the groundwork for the major conceptual breakthroughs that came next—the ideas that new species can evolve and that the mechanisms that drive within-species change and the origins of new species operate on population-level variation—and thus the paradigm-shift that revolutionized modern

---

Correspondence should be sent to Marjorie Rhodes, 6 Washington Place, rm. 301, New York University, New York, NY 10003. E-mail: marjorie.rhodes@nyu.edu

biology. A process whereby preliminary conceptual insights lay the groundwork for paradigm changes is seen throughout the history of science (Kuhn, 1962).

Potentially, a similar process might drive conceptual change in cognitive development. According to both traditional (Piaget & Inhelder, 1969) and modern constructivist (Xu, 2007) theories, children's initial conceptual framework constrains how they make sense of new evidence. For Piaget, children attempt to interpret new evidence within their current conceptual framework (assimilation), making small modifications to cope with inconsistent data (accommodation), until they are prepared to adopt a new conceptual framework (equilibration). In this view, it is not advances in child language or mental age that are crucial for conceptual change, but the attainment of earlier conceptual insights that provide the groundwork for making sense of new evidence.

Similarly, from the perspective of current rational models (Tenenbaum, Kemp, Griffiths, & Goodman, 2011; Xu, 2007), learning involves the integration of new data with children's prior beliefs (constrained by a hypothesis space), with the product being both the gradual updating of these prior beliefs—which then influence the interpretation of new data—and when necessary, a revision of the underlying hypothesis space (Gopnik et al., 2004). Consistent with this perspective, children's prior beliefs constrain how they learn from new statistical evidence (Schulz, Goodman, Tenenbaum, & Jenkins, 2008; Sobel, Tenenbaum, & Gopnik, 2004; Teglas, 2011; Xu & Tenenbaum, 2007). For example, children learned new causal relationships from statistical evidence more easily when the evidence was consistent with their prior beliefs about the principles that underlie physical causality (e.g., when the causes were spatially contiguous to the effects rather than acting at a distance; Kushnir & Gopnik, 2007). As another example, 3-year-olds infer causal relations from statistical evidence more quickly when the causal relations are consistent with their domain-specific naïve theories (e.g., they more easily learned biological causes for biological effects than psychological causes for biological effects; Schulz, Bonawitz, & Griffiths, 2007; see also, Sobel & Munro, 2009). Furthermore, developmental changes in children's naïve theories of human action influence how 4- and 6-year-olds use patterns of statistical co-variation to explain behavior (Seiver, Gopnik, & Goodman, in press). Even when statistical evidence is inconsistent with children's prior beliefs, however, preschool-age children are able to rationally update their beliefs if they are given multiple training sessions over time (Bonawitz, Fisher, & Schulz, in press). Children's prior beliefs also constrain how they explore and understand new evidence in exploratory-play and question-asking tasks (Bonawitz, van Schijndel, Friel, & Schulz, 2012; Legare, 2012). More generally, the importance of assessing children's initial beliefs about phenomena has been a common theme throughout the educational and developmental literatures (Vosniadou & Brewer, 1992).

Yet there has been surprisingly little experimental research examining these processes as they unfold over time, as opposed to within one or two experimental sessions (for an exception, see Bonawitz et al., 2012, where children participated in four sessions over the course of 2 weeks), or with respect to fundamental conceptual changes that are more akin to the paradigm changes in science described above. With respect to examining fundamental conceptual change over time, many cross-sectional studies have found that beliefs

change as children get older, but almost none have both examined the effects of new evidence experimentally and assessed progressions of conceptual change. Here, we examine these processes in the context of a fundamental conceptual change that occurs in early childhood—the development of a representational theory of mind.

Transition to a representational theory of mind, typically measured via a transition from consistently failing to consistently passing explicit false belief (FB) tasks, provides an excellent opportunity to test models of conceptual change for several reasons. First, the developmental trajectory of this conceptual change in the preschool years is well established (for meta-analysis, see Wellman, Cross, & Watson, 2001). Second, although children show some implicit understanding of FB in infancy (Onishi & Baillargeon, 2005; Scott & Baillargeon, 2009; Scott, Baillargeon, Song, & Leslie, 2010; Song & Baillargeon, 2008; Song, Onishi, Baillargeon, & Fisher, 2008; Surian, Caldi, & Sperber, 2007), the development of an explicit representational theory of mind in preschool remains an important conceptual achievement. Indeed, the gap between infants' implicit understanding of FB and the later development of explicit FB concepts makes the development of theory of mind a particularly intriguing developmental puzzle and highlights the importance of understanding the processes that underlie conceptual development in this domain.

Third, preschool theory of mind developments, as indexed by FB developments, qualitatively change how children interact with their environment—much like the paradigm changes discussed above. The central importance of the development of explicit FB concepts in early childhood is underscored by its real-world implications; the ability to pass explicit FB tasks in preschool is correlated with children's popularity with peers (Peterson & Siegal, 2002; Slaughter, Dennis, & Pritchard, 2002), teacher-rated social competence (Astonington, 2003; Peterson, Slaughter, & Paynter, 2007; Watson, Nixon, Wilson, & Capage, 1999), and skilled interactions with peers (Dunn, Cutting, & Demetriou, 2000), including abilities to play games like hide and seek (Peskin & Ardino, 2003) and social pretend play (Astonington & Jenkins, 1995).

Fourth, the preschool change from consistently incorrect FB judgments to consistently correct ones takes a year or more to accomplish in typically developing children (Wellman et al., 2001). It thus constitutes not only an important change but also a developmentally difficult one that generally requires sustained conceptual development.

Finally, and most important here, the series of conceptual insights that precede the development of FB understanding can be measured via a theory of mind scale (Wellman & Liu, 2004). This scale assesses understanding of (a) Diverse Desires (people can have different desires); (b) Diverse Beliefs (people can have different beliefs); (c) Knowledge-Access (a person will not have knowledge if he or she has not had access to the relevant information); (d) FB (someone can believe something that is false); and (v) Hidden Emotion (someone can feel one way but display a different emotion). In cross-sectional studies, typically developing children reliably proceed in order through this series of understandings (Peterson, Wellman, & Liu, 2005; Wellman & Liu, 2004) and change from one step to the next requires 3–6 months to achieve. In this study, we experimentally examine the role of these preliminary insights in the process of conceptual change to a representational theory of mind.

Although the transition from reliably incorrect to reliably correct FB performance generally takes approximately 1 year, prior research has developed several interventions that facilitate and speed up the development of FB understanding in preschool-age children (Amsterlaw & Wellman, 2006; Lohmann & Tomasello, 2003). Crucially, however, these intervention studies also reveal substantial individual variation in improvement—some children reliably passed FB at posttest, some showed moderate improvement, and some continued to fail. For example, Lohmann and Tomasello (2003) compared several training conditions with young preschoolers, all of whom failed FB at pretest. In the most successful condition, average performance on three FB tasks substantially improved after training; however, even at posttest, individual children's scores ranged from 0 to 3 (similar variation was reported by Amsterlaw & Wellman, 2006). Why do some children develop an understanding of FB following such interventions while others, exposed to the same evidence, do not? This question is fundamental to any theoretical account of the nature of cognitive change.

The possibility that we examine is that children's learning in such interventions is constrained by their previous level of conceptual understanding. That is, that children who are conceptually closer to developing an understanding of FB (i.e., as indicated by their position along the theory of mind scale) will have the prior knowledge that will enable them to make sense of the new evidence presented to them during the study, and thus that this new evidence will prompt conceptual change for these children. In contrast, children who are initially farther away from developing an understanding of FB will not have the requisite prior knowledge to make sense of the new evidence. In this way, we test whether prior knowledge both constrains and enables children's ability to learn from new evidence, in the context of a fundamental, sustained conceptual achievement of early childhood. To test these hypotheses, we recruited a group of children who had not yet developed an understanding of FB, but who varied from one another on the extent to which they had developed prior conceptual understandings. Using the theory of mind scale as our context, we predicted that those who already understood Knowledge Access (KA)—the level that reliably precedes FB understanding in scaling research—would develop an understanding of FB over the course of an extended training period, whereas those who had not yet developed an understanding of KA—but were exposed to the same training sessions—would not. These predictions and the present research help address the fundamental theoretical question of how to characterize cognitive change; in our case, change to a representational theory of mind.

Our methods combined scaling methods with a microgenetic training study. Microgenetic methods rest on fine-grained analyses of cognitive change over multiple successive sessions to provide a rich picture of development and learning as it unfolds (Siegler, 2006; Siegler & Crowley, 1992). Although microgenetic methods have most often been used to examine skill or strategy acquisition (e.g., Luwel, Siegler, & Verschaffel, 2008; Siegler & Chen, 1998; Siegler & Stern, 1998; Siegler & Svetina, 2006), they can also be fruitfully applied to conceptual development (e.g., Opfer & Siegler, 2004, 2007). Our microgenetic methods were inspired by Amsterlaw and Wellman (2006), but critically, we combined this approach with an assessment of children's initial position along the

progression of conceptual insights captured by the theory of mind scale. We included sufficient children to model variation in children's progress and attainment of FB understanding. Our microgenetic method differs from a focused training study (e.g., Lohmann & Tomasello, 2003), in that children receive no explicit teaching about FB concepts. Instead, children see people acting in accordance with FBs (evidence that is inconsistent with a non-representational theory of mind), and we assess whether and how this evidence prompts the development of a representational theory of mind.

## **1. Methods**

### *1.1. Participants*

Participants included 47 children, assigned to Experimental ( $n = 29$ , 15 male,  $M$  age = 3.77 years) or Control ( $n = 18$ , eight male,  $M$  age = 3.76 years) conditions, recruited from private preschools in a midsize city in the midwestern United States. The larger sample in the Experimental condition reflects our primary aim of testing whether children's initial understanding of KA predicted their development of an understanding of FB given microgenetic experiences. Our Control condition (described below) also included children who both did and did not already understand KA, to confirm that initial understanding of KA did not lead to an understanding of FB during the course of the experiment simply due to the passage of time (instead of due to the evidence received during the microgenetic sessions). Thus, previous understanding of KA should predict whether children develop an understanding of FB in the Experimental condition, but not in the Control condition.

### *1.2. Pretest and posttest measures*

To establish that children did not yet have an understanding of FB and to identify their initial conceptual level, all children completed a pretest battery and then completed the same battery at posttest (approximately 8 weeks later). These measures included the five tasks of the theory of mind scale developed by Wellman and Liu (2004; outlined earlier). The scale includes a Contents FB task (children predict whether an agent will think that a box contains its true contents or the contents suggested by its appearance). Children also completed two additional FB tasks: a FB-self task (children report what they previously thought a crayon box contained—crayons or a toy—before they looked inside and found a toy) and a FB-location task (children predict where a character will search for an object—where he left it or in a new location—after the object was moved while the character was not looking). Children were excluded from the final sample if they passed more than one of the three FB tasks (or if they passed one FB task and one of three other related tasks; see the Data S1); 96 children completed pretest measures, 49 were excluded because they (a) passed more than one FB task at pretest, (b) had no consent for participation beyond pre-testing, or (c) left their preschool prior to study completion.

### 1.3. Microgenetic sessions

Children in the Experimental condition ( $n = 29$ ) completed 6 weeks of microgenetic sessions between the pretest and posttest, with two sessions per week. Children in the Control condition had no microgenetic sessions; they completed only the pretest and posttest measures at the same time interval as children in the Experimental condition. In each session, children completed two FB tasks, which varied in form (FB-contents or FB-locations) and presentation (shown in storybooks or acted out with props). Each task presented new characters and scenarios, but all followed the same structure. First, children were asked to predict the thoughts or behavior of an agent who had an FB, as in a standard FB task. Following the child's prediction, however, children were shown the outcome, in which the agent acted based on his or her FB. For example, children predicted whether an agent who had not seen inside a playdoh can say there was playdoh inside (an action based on a FB) or a bouncy ball inside (an action based on reality). After the child's prediction (children who do not yet understand FB predict that the agent will say there is a ball), the experimenter described the outcome (e.g., "Sammy says there is playdoh inside!"). Thus, across the 12 sessions, children were exposed to 24 scenarios where agents acted based on FBs. As all children began the experiment without an understanding of FB, they received 24 pieces of evidence that were inconsistent with their initial theories. For analyses, children received a "1" each time they predicted an action based on an FB and a "0" each time they predicted an action based on reality.

To help children focus on the discrepancies between their initial theories and the agents' actions, children were also asked to explain the agents' behaviors (a sample script is available in the Data S1); 17 children were asked for explanations on every trial and 12 were asked every four trials. There were no differences based on this factor at pretest or posttest for scale-level or percentages of passed FB trials ( $ps > .15$ ). Thus, all children were considered as a single experimental group ( $n = 29$ ). Two independent raters coded children's explanations (see Table 1). Inter-rater agreement was excellent (92%).

#### 1.3.1. True belief

Every second microgenetic session, children answered one story involving true beliefs, to help them resist any expectation that all tasks had some sort of "trick" and to track whether increases in FB accuracy occurred at the expense of general accuracy (which

Table 1  
Coding categories for children's explanations

Explanation	Examples
Belief	"He doesn't know there is a ball in there" "He thinks there is playdoh in there, but there isn't"
Mistake	"He made a mistake"
Desire	"He wants playdoh." "He loves to play with playdoh"
Situational	"There's no playdoh there." "It moved away"
Don't know (or no response)	"I don't know"

could result in true belief decrements). For example, children saw a juice carton, which contained juice, and were asked to predict what a character would think was inside. There was 96% accuracy on True Belief trials.

## 2. Results

### 2.1. Posttest FB understanding

Our key prediction was that children whose initial conceptual level placed them closer to FB understanding would be more likely to develop an understanding of FB following exposure to relevant evidence. Thus, children who passed KA at pretest (the conceptual level immediately preceding FB) and were in the experimental group (and thus exposed to relevant evidence) should be most likely to develop FB understanding. We conducted a binomial regression predicting the number of FB trials passed at posttest, with age as a continuous predictor and condition (experimental, control) and pretest-KA as categorical predictors. As predicted, within the experimental condition, children who passed pretest-KA passed more FB trials at posttest than children who failed pretest-KA,  $p < .001$ , whereas within the control condition, children who passed and failed pretest-KA did not differ from each other (Fig. 1). The Condition \* KA interaction was reliable, Wald  $\chi^2(1) = 6.98$ ,  $p = .008$ , as was the overall model, Likelihood ratio  $\chi^2 = 44.07$ ,  $p < .001$ . Performance on FB at posttest also improved with age, Wald  $\chi^2(1) = 7.51$ ,  $p = .006$ , OR = 3.85. Comparing the standardized deviance estimates across successive models

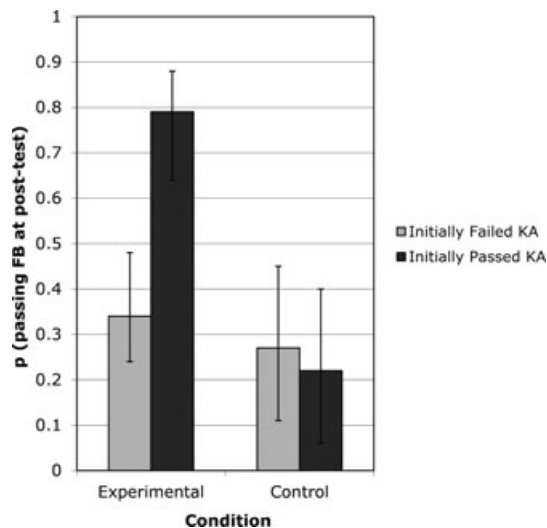


Fig. 1. The probability of passing a false belief task at posttest, with 95% confidence intervals, by condition and whether children passed the knowledge access task at pretest. At pretest, 28 children failed knowledge access (KA) (17 experimental, 11 control) and 19 passed KA (12 experimental, seven control).

(first including only age, then adding condition, then pretest-KA, and finally the condition \* pretest-KA interaction) confirmed that each successive model had improved fit,  $ps < .01$ . Across the sample, only nine children passed all three indicators of FB understanding at posttest (17 passed 0; 21 passed 1 or 2). All nine were in the experimental condition and had passed KA at pretest.

We re-ran this analysis focusing only on the 29 children in the experimental condition, including an indicator of verbal fluency as an additional control variable. Verbal fluency was assessed by tallying average explanation lengths—the number of words a child used per explanation within their 12 microgenetic sessions ( $M = 4.08$  words, range .83–9.50). The overall model was again significant, Likelihood ratio  $\chi^2(3) = 33.74$ ,  $p < .001$ . As above, controlling for age and verbal fluency, children who passed KA at pretest passed significantly more FB trials at posttest ( $M = .77$ ,  $CI = 0.57, 0.90$ ) than those who failed KA at pretest ( $M = .36$ ,  $CI = 0.22, 0.53$ ), Wald  $\chi^2(1) = 8.44$ ,  $p = .004$ ,  $OR = 5.96$ . Age increased the likelihood of passing FB trials at posttest, Wald  $\chi^2(1) = 5.53$ ,  $p = .02$ ,  $OR = 4.32$ , but there was no effect of verbal fluency.

## 2.2. Microgenetic sessions

### 2.2.1. False-belief judgments

Children in the Experimental condition received multiple FB tasks during the microgenetic sessions. Children improved on these FB judgments and therefore accumulated increasing successful experience with FB understanding. Children's prior theory of mind understanding, as indexed by pretest-KA, predicted this within-session improvement (see Fig. 2). The cumulative number of correct FB judgments differed consistently and progressively by whether children initially passed KA. A 2 (Pretest-KA: Pass, Fail) X 12

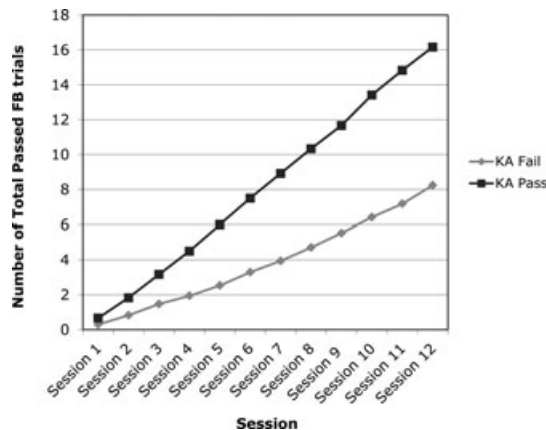


Fig. 2. The average cumulative numbers of correct false-belief (FB) judgments plotted session-by-session across the microgenetic sessions, by whether children passed the knowledge access task at pretest, for children in the experimental condition. The number plotted at session 1 shows the correct FB judgments in session 1, the plot for session 6 shows total correct FB summed across sessions 1 through 6, and so on.



(Sessions 1–12) repeated measures analysis of variance confirmed the depicted time \* pretest-KA interaction,  $F(11, 297) = 10.44$ ,  $p < .001$ ,  $\eta^2 = .28$ . Although both pretest KA failers and passers began by making consistent FB errors, with increasing microgenetic experience, pretest KA passers dramatically increased in FB accuracy. This increase was not at the expense of overall accuracy, as children maintained 96% accuracy on true belief tasks with near ceiling performance across all sessions. Accumulating microgenetic accuracy carried over into posttest; the number of FB trials that children passed during the sessions strongly predicted how many they passed on the posttest,  $r = .80$ ,  $p < .001$ .

### 2.2.2. Explanations

During microgenetic sessions, experimental children provided explanations of the characters' mistaken actions (based on the character's contrary-to-reality beliefs). Passing KA at pretest also affected these explanations. Situational explanations, which overlook the role of mental states in understanding the agents' behaviors, were given more often during their microgenetic sessions by children who failed pretest-KA ( $M = 42.69\%$ ,  $SD = 27.98\%$ ) than children who passed ( $M = 22.25\%$ ,  $SD = 20.59\%$ ),  $t(27) = 2.15$ ,  $p = .04$ . In contrast, children who passed pretest-KA were more likely to give belief-based explanations ( $M = 21.26\%$ ,  $SD = 23.44\%$ ) than children who failed ( $M = 6.85\%$ ,  $SD = 20.76\%$ ). No other explanation type varied by initial level of KA.

### 2.3. Change in ToM scale score from pretest to posttest

Table 2 shows the numbers of children demonstrating each pattern on the scale at pretest and posttest. The percent of children passing the various scale tasks conforms to the expected scale sequence (Wellman & Liu, 2004). A 2 (Time: pretest, posttest) x 2 (Condition: experimental, control) repeated measures ANOVA on scale scores (0–5 for total tasks passed) indicated children's scale scores rose over time ( $M$  pretest = 2.19,  $SE = .10$ ,  $M$  posttest = 2.64,  $SE = .19$ ),  $F(1, 29) = 5.45$ ,  $p = .03$ ,  $\eta^2 = .16$ . Binomial regression models confirmed that being in the experimental condition did not influence the odds of passing any step on the scale at posttest, aside from FB (reported in the prior analyses). Thus, there were no effects of condition,  $ps > .25$ , in this overall analysis of scale scores.

Table 2  
Number of children receiving each scale score at pretest and posttest<sup>a</sup>

Passed	Number of Children at Pretest	Number of Children at Posttest
Nothing	1	0
DD only	4	6
DD and DB	20	9
DD, DB, and KA	11	9
DD, DB, KA, and FB	0	10
Other	11	13

<sup>a</sup>The frequencies of each scale score did not vary by condition at pretest,  $p > .80$  or posttest,  $p > .29$ . DB, diverse beliefs; DD, diverse desires; FB, false belief; KA, knowledge access.

### 3. Discussion

Children's initial conceptual level determined whether and how new evidence provoked conceptual change. Specifically, understanding KA (or not) substantially predicted whether microgenetic exposure to relevant evidence provoked transition to an understanding of FB. Initial conceptual level and the passage of time alone did not provoke this change, as control children who initially understood KA did not develop an understanding of FB. Exposure to evidence alone was also not sufficient, as experimental children who did not initially understand KA also did not develop an understanding of FB. Age was associated with improved understanding of FB, but conceptual change was predicted by the interaction between initial understanding of KA and experimental exposure to relevant evidence, controlling for age.

These data go beyond other studies that have examined the influence of children's prior beliefs on conceptual learning, by examining learning processes as they unfold over time (12 sessions over 6 weeks, as opposed to within a single experimental session). Also, we examined conceptual development with respect to a fundamental conceptual change that prior research has established as an extended and difficult developmental accomplishment; namely, preschool transition to a representational theory of mind. Moreover, we tracked not only children's increased accuracy on FB alone, but their progression along a reliable sequence of preschool theory of mind understandings.

The present data are consistent with the proposal that the development of ToM involves a series of domain-specific conceptual changes. Although these findings do not preclude the possibility that factors external to a conceptual domain—such as children's general processing skills, engagement, inhibitory control, or memory—importantly contribute to conceptual change, the present data suggest that these abilities do not fully account for theory of mind development for several reasons. First, if the Experimental condition supported the development of FB understanding by facilitating these general abilities, it is unclear why it would do so only for children who had previously developed an understanding of KA. Yet children's initial conceptual understanding demonstratively enabled and constrained their potential to learn from new evidence. That is, given the same extended microgenetic experiences, children's prior understandings both enabled learning (for children closer to FB on the theory of mind scale at pretest) and constrained it (for those further away to begin with). Second, we found that our measure of verbal fluency did not predict whether children in the Experimental condition developed an understanding of FB. In future work, it would be useful to include more extensive measures of executive functioning and memory to examine more fully the contribution of these factors in shaping children's conceptual development in this context.

It is worth considering why prior understanding of KA, in particular, enabled progress to FB understanding in our study. We speculate that understanding the relationship between perceptual access and resulting mental states, early acquired in the case of understanding KA, enables recognition of the role of perceptual evidence in belief formation as well (and this also explains why KA reliably precedes FB understanding in

cross-sectional and longitudinal theory of mind scaling research, e.g., Wellman, Fang, & Peterson, 2011). Neither understanding Diverse Desires nor Diverse Beliefs requires appreciation of how mental states depend on perceptual experience. Others have speculated on the formative role of understanding perceptual access as a conceptual “prior” for understanding FB evidence (see Gopnik & Wellman, 1994). Our data underwrite further research to address this specific hypothesis.

The current data provide clear support for a basic, but often unexamined, premise of constructivist models of cognitive development—that conceptual development involves the successful building of new insights through an active process of evidence interpretation, consistent with current rational models (Goodman et al., 2006; Ullman, Goodman, & Tenenbaum, 2010). Our data clearly manifest two empirical signatures of such a process of conceptual development: (a) that prior conceptual knowledge influences whether exposure to new evidence results in conceptual change and (b) that learning proceeds in orderly conceptual progressions. Our data evidence these features in extended, sustained childhood learning of demonstrably difficult, everyday concepts. Thus, these data also illustrate the usefulness of combining scaling and microgenetic methods to empirically capture processes of conceptual change.

## Acknowledgments

We thank Betsey Hamilton, Emily Haas, Laura Dean, and Jessica Emerick for assistance with data collection, Jennifer Amsterlaw for helpful input into study design, and the families and staff at the JCC of Greater Ann Arbor Early Childhood Center and at Pooh Corner Preschool for their participation. Funding was provided by NICHD in grant HD022149 to Wellman.

## References

- 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, 139–172.
- Astington, J. W. (2003). Sometimes necessary, never sufficient: False-belief understanding and social competence. In B. Repacholi, & V. Slaughter (Eds.), *Individual differences in theory of mind: Implications for typical and atypical development* (pp. 13–38). Hove, UK: Psychology Press.
- Astington, J. W., & Jenkins, J. M. (1995). Theory of mind development and social understanding. *Cognition and Emotion*, 9, 151–165.
- Bonawitz, E., Fisher, A., & Schulz, L. (in press). Teaching three-and-a-half-year-olds to revise their beliefs given ambiguous evidence. *Journal of Cognition and Development*, 13(2), 266–280.
- Bonawitz, E., van Schijndel, T., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, 64, 215–234.
- Dunn, J., Cutting, A., & Demetriou, H. (2000). Moral sensibility, understanding others and childrens’ friendship interactions. *British Journal of Developmental Psychology*, 18, 159–177.
- Goodman, N. D., Baker, C. L., Bonawitz, E. B., Mansinghka, V. K., Gopnik, A., Wellman, H., Schulz, L., & Tenenbaum, J. B. (2006). Intuitive theories of mind: A rational approach to false belief. In R. Sun &

- N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 1–31.
- Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. Hirschfeld, & S. Gelman (Eds.), *Domain specificity in cognition and culture* (pp. 257–293). New York: Cambridge University Press.
- Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago, IL: The University of Chicago Press.
- Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology*, *43*, 186–197.
- Legare, C. (2012). Exploring explanation: Explaining inconsistent evidence informs exploratory, hypothesis-testing behavior in young children. *Child Development*, *83*, 173–185.
- Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development*, *74*, 1130–1144.
- Luwel, K., Siegler, R., & Verschaffel, L. (2008). A microgenetic study of insightful problem solving. *Journal of Experimental Child Psychology*, *99*, 210–232.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*, 255–258.
- Opfer, J., & Siegler, R. (2004). Revisiting preschoolers' living things concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology*, *49*, 301–332.
- Opfer, J., & Siegler, R. (2007). Representational change and children's numerical estimation. *Cognitive Psychology*, *55*, 169–195.
- Peskin, J., & Ardino, V. (2003). Representing the mental world in children's social behavior: Playing hide-and-seek and keeping a secret. *Social Development*, *12*, 496–512.
- Peterson, C. C., & Siegal, M. (2002). Mindreading and moral awareness in popular and rejected preschoolers. *British Journal of Developmental Psychology*, *20*, 205–224.
- Peterson, C. C., Slaughter, V. P., & Paynter, J. (2007). Social maturity and theory of mind in typically developing children and those on the autism spectrum. *Journal of Child Psychology and Psychiatry*, *48*, 1243–1250.
- Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory of mind development among children with autism, deafness or typical development. *Child Development*, *76*, 502–517.
- Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. New York: Basic Books.
- Schulz, L., Bonawitz, E., & Griffiths, T. (2007). Can being scared cause tummy aches? Naïve theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology*, *43*, 1124–1139.
- Schulz, L., Goodman, N., Tenenbaum, J., & Jenkins, A. (2008). Going beyond the evidence: Abstract laws and preschoolers' responses to anomalous data. *Cognition*, *109*, 211–223.
- Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development*, *80*(4), 1172–1196.
- Scott, R. M., Baillargeon, R., Song, H., & Leslie, A. M. (2010). Attributing false beliefs about non-obvious properties at 18 months. *Cognitive Psychology*, *61*(4), 366–395.
- Seiver, E., Gopnik, E., & Goodman, N. (in press). 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*.
- Siegler, R. (2006). Microgenetic analyses of learning. In R. Siegler (Ed.), *Handbook of child psychology, vol. 2, cognition, perception, and language*, (6th ed.) (pp. 464–510). Hoboken, NJ: John Wiley & Sons Inc.
- Siegler, R., & Chen, Z. (1998). Developmental differences in rule learning: A microgenetic analysis. *Cognitive Psychology*, *36*, 273–310.
- Siegler, R., & Crowley, K. (1992). Microgenetic methods revisited. *American Psychologist*, *47*, 1241–1243.
- Siegler, R., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General*, *127*, 377–397.
- Siegler, R., & Svetina, M. (2006). What leads children to adopt new strategies? A microgenetic/cross-sectional study of class inclusion. *Child Development*, *77*, 997–1015.

- Slaughter, V., Dennis, M., & Pritchard, M. (2002). Theory of mind and peer acceptance in preschool children. *British Journal of Developmental Psychology*, 20, 545–564.
- Sobel, D., & Munro, S. (2009). Domain generality and specificity in children's causal inferences about ambiguous data. *Developmental Psychology*, 45, 511–524.
- Sobel, D., Tenenbaum, J., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, 28, 303–333.
- Song, H., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology*, 44(6), 1789–1795.
- Song, H., Onishi, K. H., Baillargeon, R., & Fisher, C. (2008). Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition*, 109(3), 295–315.
- Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, 18(7), 580–586.
- Teglas, E. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, 332, 1054–1059.
- 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.
- Ullman, T. D., Goodman, N. D., & Tenenbaum, J. B. (2010). Theory acquisition as stochastic search. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change. *Cognitive Psychology*, 24, 535–585.
- Watson, A., Nixon, C., Wilson, A., & Capage, L. (1999). Social interaction skills and theory of mind in young children. *Developmental Psychology*, 35, 386–391.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72, 655–684.
- Wellman, H. M., Fang, F., & Peterson, C. (2011). Sequential progressions in a theory-of-mind scale: Longitudinal perspectives. *Child Development*, 80, 780–792.
- Wellman, H. M., & Liu, D. (2004). Scaling of theory of mind tasks. *Child Development*, 75, 523–541.
- Xu, F. (2007) Rational statistical inference and cognitive development. In P. Carruthers, S. Laurence & S. Stich (Eds.), *The innate mind: Foundations and the future* (Vol. 3) (pp. 199–215). New York: Oxford University Press.
- Xu, F., & Tenenbaum, J. (2007). Word learning as Bayesian inference. *Psychological Review*, 114, 245–272.

### Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Data S1.** Full list of pre-test and post-test measures and sample scripts.