Cognitive Psychology 64 (2012) 215–234

Contents lists available at SciVerse ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Children balance theories and evidence in exploration, explanation, and learning

Elizabeth Baraff Bonawitz[a,*], Tessa J.P. van Schijndel[b], Daniel Friel[c], Laura Schulz[c]

[a] University of California, 5427 Tolman Hall, Berkeley, CA 94720, United States
[b] University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands
[c] Massachusetts Institute of Technology, 77 Massachusetts Ave.f 46-4013, Cambridge, Massachusetts 02139, United States

ARTICLE INFO

Article history:
Accepted 9 December 2011
Available online 25 January 2012

Keywords:
Beliefs
Evidence
Play
Explanation
Learning

ABSTRACT

We look at the effect of evidence and prior beliefs on exploration, explanation and learning. In Experiment 1, we tested children both with and without differential prior beliefs about balance relationships (Center Theorists, mean: 82 months; Mass Theorists, mean: 89 months; No Theory children, mean: 62 months). Center and Mass Theory children who observed identical evidence explored the block differently depending on their beliefs. When the block was balanced at its geometric center (belief-violating to a Mass Theorist, but belief-consistent to a Center Theorist), Mass Theory children explored the block more, and Center Theory children showed the standard novelty preference; when the block was balanced at the center of mass, the pattern of results reversed. The No Theory children showed a novelty preference regardless of evidence. In Experiments 2 and 3, we follow-up on these findings, showing that both Mass and Center Theorists selectively and differentially appeal to auxiliary variables (e.g., a magnet) to explain evidence only when their beliefs are violated. We also show that children use the data to revise their predictions in the absence of the explanatory auxiliary variable but not in its presence. Taken together, these results suggest that children's learning is at once conservative and flexible; children integrate evidence, prior beliefs, and competing causal hypotheses in their exploration, explanation, and learning.

* Corresponding author.
E-mail address: liz_b@berkeley.edu (E.B. Bonawitz).

doi:10.1016/j.cogpsych.2011.12.002

1. Introduction

1.1. Theory theory

Many researchers have argued that science is possible because of mechanisms developed to support learning in early childhood (e.g. Carey, 1985; Gopnik & Meltzoff, 1997; Wellman & Gelman, 1992). This idea, the "theory theory," suggests that children generate abstract, coherent, theories that (as in science) are defeasible in the face of counter-evidence. The theory theory has gained considerable support from research suggesting that children's early knowledge has structural and functional similarities with scientific theories. In particular, folk theories in early childhood support prediction, explanation, intervention, and counterfactual reasoning (Carey, 1985, 2009; Gopnik & Meltzoff, 1997; Murphy & Medin, 1985; Wellman & Gelman, 1992).

However, theory theory also makes predictions about dynamic properties of theories; in particular, it predicts that children's prior beliefs and evidence should interact to affect learning and exploration (Gopnik & Meltzoff, 1997). These dynamic aspects of the analogy between science and cognitive development have been relatively less investigated. The majority of research on children's causal learning has looked not at how children learn from exploration but at how children learn from evidence provided by others (Bullock, Gelman, & Baillargeon, 1982; Chen & Klahr, 1999; Gopnik et al., 2004; Kushnir, Xu, & Wellman, 2010; Saxe, Tenenbaum, & Carey, 2005; Schulz, Bonawitz, & Griffiths, 2007; Shultz, 1982; Sobel, Tenenbaum, & Gopnik, 2004; Sodian, Zaitchik, & Carey, 1991). The few studies that have looked at the relationship between causal learning and exploration have focused on novel contexts, in which children do not have prior domain-specific beliefs about the competing hypotheses (Gweon & Schulz, 2008; Legare, Gelman, & Wellman, 2010).

1.2. Exploration and learning in science education

Exceptions to this trend are studies that have looked not at exploratory behavior during early childhood (when it is widely believed that play is critical to learning), but at exploration in school-age children as a component of formal science education. Research in science education has found that self-directed exploration does little to support students' learning (e.g. Dunbar & Klahr, 1989; Klahr & Nigam, 2004). Indeed, considerable research suggests that both children and adults have a poor metacognitive understanding of principles of experimental design, have difficulty designing informative, controlled interventions and are poor at anticipating the type of evidence that would support or undermine causal hypotheses (Inhelder & Piaget, 1958; Koslowski, 1996; Kuhn, 1989; Kuhn, Amsel, & O'Laughlin, 1988; Masnick & Klahr, 2003). Additionally, students often fail to seek disconfirming evidence (e.g. Karmiloff-Smith & Inhelder, 1974; Wason, 1960) and fail to learn from disconfirming evidence when it is provided (e.g. Messer, Mohamedali, & Fletcher, 1996; Messer, Norgate, Joiner, Littleton, & Light, 1996; Nickerson, 1998).

However, there are several reasons to think that exploratory components of formal science education may not be particularly informative about the nature of exploratory learning in early childhood. First, science education research tends to focus on students' exploration of relatively complex, multivariate problems (e.g., Kuhn, 1989; Masnick & Klahr, 2003). Such problems are of course appropriate as indicators of classroom performance, but the task complexity may lead to underestimates of children's learning in simpler contexts. Second, passing tests of scientific reasoning typically requires an explicit, metacognitive understanding of the principles involved in causal inference and experimental design; this understanding is presumably absent in early childhood and may well be the exclusive purview of formal education. Finally, science education research often pits children's folk theories against statistical evidence; children are typically credited with "success" only insofar as they suspend their prior beliefs and reason exclusively from the statistical data (e.g. Dunbar & Klahr, 1989; Inhelder & Piaget, 1958; Klahr & Nigam, 2004; Kuhn, 1989; Kuhn, Amsel, & O'Laughlin, 1988). Everyday inference however, typically requires integrating prior beliefs and evidence. Considerable research suggests that even preschool children are capable of accurate causal judgments of this nature (see e.g., Kushnir & Gopnik, 2007; Schulz & Gopnik, 2004; Schulz et al., 2007; Sobel & Munro, 2009).

1.3. Bayesian inference models and ambiguous evidence

Bayesian inference is one approach that describes how statistical evidence interacts with domain-specific theories, and an increasing number of studies have argued that people act in ways consistent with optimal Bayeisan inference (Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Griffiths & Tenenbaum, 2009; Kording & Wolpert, 2004; Weiss, Simoncelli, & Adelson, 2002; Xu & Tenenbaum, 2007). Bayesian inference indicates how the learner updates her beliefs about a set of hypotheses following the data. The learner begins with different degrees of belief about the truth of the these hypotheses, known as the prior, $p(h)$. The probability of any particular hypothesis, h, given some observed data, d, is the posterior, $p(h|d)$. Bayes rule tells us that $p(h|d) \propto p(d|h)p(h)$, where $p(d|h)$, the likelihood, is the probability of observing the data if h were true.

The details of Bayesian inference will not be critical here. However, we emphasize the framework because it provides an intuitive account of how prior beliefs and evidence might interact to affect curiosity and exploration. In particular the learner will be uncertain about the causal structure of an event if the posterior probability of two or more hypotheses is roughly equivalent: $p(h_1|d) \approx p(h_2|d)$. This could occur in two ways: (1) if hypotheses are both *a priori* equally likely and also both equally likely to have generated the observed data (i.e., evidence is confounded); (2) if data provides *strong* support for a hypothesis whose prior probability (given the learner's current theory) is *low* (i.e., evidence violates the learner's expectations) and *weak* support for a hypothesis whose prior probability is *high*. We suggest that cases where a small number of hypotheses have equivalent posterior probabilities should promote curiosity and exploration. One virtue of this account is that it makes it clear that the curiosity provoked by confounding and the curiosity provided by theory violation result from a common inferential process that generates equivalent outcomes: competing posterior probabilities among hypotheses.

Previous work (Schulz & Bonawitz, 2007) has shown that, given a familiar toy with multiple *a priori* equally plausible causal structures, children are more likely to continue exploring it (rather than a novel toy) when evidence is confounded than when evidence is unconfounded (see also Gweon & Schulz, 2008). That is, children engage in selective exploration when the prior probability of competing hypotheses and the likelihood of the evidence under the hypotheses are matched (Cook, Goodman, & Schulz, 2011; Gweon & Schulz, 2008; Schulz & Bonawitz, 2007). Here we investigate the other case of ambiguity: whether children show different patterns of exploratory play when they observe identical evidence but have different prior beliefs. Put formally, the current study looks at whether children engage in selective exploration when the prior probability of hypothesis A is greater than hypothesis B but hypothesis A is less likely given the evidence than hypothesis B.

1.4. Exploration and learning in early childhood

The basic premise, that children will attend more to belief-violating than belief-consistent evidence, is of course not new: indeed, it is central to violation-of-expectation paradigms in infancy. Strikingly however, such paradigms have rarely been extended through early and middle childhood. Although both classic and contemporary research have provided elegant accounts of children's exploratory play, the vast majority of this work has been descriptive rather than experimental (Berlyne, 1969; Bruner, Jolly, & Sylva, 1976; Hutt & Bhavnani, 1972; Piaget, 1962; Power, 2000; Rubin, Fein, & Vandenberg, 1983; Singer, Golinkoff, & Hirsh-Pasek, 2006). Thus relatively little is known about the exploratory behavior of older children. Critically, exploratory play allows children to access different kinds of information than they could get merely through visual inspection; in particular, active exploration might allow children to discover otherwise hidden, but potentially explanatory, variables. Thus a second consideration of the current work is whether children identify hidden variables in the course of exploration and if so, how the discovery of auxiliary variables interacts with children's prior beliefs to affect learning.

1.5. Children's beliefs about balance relationships

Looking at instances of theory violation requires looking at a domain in which children have well-defined prior beliefs. For this purpose, we focus the current investigation on children's beliefs about balance relationships. We chose this domain both because children's beliefs about balance have been well established by prior research (e.g., Case, 1985; Halford, Andrews, Dalton, Boag, & Zielinski, 2002; Jansen & van der Maas, 2002; Karmiloff-Smith & Inhelder, 1974; McClelland, 1989, 1995; Normandeau, Larivee, Roulin, & Longeot, 1989; Pine & Messer, 2000; Raijmakers, van Koten, & Molenaar, 1996; Shultz & Takane, 2007; Siegler, 1976; Siegler & Chen, 1998, 2002) and because balancing blocks are natural and accessible stimuli for exploratory play.

Our research takes off from a seminal study looking at children's folk theories of balance (Karmiloff-Smith & Inhelder, 1974). Karmiloff-Smith and Inhelder (1974) showed that children younger than six ("No Theory[1]" children) balance blocks by trial and error and are thus equally successful (after several attempts) in balancing blocks whether the blocks are symmetrically or asymmetrically weighted. Around age six, children develop a "Center Theory" and believe that blocks will balance if the point of balance bisects the base of the block. These children succeed even on a first attempt at balancing symmetric blocks but fail, even after repeated attempts, with asymmetric blocks. They perseverate on the block's geometric center and are reluctant to adjust towards the center of mass. Thus older "Center Theorists" can take longer to balance an asymmetric block than younger "No Theory" children. Still older children (7- and 8-year-olds) accurately consider the distribution of weight ("Mass Theory") and are likely to succeed, even on their first attempt, at both symmetrically and asymmetrically weighted blocks.

Children's understanding of balance has subsequently been investigated by many researchers. These studies have typically focused on transitions in children's use of rules and strategies in balance scale tasks (e.g. Case, 1985; Halford et al., 2002; Jansen & van der Maas, 2002; Normandeau et al., 1989; Pine & Messer, 2000; Raijmakers et al., 1996; Siegler, 1976; Siegler & Chen, 1998, 2002). Consistent with the Karmiloff-Smith and Inhelder (1974) study, such research shows a consistent progression in children's ability to consider both weight and distance relations in predicting the outcome of balance relations (e.g., from initially considering both variables only when a single variable fails to distinguish the outcome to gradually integrating the two variables). Various formal accounts (see e.g., McClelland, 1989, 1995; Shultz & Takane, 2007) have been advanced to explain the periods of stability and transition in children's rule use.[2]

1.6. Theories, evidence, and exploration

The current work is distinct from this tradition. We do not focus either on the particular content of children's domain-specific beliefs or on transitions in children's domain-general ability to integrate information across multiple variables. Rather, we use children's beliefs about balance as a content domain in which children's beliefs might affect exploration and learning. That is, here we look at whether children's prior beliefs affect the actions they take, the evidence they get as a result, and children's responses to that evidence (in particular, whether they learn from the evidence or try to explain it away).

Note that although some research traditions emphasize children's ability to engage in flexible associative learning from data (with and without feedback; Gruen & Weird, 1964; Weir, 1964; Weir & Stevenson, 1959), other research stresses children's difficulty with belief revision and learners' tendency to misrepresent evidence that conflicts with strongly held beliefs (see e.g. Dunbar & Klahr, 1989; Kuhn, Amsel, & O'Laughlin, 1988; Lord, Ross, & Lepper, 1979; Schauble, 1990; Sloutsky & Spino, 2004). Indeed, some studies suggest that learners may recall evidence better and give it more weight

[1] Because we used the Karmiloff-Smith and Inhelder (1974) study as a starting point, we adopt their practice of referring to the children's beliefs about balance relations as "theories". However, we recognize that folk theories of physics, psychology, and biology (see e.g., Carey, 1985; Wellman & Gelman, 1992) call on a richer, more integrated set of beliefs than the balance relationships we investigate here.

[2] Infancy research has also looked at children's beliefs about balance relations, showing that over the first year of life, infants come to understand that the continuous proportion of the bottom surface left unsupported predicts whether an object will fall or remain supported (i.e., infants will accept that an object can be supported by a small support in the middle, like a "T" but will reject a small support at one end, "Γ"; Needham & Baillargeon, 1993). However, in these cases infants do not need to consider either the relative proportion of unsupported area on either side of a point of balance (as required even for Center Theory) or the distribution of weight as required in traditional balance scale tasks. Indeed, in the one experiment involving asymmetric objects (triangles) infants failed to look longer at impossible events (Baillargeon & Hanko-Summers, 1990). Thus, the infancy research is not in tension with the Karmiloff-Smith and Inhelder results.

when they are not committed to particular theories than when they are (Cooper, 1981; Hastie & Kumar, 1979; Wright & Murphy, 1984). One advantage of the theory theory approach is that it predicts both flexibility and intransigence in children's learning. Theory theory suggests that learners' sensitivity to evidence, including their tendency to accurately represent or explain away data, depends on how evidence (including the availability of auxiliary explanatory variables) is integrated into children's prior knowledge.

Here we predict that children's prior beliefs and evidence should interact to affect children's behavior in three respects. First, consistent with the qualitative predictions of our Bayesian analysis of ambiguity, we predict that, given identical evidence, children with different prior beliefs about the evidence will show different patterns of exploratory behavior. Second, we predict that children will be sensitive to discoveries they make in the course of exploration that might explain away belief-violating evidence. Even if a given variable suffices to explain all the observed evidence, we predict that children will selectively appeal to the variable to explain belief-inconsistent but not belief-consistent evidence. In this respect we suggest that children's beliefs are resistant to anomalous data. Finally, we predict that if children's exploration fails to uncover factors that might explain away belief-violating evidence, children will make new generalizations that are consistent with the evidence rather than with their prior beliefs. That is, if the evidence cannot be explained away, children will learn from it. Such empirical generalizations may not themselves be tantamount to belief-revision but could help lay the groundwork for later theory-change.

In the following studies, we investigate the effect of evidence and prior beliefs on children's exploration, explanation, and learning using a paradigm similar to Schulz and Bonawitz (2007). We familiarize children with an asymmetrically weighted block and then balance the block in a manner consistent with, or in violation of, their prior beliefs. We also introduce a novel distracter toy. We compare children's relative interest in exploring the balance relationship against their interest in exploring the new toy.[3] We predict that children's relative interest in the balancing block will be mediated by their prior beliefs such that they explore the block more when the evidence is inconsistent with their beliefs. A magnet actually holds the block in position in all cases, and the magnet is, of course, always a sufficient explanation for why the block "stays up." We expect the majority of children (in all conditions) to discover the magnet, but we expect children to explain the evidence by referring to the magnet more often in belief-violating than in belief-consistent conditions. Finally, we look at the conditions under which children learn from their exploration.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Ninety-five 6- and-7-year-olds ($M = 84$ months; range = 73–97 months) and thirty-one 4- and-5-year-olds ($M = 62$ months; range = 51–68 months) were recruited from a local urban science museum. Two 2-year-olds and two 5-year-olds were dropped from the study and replaced due to parental interference; one 6-year-old was dropped and replaced because he failed the initial familiarization. An additional thirty-five 6- and-7-year-olds participated in the belief classification task but were not included in additional analyses due to ambiguous belief classification (see below). Approximately equal number of boys and girls participated (46% girls).

2.1.2. Materials

A total of nine asymmetric blocks, made of Styrofoam and colored tape were used: four (each unique in color and shape) were used for the belief classification task; three blue blocks (identical to each

[3] We pit the balanced block against a closely matched novel toy (rather than directly comparing children's play with two blocks balanced in different ways) in order to assess whether the evidence induces causal uncertainty in itself, rather than only relative to an explicitly presented alternative. This methodological advantage, rather than theoretical considerations, motivates this comparison. That is, we believe that it is rational for children to explore novel stimuli; we only suggest that given stimuli otherwise well-matched for salience, children can override novelty preferences to explore evidence that induces causal uncertainty.

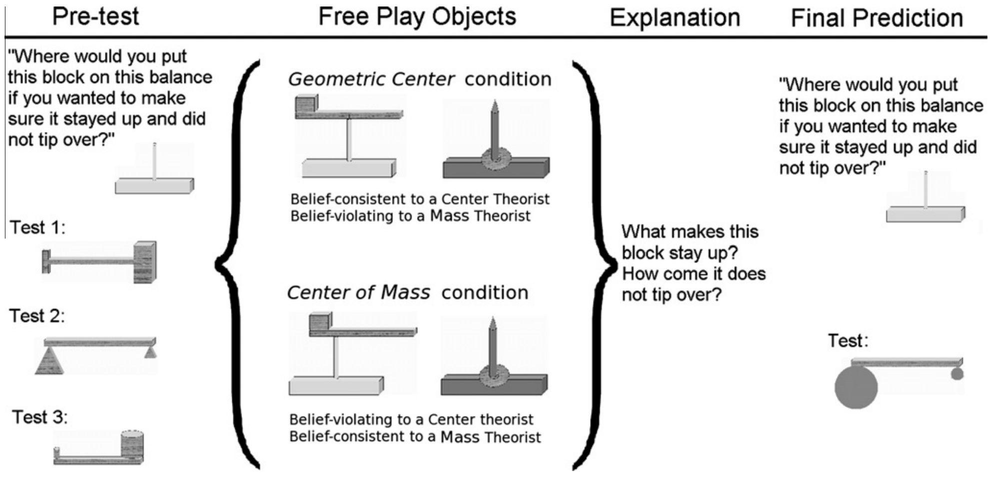

Fig. 1. Method for Experiments 1 and 3.

other), were used for familiarization, and two blue test blocks (identical to the familiarization blocks but containing a magnet either at their geometric center or center of mass) were used in the test condition. The base for the balancing blocks was a rod inserted into a rectangular wooden platform. The novel toy was a metal key ring with several charms; the ring was placed on platform similar to those of the balances. See Fig. 1. An opaque bag was used to cover the novel toy.

2.2. Procedure

2.2.1. Belief-classification

The experiment began with a belief-classification task. Children were presented with three of the four classification blocks (chosen at random) and were asked to try to balance each block on the post. The experimenter watched to see whether the child attempted to balance the block at its geometric center or towards the center of mass. The experimenter took hold of the block just as the child set it on the post so that children never observed the outcome of their balancing attempts.

2.2.2. Familiarization

The experimenter then set aside the balancing post and introduced the children to the three familiarization blue blocks, one at a time. Children were encouraged to explore each of the blocks and were asked to point to the heavier side of each block. Throughout the classification and familiarization periods, the novel toy was on the table, covered and off to the side, out of the child's view.

2.2.3. Play

Children were assigned (randomly or pseudo-randomly; see below) to a *Geometric Center* or *Center of Mass* condition. The experimenter said, "I'm going to try to balance my block here very carefully," and 'balanced' the test block (helped by the magnet) either in the geometric center of the block or over the center of mass. The experimenter then uncovered the novel toy and placed both the balanced block and the novel toy within the child's reach, approximately equidistant from the child. She told the child, "Go ahead and play with which ever toy you want until I come back." The instructions were designed to discourage simultaneous play with both toys to facilitate coding. However, coders looked separately at time with each toy, thus simultaneous play was credited to both of the toys. The base and block of the balance could be separated, and the novel toy ring and platform could also be separated. Play with either part of the balance was considered play with the balance, and play with either part of the novel toy is referred to as play with the novel toy. Children were given one minute to play.

2.2.4. Explanation

After 60 s, the experimenter returned to the table and covered up the novel toy. She returned the test block to its original balanced position and asked, ‘‘Can you tell me, why is this block staying straight? How come it's not tipping over?’’ If a child responded, ‘‘I don't know’’ she was prompted: ‘‘It's okay to take a guess, how come it stays up like this and isn't tipping over?’’

2.2.5. Final prediction

Following the child's explanation, the experimenter removed the blue block and presented the fourth classification block to the child asking: ‘‘Can you balance this very carefully for me, so that it does not tip over?’’

3. Results and discussion

3.1. Belief classification

Consistent with the previous research by Karmiloff-Smith and Inhelder (1974), and corroborated by the current findings (see analysis to follow), all of the 4 and 5-year-olds were classified as ‘‘No Theory’’ children.[4] The 6- and 7-year-old children were classified as ‘‘Center Theorists’’ or ‘‘Mass Theorists’’ based on where they attempted to balance the classification block on all three of the trials; older children who produced inconsistent balance attempts were dropped from further analyses. All attempts within 1'' of the geometric center of the block (a 10% margin of error) were classified as Center balances. All balances outside of this margin of error and towards the heavy side of the block were coded as Mass balances. Children's initial predictions were coded by a research assistant blind to hypotheses and conditions, and 88% were reliability coded by a second researcher blind to condition; reliability was high (Kappa = .94).[5]

Consistent with previous research (Karmiloff-Smith & Inhelder, 1974), the initial predictions of the 4- and-5-year-olds were quite variable: 65% of the children split their responses between the geometric center and the center of mass predictions on the three classification trials (i.e., they gave one response on two of the trials and a different response on a third). Additionally, 34% of the children made at least one balance attempt that was towards the lighter end of the block and thus inconsistent with both Center Theory and Mass Theory.

By contrast, the older children showed somewhat more systematic patterns: none of the older children ever attempted to balance the block towards the lighter side, and 63% of the children consistently balanced the blocks on all three trials. Because the older children who performed inconsistently could not be readily distinguished from the No Theory children, they were dropped from further analysis. Of the 60 children who balanced consistently on all three classification trials, 32 were classified as Center Theorists, and 28 were classified as Mass Theorists. Consistent with Karmiloff-Smith and Inhelder's (1974) original findings, the Center Theorists (M = 81 months) were younger than the Mass Theorists (M = 88 months; $t(58) = 4.9$, $p < .0001$).

3.2. Play

The No Theory children were randomly assigned to a *Geometric Center* condition (n = 16) or a *Center of Mass* condition (n = 15). There were no age differences between conditions (*Geometric Center* mean age = 62.0 months; *Center of Mass* mean age = 62.8 months; $t(29) = 0.43$, $p = ns$). In order to match the number of children in each cell, Center Theorists and Mass Theorists were pseudo-randomly assigned to the two conditions. The child's theory classification was ultimately determined by coding the classification task from videotape; however, the experimenter used the child's performance during the classification task to make her best guess about the child's ultimate classification. The experimenter's

[4] ‘‘No Theory’’ here is shorthand for no differential theory favoring the geometric center or center of mass. The younger children do of course have some theories relevant to balance relations (e.g., that the block must be supported from below).

[5] To ensure children in the older group were theory-consistent, a third researcher, blind to hypotheses and condition served as ‘‘tie-breaker’’ for the few responses in which there was balance classification disagreement between coders.

Table 1
Results from Experiment 1. Seconds of play, percent explanations and final predictions of 4 and 5-year-old "No Theory" children in the *Geometric Center* and *Center of Mass* conditions.

		Geometric center ($n = 16$)	Center of mass ($n = 15$)
Play	Block	14.3s (*13.8*)	13.4s (*15.7*)
	Novel Toy	25.7s (*13.4*)	26.2s (*18.6*)
Explanations	Other	56	47
	Center-consistent	0	7
	Mass-consistent	13	20
	Magnet	31	27
Final predictions	Other	0	7
	Center-consistent	94	67
	Mass-consistent	6	27

Note: Standard Deviation in parentheses.

online judgment was consistent with the videotape classification in all cases. Within each classification, children were then randomly assigned to either the *Geometric Center* or *Center of Mass* condition, resulting in 16 Center Theorists ($M = 80.3$) and 14 Mass Theorists ($M = 90.6$) in the *Geometric Center* condition and 16 Center Theorists ($M = 80.8$) and 14 Mass Theorists ($M = 86.2$) in the *Center of Mass* condition. Center Theory children were matched for age between conditions ($t(30) = 0.76$, $p = ns$) as were the Mass Theory children ($t(26) = 1.7$, $p = ns$).

Children were counted as playing with the toys as long as they were actively engaged with the toys[6]; two coders blind to condition and hypotheses coded the total amount of time each child[7] played with each toy (reliability was high; Balance Toy: $r^2 = .96$, Novel Toy: $r^2 = .97$). We analyzed children's play by looking at children's mean length of the play with the balance and block. Additionally, we looked at whether children discovered the magnet in the course of free play. The magnet was not visible, but pilot work suggested that its force was strong enough that it could be readily felt whenever the block was pulled from the stand. Thus, children were counted as discovering the magnet if during the play period, they pulled the block from the stand at the point where it was held in place by the magnet.

3.2.1. Play results for 4- and 5-year-old "No Theory" children

We compared how long the children played with each toy in each condition by doing a 2×2 mixed ANOVA with play time on each toy as the within-subjects variable and condition as the between-subjects variable. Two-tailed tests are used throughout. For the 4- and 5-year-old "No Theory" children, the analysis revealed a main effect of toy type (averaging across the two conditions, children significantly preferred the novel toy over the balance toy ($F(1,29) = 9.43$, $p < .01$)), but no main effect of condition (overall, children played for the same amount of time in each condition ($F(1,29) = 0$, $p = ns$)) and no interaction ($F(1,29) = .03$, $p = ns$). This preference for the novel toy held up by condition: 4- and 5-year-olds were more likely to play with the novel toy than the balance in the *Geometric Center* condition ($t(15) = 1.88$, $p < .05$) and marginally more likely in the *Center of Mass* condition ($t(14) = 1.64$, $p = .06$). In both conditions, a non-significant majority of children played most with the novel toy. See Table 1.

There were no other differences between conditions. Four- and 5-year-olds spent the same amount of time playing with the balance toy in the *Geometric Center* condition as in the *Center of Mass* condition ($t(29) = .17$, $p = ns$). Additionally, individual children were no more likely to prefer the balance toy in the *Geometric Center* condition than in the *Center of Mass* condition (Fisher Exact ($N = 31$) = ns).

[6] Coders were asked to use their intuitions about what constituted active engagement throughout. For example, if the child's hand was limply resting on the balance toy while they were clearly looking at and touching the novel toy with their other hand, the balance toy was not credited with play, and the novel toy was counted. Alternatively, for instance, if a child used the block from the balance toy to build a structure with the novel toy, play was counted towards both objects.

[7] Two of the 93 children who participated were unable to be reliability coded for play due to technical malfunction of the recordings.

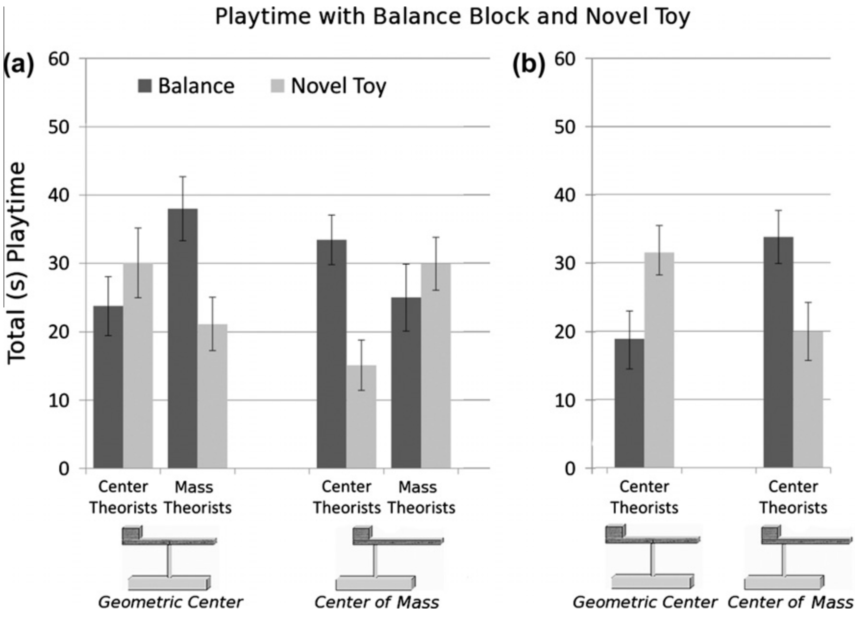

Fig. 2. (a) Play results for Center Theorists and Mass Theorists in Experiment 1. (b) Play results for Center Theorist in Experiment 3.

Table 2
Results from Experiment 1. Seconds of play, percent explanations and final predictions of Center and Mass Theorists in the *Geometric Center* and *Center of Mass* conditions.

		Center Theorists		Mass Theorists	
		Geometric center (n = 16)	Center of mass (n=16)	Geometric center (n=14)	Center of mass (n = 14)
Play	Block	23.8 (*19.0*)	33.4 (*15.9*)	38.0 (*16.1*)	25.0 (*20.2*)
	Novel Toy	30.1 (*19.5*)	15.1 (*16.0*)	21.1 (*13.2*)	29.9 (*19.6*)
Explanations	Other	19	6	0	14
	Center	13	6	0	0
	Mass	19	31	29	71
	Magnet	50	56	71	14
Final predictions	Other	0	0	7	0
	Center	100	63	29	7
	Mass	0	38	64	93

Note. Standard Deviation in parentheses.

Finally, 63% of children in the *Geometric Center* condition and 60% of children in the *Center of Mass* condition were coded as discovering the magnet (Fisher Exact (N = 31) = *ns*).

3.2.2. Play results for 6- and 7-year-old "Center Theory" and "Mass Theory" children

As predicted, 6- and 7-year-old children were more likely to explore the balance and block when the evidence conflicted with their beliefs than when it confirmed their beliefs (see Fig. 2a, Table 2). To compare the amount of time playing with the blocks, we ran a two-way-between subjects ANOVA with theory and condition as the between subjects variables and time spent playing with the blocks as the dependent measure; the analysis revealed no main effect of theory (averaging across the two conditions, Center Theorists and Mass Theorists played for equal amounts of time) and no main effect of evidence type (averaging across Center and Mass Theorists, children played with the block for equal amounts of time regardless of where it was balanced). However, comparisons revealed a significant interaction: children spent more time playing with the block when the evidence conflicted with their theories than when the evidence confirmed their theories ($F(1,59) = 6.02$, $p < .05$).

Within the *Geometric Center* condition, Center Theory children played equally long with both toys ($t(15) = 0.71$, $p = ns$), whereas Mass Theory children played significantly longer with the balancing block than the novel toy ($t(13) = 2.48$, $p < .05$). Within the *Center of Mass* condition, Center Theorists played longer with the balancing block than the novel toy ($t(15) = 2.47$, $p < .05$), whereas Mass Theorists play equally long with both toys ($t(13) = 0.49$, $p = ns$). We also analyzed whether more children were more likely to play with the novel toy or the balancing block using a log-linear analysis on theory by condition by toy type. Six- and 7-year-olds were more likely to prefer the block to the novel toy when evidence conflicted with beliefs than when it confirmed them ($G^2(2) = 6.12$, $p < .05$).

Finally, we looked at 6- and 7-year-olds' tendency to discover the magnet by belief and condition: 75% of Center Theorists and 100% of Mass Theorists discovered the magnet in the *Geometric Center* condition; 94% of Center Theorists and 86% of Mass Theorists discovered the magnet in the *Center of Mass* condition. There were no differences between any of these conditions or groups (Fisher Exact ($N = 60$) = ns). Although almost all of the children discovered the magnet in the course of free play, we predict that children will only appeal to the magnet as an explanatory variable when data are surprising with respect to their beliefs.

3.2.3. Explanation

All children generated a single explanation, and their explanations uniquely and unambiguously fell into one of four mutually exclusive categories: children who generated Center Theory-consistent explanations (e.g. "It balances because it's in the middle; there's the same length on both sides"); children who generated Mass Theory-consistent explanations (e.g. "There's equal amount of weight on both sides"); children who appealed to the hidden cause, the magnet (e.g. "There's something sticky there holding it up, like a magnet"); or children who provided Uninformative explanations (e.g. "It's flat"; "You balanced it slowly and carefully"). Two research assistants, blind to condition and hypotheses, coded children's explanations; reliability was high (Kappa = .97). The two explanations that were coded discrepantly were resolved by the first author, blind to condition.

3.2.4. Explanation results for 4 and 5-year-old, "No Theory" children

The majority (52%) of the No Theory children gave Uninformative explanations.[8] Of the remaining children, 3% were coded as providing Center-consistent explanations; 16% as mass-consistent, and 29% as appealing to the magnet. Comparing the *Geometric Center* and *Center of Mass* conditions, children were equally likely to appeal to each of the four explanation types (Fisher Exact ($N = 31$) $p = ns$, see Table 1). In particular, comparing children who generated magnet explanations to children who generated any of the other explanations did not reveal any differences among the conditions (Fisher Exact ($N = 31$) = ns); see Table 1.

3.2.5. Explanation results for 6 and 7-year-old, "Center Theory" and "Mass Theory" children

Comparing 6- and 7-year-old children's explanations across conditions and theory type revealed a significant interaction $\chi^2(9, N = 60) = 17.8$, $p < .05$ (see Table 1). This was driven by two factors. First, Mass Theorists were marginally more likely to appeal to a Mass consistent explanation (50% of their explanations) than were Center Theorists (25% of their explanations (Fisher Exact ($N = 60$) = .06). There were no other differences between Mass and Center theorists collapsing by condition. See Table 2.

The second factor driving the interaction is that Mass theorists in the two conditions differentially appealed to the magnet. Counter to our predictions, the Center Theorists were equally likely to appeal to the magnet as the explanatory variable in both conditions (*Geometric Center:* 50%; *Center of Mass:* 56%; Fisher Exact ($N = 32$) = ns). Consistent with our predictions, however, the Mass Theorists were

[8] One child in the *Center of Mass* condition responded, "I don't know", which was also coded as uninformative.

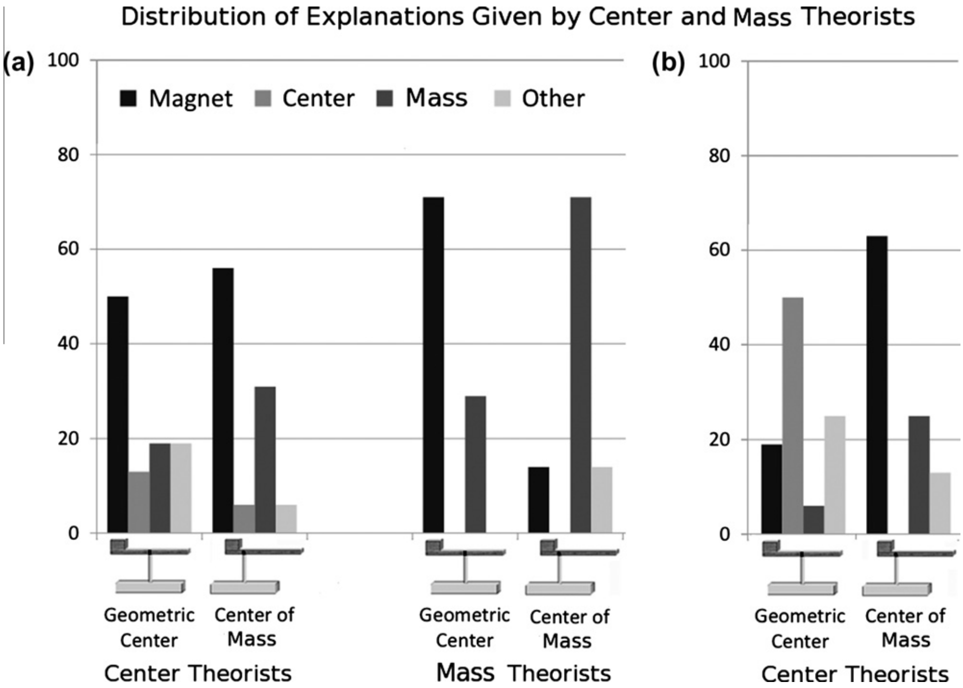

Fig. 3. (a) Distribution of explanations in Experiment 1: Center Theorists were equally likely to invoke the magnet in both conditions; Mass Theorists were more likely to invoke the magnet given theory-violating than theory-consistent evidence. (b) Distribution of explanations by Center Theorists in Experiment 2: children were more likely to invoke the magnet given theory-violating than theory-consistent evidence.

significantly more likely to appeal to the magnet in the *Geometric Center* condition (71%) than the *Center of Mass* condition (14%) (Fisher Exact ($N = 28$) < .01). See Fig. 3a.

3.2.6. Final prediction

Children's final balance attempts were coded as Center consistent, Mass consistent, or Other. A research assistant blind to hypotheses and conditions coded children's final predictions and 81% of the clips were reliability coded by a second researcher; reliability was perfect (Kappa = 1).

3.2.7. Final prediction results for 4 and 5-year-old, "No Theory" children

Consistent with Karmiloff-Smith and Inhelder's finding that even after a relatively short play period, younger children can move from 'No Theory' to a 'Center Theory', the majority of the youngest children's final predictive balances were consistent with the Center Theory prediction (83%). Four and 5-year-olds were significantly more likely to place the block in the center of the stand on their final predictive balance attempt than on their initial classification trials (69%; Binomial, $p = .05$). The move to Center Theory consistent balances seemed to be driven by the evidence that children observed: marginally more children made a Center-consistent final balance prediction in the *Geometric Center* condition (93%) than in the *Center of Mass* condition (67%) (Fisher Exact ($N = 31$) $p = .08$); see Table 1.

3.2.8. Final prediction results for 6 and 7-year-old center theory and mass theory children

Overall, older children who received evidence consistent with their initial beliefs (Center Theory children in the *Geometric Center* condition and Mass Theory children in the *Center of Mass* condition) showed remarkable consistency between their initial classification and final predictions. Only one child of the 30 (3%) changed predictions after observing belief-consistent evidence. In contrast, 11 of the 30 children (37%) changed predictions on the final balance attempt after observing evidence contrasting with their initial beliefs (Center Theory children in the *Center of Mass* condition and Mass Theory children in the *Geometric Center* condition) (Fisher Exact ($N = 60$) < .01). These results held up within theory type: Center Theorists in the *Center of Mass* condition were significantly more likely than Center Theorists in the *Geometric Center* condition to make a correct mass-consistent prediction on the final balance attempt (Fisher Exact ($N = 60$), $p < .05$), and Mass Theorists in the *Geometric Center* condition were marginally more likely than Mass Theorists in the *Center of Mass* condition to make mass-inconsistent balances (center or other) on the final balance attempt (Fisher Exact ($N = 60$), $p = .08$); see Table 2.

3.3. Discussion

The exploratory play data for the 4-and-5-year-old "No Theory" children support the well established finding that children preferentially explore novel objects over familiar ones. They also support the idea that these children do not have strong (evidence differentiating) beliefs about balance. Contrasting these results with those of the 6- and 7-year-old children suggests the influence that children's beliefs can have in overcoming a preference for stimulus novelty. Not only did the older children have different beliefs, their beliefs shaped their choices in play. The Center and Mass Theory children were only a few months apart in age (e.g., could all have been in the same classroom) and, within condition, observed the very same stimuli. Nonetheless, the children systematically differed in their play behavior: given identical evidence, 6-and-7-year-olds showed distinctive patterns of exploratory play depending on their prior theories.

Because the magnet was easy to discover, children's different patterns of exploration did not make them any more or less likely to discover the magnet. As predicted however, the children were not equally likely to appeal to the magnet as an explanatory variable in all conditions: Mass Theorists were more likely to invoke the magnet when the block balanced at the *Geometric Center* than when it balanced at the *Center of Mass*. Indeed, of the five Mass Theorists who did not appeal to the magnet in the *Geometric Center* condition, four tried to explain away the surprising data in other ways, including questioning the evidence of their own senses ("Even though this side is smaller, it must weigh the same"; "If this is the middle, it must weigh the same on both sides somehow"; "Maybe this heavier side (pointing to the geometric center) is actually closer"). Such responses suggest that children are relatively conservative about abandoning their prior beliefs when faced with apparent counter-evidence evidence.

The responses of the Center Theorists however, violated our predictions: they were no more likely to appeal to the magnet given belief-violating evidence than belief-consistent evidence. One possibility is that because Center Theorists were slightly younger, they were more excited by the discovery of the magnet during free play, and thus had more difficulty inhibiting reference to it during the explanation phase. A second possibility is that the younger children had slightly more fragile explanatory abilities than the older children and thus found it easier to mention the magnet than to construct a response in terms of their prior beliefs (e.g., "Because the block is in the center of the post."). We follow-up these possibilities in Experiment 2.

4. Experiment 2

In Experiment 1, the Mass Theorists selectively appealed to alternative explanatory variables given theory-violating evidence but the slightly younger Center Theorists did not. As discussed, this might be due to the demands of inhibiting a salient alternative variable. We hypothesized that if 6-year-old children were made aware that magnets might be present but did not spontaneously discover the magnets themselves, the impulse to refer to the magnet across the board would be reduced. We also hypothesized that children's explanations might be more selective if they had an opportunity to practice giving explanations over the course of the experiment. Finally, we removed the possibility that Center Theorists' explanations might be affected by belief revision in the course of the experiment itself; children were not given the opportunity to play freely with the block so they could not observe balances generated by their own interventions. In Experiment 2, we look at whether Center Theorists selectively appeal to auxiliary variables in the face of anomalous evidence under these task conditions.

4.1. Methods

4.1.1. Participants

Fifty-one 6- and-7-year-olds were recruited from a local science museum. Eight children were classified as Mass Theorists (see below) and were not included in these analyses and eleven children were dropped and replaced for failing to balance the block at the geometric center on all three familiarization trials. The remaining 32 Center Theory children (mean = 83 months; range = 72–97 months) were randomly assigned to a *Geometric Center* (16 children) or *Center of Mass* condition (16 children). Equal number of boys and girls participated (50% girls).

4.1.2. Materials

The materials were identical to those used in Experiment 1 except that a third blue test block (identical to the previous two but without any magnets) was also used. Additionally, three sets of warm-up toys were used: 2 bells (identical except that one made noise and one did not); 2 small toy cars (one that rolled and one that did not); and 2 identical 1” cubes (one magnetic and one not). Three paper clips were also used.

4.2. Procedure

4.2.1. Belief classification task

The Classification task was identical to Experiment 1.

4.2.2. Warm-up task

Children were given a warm-up task to help them practice generating explanations and to familiarize them with magnets. First, the experimenter brought out the bells and showed children that one bell rang but the other did not. The experimenter asked the child “Why do you think this bell rings and this one doesn’t?” If the child offered an explanation the experimenter moved onto the toy cars. If the child could not explain the evidence, the experimenter prompted the child: “Can you come up with any ideas for why this bell works and this one doesn’t?” If the child still failed to answer, the experimenter said, “Maybe it’s because this one does not have the clapper and this one does. Or maybe because the clapper in this one is stuck.” The child was then shown that one toy car rolled and one toy car did not and was asked “Why do you think this one rolls and this one does not?” Again, if children generated an explanation, the experimenter moved on (to the magnets warm-up); if not the child was again prompted and finally provided with feedback: “Maybe it’s because the bottom of the car is sticky. Or maybe because the wheels are glued so that they can’t spin.”

In the last warm-up task, children were shown the two cubes and the clips. One cube attracted the clips; the other did not. Children were told, “See how this block picks up the clips and this one does not? That because this block has a magnet in it which makes the clips stick. This one doesn’t. I’m going to put the magnet block over here (to the right) and the non-magnet block over here (to the left)”. The experimenter then brought out two of the blue test blocks (the inert one and one which had a magnet). The experimenter demonstrated that one of the test blocks picked up the clips and the other did not. The experimenter asked, “Can you tell me which block has a magnet in it and which does not have a magnet in it?” After the child correctly identified the magnetic and non-magnetic blocks, the experimenter asked the child to sort the blocks; a correct response involved placing the magnetic block with the magnetic cube (to the right) and the inert block with the inert cube (to the left). The magnetic and inert object piles remained on the right and left of the table for the remainder of the experiment.

4.2.3. Test phase

The experimenter pulled out the third blue test block and showed it to the child, asking “Can you show me, which is the big, heavy side and which is the light side?” The experimenter then carefully

Table 3
Results from Experiment 2. Percentages of explanations and sorting responses of Center Theorists in the *Geometric Center* and *Center of Mass* conditions.

Condition	Explanations				Sorting	
	Magnet	Center	Mass	Other	Magnetic	Non-magnetic
Geometric center $n = 16$	19	50	6	25	36[a]	64[a]
Center of mass $n = 16$	63	0	25	13	81	19

[a] $n = 14$.

'balanced' the block on the stand for the child. (As in Experiment 1, in both conditions, the block was held in place by the magnet.) Half the children saw the block balancing at the geometric center (*Geometric Center* condition) and half balancing at the center of mass (*Center of Mass* condition). Children were then asked for an explanation: "Can you tell me, why is this block staying up? How come it's not falling over?" Following the child's explanation, the experimenter asked the child "Can you tell me, which group do you think this belongs in? The group with the magnetic blocks or the group that is not magnetic?"

4.3. Results and discussion

Children's initial prediction results were coded as in Experiment 1. Only children balancing the block at the geometric center on all three classification trials were included (74% of those tested). Children's initial predictions were coded by a research assistant blind to hypotheses and conditions and 88% of the clips were reliability coded by a second researcher; reliability was high (96% agreement[9]). Sixteen Center Theorists were randomly assigned to the *Geometric Center* condition, and 16 to the *Center of Mass* condition. There were no age differences between conditions (*Geometric Center* mean age: 82.2 months; *Center of Mass* mean age: 84.6 months; $t(30) = 0.88$, $p = ns$). All children successfully generated explanations in the warm-up tasks and successfully sorted the magnetic and nonmagnetic cubes and test blocks. Children provided a single explanation and children's explanations were coded by two research assistants, blind to condition and hypotheses; 94% of the clips were reliability coded for explanation and reliability was high (Kappa = .77)[10]; disagreements were resolved by the first author, blind to condition.

As hypothesized, significantly more children appealed to the magnet in the *Center of Mass* condition (63%) than in the *Geometric Center* condition (13%) (Fisher Exact ($N = 32$) < .01; see Fig. 3b). Significantly more children also made a Center consistent explanation in the *Geometric Center* condition (50%), than in the *Center of Mass* condition (0%), Fisher Exact ($N = 32$), $p < .01$. There were no differences in the proportions of children who generated a Mass consistent explanation (*Geometric Center*: 6%; *Center of Mass*: 25%) or Uninformative explanation (*Geometric Center*: 31%; *Center of Mass*: 13%) between conditions (Fisher Exact ($N = 32$), $p = ns$ for both conditions); see Table 3.

Two children in the *Geometric Center* condition did not complete the sorting task and were dropped from subsequent analyses. As predicted, and consistent with the explanation results, children in the *Center of Mass* condition were more likely than children in the *Geometric Center* condition to sort the balanced block with the magnetic objects. In the *Center of Mass* condition the majority of children (81%) classified the block as magnetic; significantly fewer of the children (36%) did so in the *Geometric Center* condition (Fisher Exact ($N = 30$) = .01); see Table 3.

The results of Experiment 2, together with the performance of the Mass Theorists in Experiment 1, suggest that 6- and 7-year-old children selectively explain away surprising evidence by appealing to hidden variables. Even though the magnets were present throughout, children appealed to the

[9] Because virtually all of the responses fell into a single category (center balances) we report the overall correlation; Kappa is not appropriate given the distribution of the data.

[10] The moderate Kappa score is driven by our small N's. Coders agreed on 100% of the magnet explanations, our principal measure of interest this Experiment. Thus disagreements in coding do not affect the primary results reported above.

alternative explanatory variable in belief-violating but not in belief-consistent conditions. Children's ability to invoke hidden variables to explain away belief-violating evidence suggests a rational process by which children might maintain their beliefs in the face of apparent counter-evidence. One paradoxical implication of these results is that in some contexts, reducing children's physical engagement in play (see e.g., Bjorklund & Brown, 1998; Pellegrini & Smith, 1998), and thus reducing the salience of auxiliary variables, can improve children's inferential reasoning, allowing them to make more selective judgments about the causal explanatory role of these variables.

5. Experiment 3

In Experiment 3, we look at what children might be capable of learning from evidence generated in exploratory play if no auxiliary variables are provided. We replicate the procedure of Experiment 1 but never give children a chance to observe evidence they can explain away. To do this, we test Center Theorists using two new asymmetric blocks. (We test Center Theorists and not Mass Theorists to avoid deceptively teaching children an incorrect theory of balance.) One block only balances over its center of mass (e.g., like any asymmetrically weighted block) and one balances only over the geometric center (because it is surreptitiously weighted so that although one end looks heavier, the center of mass is in the middle of the block). In neither case is a magnet involved in the balancing. Thus, regardless of what actions children take during play, they can only generate evidence consistent with their condition: in the *Geometric Center* condition the block balances only at the geometric center; in the *Center of Mass* condition the block balances only at the center of mass. We predict that, replicating Experiment 1, Center theorists will preferentially explore the block only in the Center of Mass condition. However, because here the children cannot explain away the anomalous evidence by appealing to the magnet, we predict that, relative to the children Experiment 1, children in the Center of Mass condition of Experiment 3 will be more likely to change their predictions and generate mass-consistent final balances.

5.1. Methods

5.1.1. Participants

Thirty-two 6- and 7-year-olds (mean = 84 months; range = 74–94 months) were recruited from a local urban area science museum. An additional nine children were dropped and replaced for failing to balance the block at the geometric center on all three familiarization trials; one child in the *Geometric Center* condition was dropped and replaced for failure to interact with the stimuli during the play period. Children were randomly assigned to either the *Geometric Center* condition ($n = 16$) or a *Center of Mass* condition ($n = 16$). Equal number of boys and girls participated (50% girls).

5.1.2. Materials

The materials were identical to Experiment 1, except that the magnet was placed at the top surface of the block, so that it would not interfere with balancing and could not be used to explain away surprising evidence. The test block for the *Geometric Center* condition was surreptitiously weighed so that although it looked heavier on one side, it actually balanced over its geometric center. Additionally, six new colored blocks, three of which were clearly equally weighed and balanced at the center, three of which were clearly unevenly weighed and balanced towards one end were used for a final sorting task in the *Geometric Center* condition.

5.1.3. Procedure

The procedure was identical to Experiment 1, with one exception. After the final prediction at the end of the experiment, children in the *Geometric Center* condition were asked to sort the six new colored blocks into two piles: blocks that had a heavier side and blocks that were equally heavy on both sides. Children were then asked to sort the test block. This allowed us to make sure that children had not discovered the surreptitious weighting during play and continued to believe that the block was heavier on the larger side (like the block in the *Center of Mass* condition). All children 'passed' the final

Table 4
Results from Experiment 3. Seconds of play, percent explanations and final predictions of Center Theorists in the *Geometric Center* and *Center of Mass* conditions.

		Geometric center (n = 16)	Center of mass (n = 16)
Play	Block	19.4 (*17.6*)	33.8 (*15.2*)
	Novel Toy	30.7 (*18.8*)	19.9 (*17.0*)
Explanations	Other	50	38
	Center-consistent	31	6
	Mass-consistent	19	50
	Magnet	0	6
Final predictions	Other	0	0
	Center-consistent	100	38
	Mass-consistent	0	63

Note: Standard Deviation in parentheses.

sort (i.e., failed to discover that the surreptitiously weighted block was actually evenly weighted); that is, they sorted the test block with the other objects that were heavier towards the larger side.

5.1.4. Results

There were no age differences between groups ($t(30) = -.85$, $p = ns$). Children's initial prediction results were coded as in Experiment 1 and 2; as stated above, only children balancing the block at the geometric center on all three classification trials were included (78%). Children's initial predictions were coded by a research assistant blind to hypotheses and conditions and 88% of the clips were reliability coded by a second researcher also blind to hypothesis and condition; reliability was high (98% agreement Footnote 9).

5.1.5. Play

Two coders blind to condition and hypotheses coded the total amount of time each child played with each toy (reliability was high: $r^2 = .97$). Replicating the results of Experiment 1, Center Theorists were more likely to explore the balancing block when the evidence conflicted with their beliefs than when it confirmed their beliefs. We ran a two-way-between subjects ANOVA on playtime with type of evidence as the between subjects variables and time spent playing with the blocks and novel toy as the dependent measures. Comparisons between conditions revealed no main effect of condition (children in the *Geometric Center* condition played as long as children in the *Center of Mass* condition), and no main effect of toy type (averaging across conditions, children played as long on average with the blocks as with the novel toy). However, comparisons revealed a significant interaction ($F(1,60) = 8.51$, $p < .01$); children spent more time playing with the block over the novel toy when evidence conflicted with beliefs than when evidence was consistent with beliefs. Children in the *Geometric Center* condition played longer with the novel toy than the balancing block ($t(15) = 1.84$, $p < .05$), while children the *Center of Mass* condition played significantly longer with the balancing block than the novel toy ($t(15) = 1.83$, $p < .05$). See Fig. 2b and Table 4.

5.1.6. Explanations

Children's explanations were coded as in Experiment 1. One child in the *Center of Mass* condition refused to provide an explanation. Two research assistants, blind to condition and hypotheses, coded children's explanations; reliability was high (Kappa = .95). The one explanation that was coded discrepantly was resolved by the first author, blind to condition.

Although there was no magnet present in the block at the point of balance, one child in the *Center of Mass* condition spontaneously explained that the block stayed up because of a magnet. For the reported analyses, this child was coded as providing an ''Uninformative'' explanation. The remaining children fell uniquely and unambiguously into the Mass consistent, Center consistent, or Uninformative categories. There was a marginally significant effect of condition on explanation: more children in the *Center of Mass* condition appealed to a Mass consistent explanation following play (50%) than did children in *Geometric Center* condition (19%); in contrast, more children made Center consistent explanations in the Geometric Center condition (31%) than the *Center of Mass* condition (6%) (Fishers Exact ($N = 32$), p = .07). There were no differences between conditions with respect to Uninformative explanations (*Geometric:* 50%; *Mass:* 44%); see Table 4.

5.1.7. Final predictions

As in Experiment 1, children's final balance attempts were coded as either Mass consistent, Center consistent, or Other. The majority (63%) of Center Theorists in the *Center of Mass* condition changed their final prediction to a Mass consistent prediction. In sharp contrast, no child in the *Geometric Center* condition changed his or her predictions on the final balance; all children made a Center consistent prediction. Significantly more children changed predictions on the final balance attempt after observing conflicting evidence than confirming evidence (Fisher Exact ($N = 32$), $p < .001$). Following the conflicting evidence, the majority of children who made a Mass consistent final prediction also made a Mass consistent explanation (7 of 10 children), while only 1 of the 6 children who made a Center consistent final prediction gave a Mass explanation (Fisher Exact ($N = 16$), $p = .06$); see Table 4.

Given that the Center Theorists in the *Center of Mass* condition observed evidence counter to their beliefs and those in the *Geometric Center* condition did not, it is perhaps unsurprising that children in the *Center of Mass* condition changed their final predictions and those in the *Geometric Center* condition did not. However, it is interesting to compare the Center Theorists in the *Center of Mass* condition of Experiment 3 with those in Experiment 1. In Experiment 1, eight Center Theorists in the *Center of Mass* condition generated evidence (in free play) that they could not explain away (i.e., they tried to balance the block over the geometric center and failed because the magnet was located under the center of mass); 75% of these children made mass consistent final predictions, comparable to the 63% of Center Theorists in the *Center of Mass* condition of Experiment 3 who also observed evidence they could not explain away (Fisher Exact = *ns*). By contrast, eight Center Theorists in the *Center of Mass* condition of Experiment 1 observed only evidence that they could explain away (with the magnet); none of these children made a Mass consistent final prediction. Comparing the Center Theorists who could explain away the data (the 8 children from Experiment 1 and 16 from Experiment 3) with those who could not (8 children in Experiment 1), suggests that children were more likely to change their final balance in response to the data in the absence of a potentially explanatory auxiliary variable (Fisher Exact ($N = 32$), $p < .01$).

5.2. Discussion

Experiment 3 replicated the finding that, given identical evidence, 6-and 7-year-olds' prior beliefs about balance relationships mediate their pattern of exploratory play: children were more likely to explore the familiar balance when evidence was surprising with respect to their prior beliefs than when it was consistent with them. Additionally, both children's explanations and their final predictions support the claim that children generalized from the evidence they generated in exploratory play. Taken together, the Center Theorists' patterns of responding in Experiments 1 and 3 suggest a dynamic relationship between prior beliefs, the presence of auxiliary variables, and learning. When an auxiliary variable was present, Center Theorists who observed belief-violating evidence were just as resistant to changing their beliefs as Center Theorists who observed only belief-consistent evidence. However, when no auxiliary variables were available to explain away the surprising evidence, Center Theorists produced mass consistent explanations and mass consistent final predictions. These results suggest that young children respond rationally to anomalous evidence they observe during play; they explain it away when they can and revise their predictions when they cannot.

5.2.1. General discussion

We began by considering the metaphor of the "child as scientist". We suggested that children's folk theories should play a critical role not only in supporting their causal judgments but also in guiding their exploratory behavior and their tendency to learn from or explain away theory-violating evidence. Consistent with this idea, we found that children's prior beliefs mediate their exploratory play; children were more likely to explore the familiar balance when they observed evidence that conflicted with their prior beliefs than when they were presented with belief-consistent evidence or when they lacked strong differential beliefs. We also found that 6- and 7-year-old children selectively appealed to auxiliary hypotheses to explain away evidence that conflicted with their prior beliefs. Finally we showed that children were more likely to learn from theory-violating data when a potentially explanatory auxiliary variable was absent than when it was present. Taken together, these results suggest that children's learning is at once conservative and flexible; children integrate evidence, prior beliefs, and competing causal hypotheses in their exploration, explanation, and learning.

In these respects, the current study is consistent with other recent research suggesting that children selectively explore and seek to explain belief-violating evidence (e.g., Legare et al., 2010). In contrast to previous studies however, this study extends beyond investigation of novel, arbitrary causal relationships (e.g., i.e., blocks that activate toys) to look at children's real world beliefs. Strikingly, children quite close in age, given identical task instructions, looking at identical evidence, can interpret it differently and thus exhibit different patterns of both exploration and explanation.

Moreover, although many studies suggest that children try to explain away evidence that violates their prior beliefs, and will even invoke unobserved variables to do so (Bullock et al., 1982; Koslowski, 1996; Kuhn, 1989; Kushnir & Gopnik, 2007; Legare et al., 2010; Schulz, Goodman, Tenenbaum, & Jenkins, 2008; Schulz & Sommerville, 2006; Sobel, Yoachim, Gopnik, Meltzoff, & Blumenthal, 2007), to our knowledge, this study is the first to show that children's ability to learn from theory-violating evidence trades-off with the accessibility of auxiliary variables. Six and 7-year-olds were much more likely to revise their predictions when unexpected evidence could not be easily accounted for than when it could. Additionally, while other studies suggest that children appeal to auxiliary variables when evidence violates their prior beliefs, this study shows that they do so even when auxiliary variables are uniformly good explanations for the observed evidence. (Magnets after all, can explain why objects stay up even in expected locations.) Future research might investigate general principles illustrating how prior belief in a theory, the strength of the anomalous evidence, and the availability of alternative hypotheses interact to affect children's responses to belief-violating evidence.

A learner might experience uncertainty and choose to explore for different reasons: because the observed evidence fails to distinguish plausible causal hypotheses or because the observed evidence violates the learner's prior beliefs. Bayesian inference offers an account of curiosity in which both these cases of uncertainty are due to a common inferential process: the integration of the learner's prior beliefs in the probability of the hypotheses (the *prior*) and the probability that the learner would observe the evidence if the hypothesis were true (the *likelihood*). Previous studies showed that children selectively explored evidence when the priors and likelihoods of competing hypotheses were roughly equivalent (Cook et al., 2011; Gweon & Schulz, 2008; Schulz & Bonawitz, 2007); the current work extends these ideas by showing that children engage in selective exploration when the prior probability of one hypothesis is higher, but its likelihood is lower, than another.

A complete understanding of the relationship between theory development, exploratory play, and theory change remains a challenge to the field. However, this work suggests that belief guided exploration may play an important role in helping children generate evidence to support causal learning. Insofar as exploration can lead to the discovery of potentially explanatory hidden variables, it might help children rationally maintain their beliefs in the face of spurious counter-evidence. Insofar as exploration can reveal new, unexplained, and previously unexpected causal relationships, it might support belief revision. Thus, in looking at how children investigate mass and balance in the physical world, we may also learn something about how children weigh evidence and balance it against their prior beliefs in everyday learning.

Acknowledgments

Special thanks to the participating daycares in Cambridge, MA, to the Museum of Science, Boston, and the Boston Children's Museum, to Suejean Lim for substantial contributions to data-collection, coding, and discussion, and to Irene Headen, Yunji Wu, Noah Goodman, Darlene Ferranti, Swe Tun, Si Wang, Meekerley Sanon, and members of the Early Childhood Cognition Lab. This research was additionally supported by an American Psychological Foundation Koppitz Fellowship to E.B., a John S. Templeton Foundation award and an NSF Faculty Early Career Development Award to L.S., and a James S. McDonnell Foundation Collaborative Interdisciplinary Grant on Causal Reasoning to E.B. and L.S.

References

Baillargeon, R., & Hanko-Summers, S. (1990). Is the top object adequately supported by the bottom object? Young infants' understanding of support relations. *Cognitive Development, 5*, 29–53.

Berlyne, D. (1969). *Conflict, arousal and curiosity*. New York: McGraw-Hill.

Bjorklund, D. F., & Brown, R. D. (1998). Physical play and cognitive development: Integrating activity, cognition, and education. *Child Development, 69*(3), 604–606.

Bruner, J., Jolly, A., & Sylva, K. (1976). *Play – Its role in development and evolution*. New York: Basic Books, Inc.

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (Ed.), *The developmental psychology of time* (pp. 209–254). New York, NY: Academic Press.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press/Bradford Books.

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Case, R. (1985). *Intellectual development: Birth to adulthood*. London: Academic Press.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*(5), 1098–1120.

Cook, C., Goodman, N., & Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition, 120*(3), 341–349.

Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin, 90*(2), 218–244.

Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery processes. In D. Klahr & K. Kotovsky (Eds.), *The 21st Carnegie-Mellon symposium on cognition: Complex information processing: The impact of Herbert A. Simon*. Hillsdale, NJ: Lawrence Erlbaum.

Goodman, N., Tenenbaum, J., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science, 32*(1), 108–154.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*(1), 1–31.

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

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116*(4), 661–716.

Gruen, G. E., & Weird, M. W. (1964). Effect of instructions, penalty, and age on probability learning. *Child Development, 35*(1), 265–273.

Gweon, H., Schulz, L. E. (2008). Stretching to learn: Ambiguous evidence and variability in preschoolers' exploratory play. In *Proceedings of the 30th annual conference of the cognitive science society* (pp. 570–574).

Halford, G. S., Andrews, G., Dalton, C., Boag, C., & Zielinski, T. (2002). Young children's performance on the balance scale: The influence of relational complexity. *Journal of Experimental Child Psychology, 81*, 417–445.

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology, 37*(1), 25–38.

Hutt, C., & Bhavnani, R. (1972). Predictions from play. *Nature, 237*, 171–172.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Jansen, B. R. J., & van der Maas, H. L. J. (2002). The development of children's rule use on the balance scale task. *Journal of Experimental Child Psychology, 81*, 383–416.

Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory. *Cognition, 3*(3), 195–212.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15*, 661–667.

Kording, K., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature, 427*, 244–247.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*(4), 674–689.

Kuhn, D., Amsel, E., & O'Laughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of others. *Psychological Science, 21*, 1134–1140.

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, 44*, 186–196.

Legare, C. H., Gelman, S. A., & Wellman, H. M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Development, 81*, 929–944.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*(11), 2098–2109.

Masnick, A. M., & Klahr, D. (2003). Error Matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development, 4*(1), 67–98.

McClelland, J. L. (1989). Parallel distributed processing: Implications for cognition and development. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 8–45). Oxford, UK: Oxford University Press.

McClelland, J. L. (1995). A connectionist perspective on knowledge and development. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 157–204). Hillsdale, NJ: Lawrence Erlbaum.

Messer, D. J., Mohamedali, M., & Fletcher, B. (1996). Using computers to help pupils tell the time: Is feedback necessary? *Educational Psychology, 16*(3), 281–296.

Messer, D. J., Norgate, S., Joiner, R., Littleton, K., & Light, P. (1996). Development without learning? *Educational Psychology, 16*(1), 5–19.

Murphy, G., & Medin, D. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289–316.

Needham, A., & Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition, 47*, 121–148.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of Genetic Psychology, 2*(22), 175–220.

Normandeau, S., Larivee, S., Roulin, J., & Longeot, F. (1989). The balance scale dilemma: Either the subject or the experimenter muddles through. *Journal of Genetic Psychology, 150*(3), 237–250.

Pellegrini, A. D., & Smith, P. K. (1998). Physical activity play: The nature and function of a neglected aspect of play. *Child Development, 69*, 577–598.
Piaget, J. (1962). *Play, dreams and imitation in childhood*. New York: Norton.
Pine, K. J., & Messer, D. J. (2000). The effect of explaining another's actions on children's implicit theories of balance. *Cognition and Instruction, 18*(1), 35–51.
Power, T. G. (2000). *Play and exploration in children and animals*. Mahwah, NJ: Lawrence Erlbaum.
Raijmakers, M. E. J., van Koten, S., & Molenaar, P. C. M. (1996). On the validity of simulating stagewise development by means of PDP Networks: Application of catastrophe analysis and an experimental test of rule-like behavior. *Cognitive Science, 20*(1), 101–136.
Rubin, K. H., Fein, G. G., & Vandenberg, B. (1983). Play. In E. M. Hetherington (Ed.), P. H. Mussen (Series Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social development* (pp. 693–774). New York: John Wiley & Sons.
Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science, 16*(12), 995–1001.
Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49*, 31–57.
Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers play more when evidence is confounded. *Developmental Psychology, 43*(4), 1045–1050.
Schulz, L. E., Bonawitz, E. B., & Griffiths, T. (2007). Can being scared cause tummy aches? Naïve theories, ambiguous evidence and preschoolers' causal inferences. *Developmental Psychology, 43*(5), 1124–1139.
Schulz, L. E., Goodman, N. D., Tenenbaum, J. B., & Jenkins, C. A. (2008). Going beyond the evidence: Abstract laws and preschoolers' responses to anomalous data. *Cognition, 109*(2), 211–223.
Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology, 40*(2), 162–176.
Schulz, L. E., Gopnik, A., & Glymore, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science, 10*(3), 322–332.
Schulz, L. E., & Sommerville, J. (2006). God does not play dice: Causal determinism and children's inferences about unobserved causes. *Child Development, 77*(2), 427–442.
Shultz, T. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47*(1).
Shultz, T. R., & Takane, Y. (2007). Rule following and rule use in simulations of the balance-scale task. *Cognition, 103*(3), 460–472.
Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8*, 481–520.
Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: A microgenetic analysis. *Cognitive Psychology, 36*(3), 273–310.
Siegler, R. S., & Chen, Z. (2002). Development of rules and strategies: Balancing the old and the new. *Journal of Experimental Child Psychology, 81*(4), 446–457.
Singer, D. G., Golinkoff, M. R., & Hirsh-Pasek, K. (2006). *Play = Learning: How play motivates and enhances children's cognitive and social-emotional growth*. New York: Oxford University Press.
Sloutsky, V. M., & Spino, M. A. (2004). Naive theory and transfer of learning: When less is more and more is less. *Psychonomic Bulletin and Review, 11*(3), 528–535.
Sobel, D. M., & Munro, S. A. (2009). Domain generality and specificity in children's causal inferences about ambiguous data. *Developmental Psychology, 45*, 511–524.
Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science, 28*(3), 303–333.
Sobel, D. M., Yoachim, C. M., Gopnik, A., Meltzoff, A. N., & Blumenthal, E. J. (2007). The blicket within: Preschoolers' inferences about insides and causes. *Journal of Cognition and Development, 8*, 159–182.
Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development, 62*, 753–766.
Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12*, 129–140.
Weir, M. W. (1964). Effect of patterned partial reinforcement on children's performance in a two-choice task. *Child Development, 35*(1), 257–264.
Weir, M. W., & Stevenson, H. W. (1959). The effect of verbalization in children's learning as a function of chronological age. *Child Development, 30*(1), 143–149.
Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*(6), 598–604.
Wellman, H., & Gelman, S. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology, 43*, 337–375.
Wright, J. C., & Murphy, G. L. (1984). The utility of theories in intuitive statistics: The robustness of theory-based judgments. *Journal of Experimental Psychology: General, 113*(2), 301–322.
Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*(2), 245–272.